

The Jackson Laboratory

The Mouseion at the JAXlibrary

Faculty Research 2021

Faculty Research

8-2021

In situ Chromatin Interaction Analysis Using Paired-End Tag Sequencing.

Ping Wang

Yuliang Feng

Kun Zhu

Haoxi Chai

Ya-Ting Chang

See next page for additional authors

Follow this and additional works at: <https://mouseion.jax.org/stfb2021>



Part of the [Life Sciences Commons](#), and the [Medicine and Health Sciences Commons](#)

Authors

Ping Wang, Yuliang Feng, Kun Zhu, Haoxi Chai, Ya-Ting Chang, Xiaofei Yang, Xiyuan Liu, Chen Shen, Eva Gega, Byoungkoo Lee, Minji Kim, Xiaoan Ruan, and Yijun Ruan

In situ Chromatin Interaction Analysis Using Paired-End Tag Sequencing

Ping Wang,¹ Yuliang Feng,^{1,2} Kun Zhu,^{1,3,4} Haoxi Chai,¹ Ya-ting Chang,^{1,5} Xiaofei Yang,^{1,6} Xiyuan Liu,^{1,7} Chen Shen,^{1,8} Eva Gega,¹ Byoungkoo Lee,¹ Minji Kim,¹ Xiaoan Ruan,¹ and Yijun Ruan^{1,9,10}

¹The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut

²Key Laboratory of Molecular Target and Clinical Pharmacology, School of Pharmacy, Guangzhou Medical University, Guangzhou, Guangdong, China

³Institute for Diabetes, Obesity, and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

⁴Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

⁵Department of Translational Research and Cellular Therapeutics, City of Hope, Duarte, California

⁶Translational Medicine Collaborative Innovation Center, The Second Clinical Medical College (Shenzhen People's Hospital), Jinan University, Shenzhen, China

⁷State Key Laboratory of Ophthalmology, Optometry and Vision Science, Eye Hospital and School of Ophthalmology and Optometry, Wenzhou Medical University, Wenzhou, Zhejiang, China

⁸Department of Dermatology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology (HUST), Wuhan, China

⁹Department of Genetics and Genome Sciences, University of Connecticut Health Center, Farmington, Connecticut

¹⁰Corresponding author: yijun.ruan@gmail.com

Chromatin Interaction Analysis Using Paired-End Tag Sequencing (ChIA-PET) is an established method to map protein-mediated chromatin interactions. A limitation, however, is that it requires a hundred million cells per experiment, which hampers its broad application in biomedical research, particularly in studies in which it is impractical to obtain a large number of cells from rare samples. To reduce the required input cell number while retaining high data quality, we developed an *in situ* ChIA-PET protocol, which requires as few as 1 million cells. Here, we describe detailed step-by-step procedures for performing *in situ* ChIA-PET from cultured cells, including both an experimental protocol for sample preparation and data generation and a computational protocol for data processing and visualization using the ChIA-PIPE pipeline. As the protocol significantly simplifies the experimental procedure, reduces ligation noise, and decreases the required input of cells compared to previous versions of ChIA-PET protocols, it can be applied to generate high-resolution chromatin contact maps mediated by various protein factors for a wide range of human and mouse primary cells. © 2021 The Authors. Current Protocols published by Wiley Periodicals LLC.

Basic Protocol 1: Sample preparation and data generation

Support Protocol: Bridge linker preparation

Basic Protocol 2: Data processing and visualization

Keywords: ChIA-PET • chromatin interaction • *in situ* ChIA-PET

How to cite this article:

Wang, P., Feng, Y., Zhu, K., Chai, H., Chang, Y., Yang, X., Liu, X., Shen, C., Gega, E., Lee, B., Kim, M., Ruan, X., & Ruan, Y. (2021). *In situ* chromatin interaction analysis using paired-end tag sequencing. *Current Protocols*, 1, e174. doi: 10.1002/cpz1.174

INTRODUCTION

Mammalian genomes are spatially organized into compartments, contact domains, and chromatin loops. The three-dimensional (3D) chromosomal organization in the nucleus provides the structural framework for many fundamental biological processes such as transcription (Tang et al., 2015), DNA replication (Sima et al., 2019), development (Bonev et al., 2017), and cell fate (Dixon et al., 2015). Our understanding of the 3D spatial organization of chromatin relies mostly on high-throughput chromosome conformation capture assays such as Hi-C (Lieberman-Aiden et al., 2009) for genome-wide chromatin contacts and ChIA-PET (Fullwood et al., 2009) for specific chromatin interactions mediated by protein factors. The original ChIA-PET protocol extracts only 20-bp tags from the interacting chromatin loci, for which tag mapping efficiency is limited. Because of this, an improved version of ChIA-PET that extracts longer reads (up to 150 bp) from interacting fragments, called “long-read ChIA-PET,” was developed, which increases the tag mapping efficiency and enables the identification of haplotype-specific chromatin interactions (Li et al., 2017; Tang et al., 2015).

To further uncover the important roles of 3D chromatin structure in different mammalian processes and diseases, it is necessary to map chromatin interactions *in vivo* and identify functional elements, for instance, in primary cells and/or clinical samples, often from a limited numbers of cells. Long-read ChIA-PET, however, cannot readily fulfill these needs because it requires 100 million cells per experiment, and it is challenging to obtain that number of cells from limited tissue material (for primary cell cultures) or clinical samples. Inspired by the *in situ* Hi-C protocol (Rao et al., 2014), we adopted *in situ* digestion and ligation steps into the ChIA-PET procedure, which significantly improve the efficiency of proximity ligation for genuine chromatin interactions and greatly reduce random ligation noise. This modification not only significantly reduces the required cell number to as few as 1 million, but also generates high-quality chromatin interaction data, comparable to that produced by the long-read ChIA-PET protocol using 100 million cells.

The *in situ* ChIA-PET protocol, described here, has been established as a standard data production protocol in the 4D Nucleome and ENCODE consortia, and has been successfully applied to generate hundreds of CTCF and RNAPII datasets from human cell lines and primary cells as part of ENCODE phase 4 projects (<https://www.encodeproject.org/>). Here, we describe detailed step-by-step procedures for performing *in situ* ChIA-PET from cultured cells, including both an experimental protocol (Basic Protocol 1) for sample preparation and data generation and a computational protocol (Basic Protocol 2) for data processing and visualization using the ChIA-PIPE pipeline (Lee et al., 2020). In Basic Protocol 1, we describe the detailed procedures for cell crosslinking (for both suspension and adherent cells) and *in situ* ChIA-PET library construction, followed by high-throughput sequencing. We use bridge linkers to connect two interacting chromatin fragments during proximity ligation. The preparation procedure for bridge linkers is described in the Support Protocol. In Basic Protocol 2, we describe the necessary scripts and tools for data analysis, quality evaluation, and data visualization. A general outline of the complete ChIA-PET workflow is shown in Figure 1.

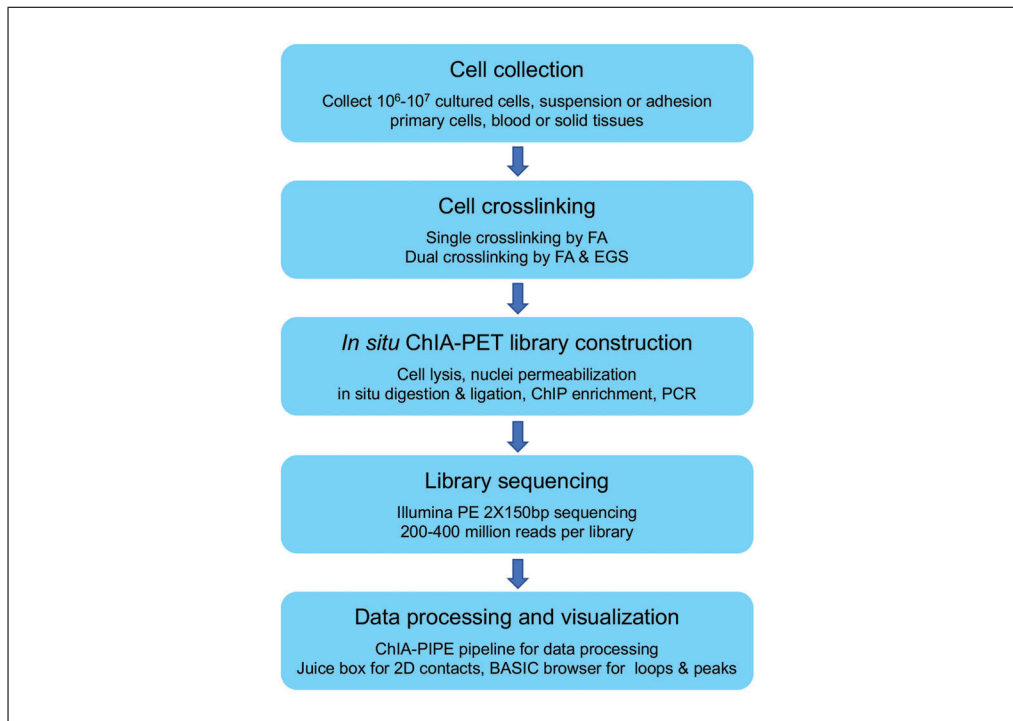


Figure 1 Outline of the *in situ* ChIA-PET workflow.

STRATEGIC PLANNING

When planning a research project that involves *in situ* ChIA-PET, there are several elements to be considered before starting, including the cells to be studied, sonication conditions for chromatin sample preparation, design and assembly of the bridge linker used in proximity ligation (Support Protocol), the protein factors of interest and the availability of antibodies against them, and capacity for high-throughput DNA sequencing and high-volume data analysis.

First, the current version of the *in situ* ChIA-PET protocol requires 10^6 - 10^7 cells as starting material. Most cell line cultures (suspension or adhesion) and some primary tissues (e.g., blood for immune cells, liver for hepatocytes, etc.) can provide sufficient numbers of cells for this. However, cell-type-specific optimization of chromatin sample preparation may be required, for example, in the case of suspension versus adhesion cells (see details in specific protocols below). The chromatin contact conformations within the nuclei of cells are often preserved by formaldehyde (FA) treatment, forming covalent bonds between DNA and associated proteins. To further capture chromatin-associated protein-protein interactions, ethylene glycol bis(succinimidyl succinate) (EGS) has been used in the ChIA-PET protocol.

Second, the optimal sonication parameters for each individual cell type should be empirically determined. This step is critical to obtaining suitable chromatin fragments to maximize antibody capture efficiency during the chromatin immunoprecipitation (ChIP) step. Different sonication parameters, such as shearing time and power setting, should be optimized to get the ideal chromatin fragments (around 2-3 kb for this experiment) for each cell line.

Third, the most important consideration in a ChIA-PET experiment is the target protein factor to be studied and the availability of a robust antibody for it. Many proteins, such as CTCF, RNA polymerase II (RNAPII), cohesin, and transcription factors (Ji et al., 2016; Tang et al., 2015; Wang et al., 2020; Weintraub et al., 2017) can be studied by ChIA-PET for their roles in mediating chromatin interactions, as long as the target protein

has a suitable antibody for ChIP-seq assays. Therefore, the quality and availability of a ChIP-grade antibody is critical. Another important consideration regarding the antibody is the specific “batch” of antibody used, due to quality variations in antibody production. Once a robust antibody batch is identified, what we typically do is stockpile enough of this specific batch for a given ChIA-PET project. The robustness of an antibody for ChIP enrichment in a ChIA-PET experiment can be assessed by ChIP quantitative PCR (ChIP-qPCR; Mukhopadhyay, Deplancke, Walhout, & Tissenbaum, 2008). In ChIP-qPCR, at least two positive target loci and one negative target locus (binding sites) in a genome for the protein factor of interest should be selected, and pairs of PCR primers should be designed based on the specific sequences at the flanking regions of the selected loci. ChIP-enriched chromatin DNA samples along with the non-enriched input genomic DNA can then be quantitatively amplified by qPCR using these primer pairs, so as to accurately measure the fold of ChIP enrichment by a specific antibody in a ChIA-PET experiment.

Fourth, once a quality *in situ* ChIA-PET library is generated, a large enough number of paired-end-tag (PET) reads (2×150 bp) by high-throughput DNA sequencing is required to provide enough data for genome-wide coverage, which ranges from 100 million to 1 billion paired-end-tag (PET) reads. Routinely, we produce 200-500 million PET reads per library, which should provide adequate coverage of protein binding peaks and chromatin loops between peak loci. Furthermore, for quantitative measurement, biological or technical replicates (at least 2 replicates) of *in situ* ChIA-PET data are desired for reliable data analysis and conclusions. Therefore, a sufficient budget for DNA sequencing should be set aside.

Lastly, one should also be prepared for high-volume *in situ* ChIA-PET data analysis. There are several publicly available computational pipelines (Lee et al., 2020; Li et al., 2010, 2017; Phanstiel, Boyle, Heidari, & Snyder, 2015), including the most recent computational pipeline, called ChIA-PIPE (Lee et al., 2020), for ChIA-PET data processing and result visualization, which we describe in Basic Protocol 2.

BASIC PROTOCOL 1

SAMPLE PREPARATION AND DATA GENERATION

The following *in situ* ChIA-PET protocol describes step-by-step procedures for preparation of cellular samples and library construction for sequencing on the Illumina platform (Fig. 2). For convenience, we describe the experimental protocol in two parts: cell crosslinking (Part 1) and library preparation (Part 2). In Part 1, to ensure the capture

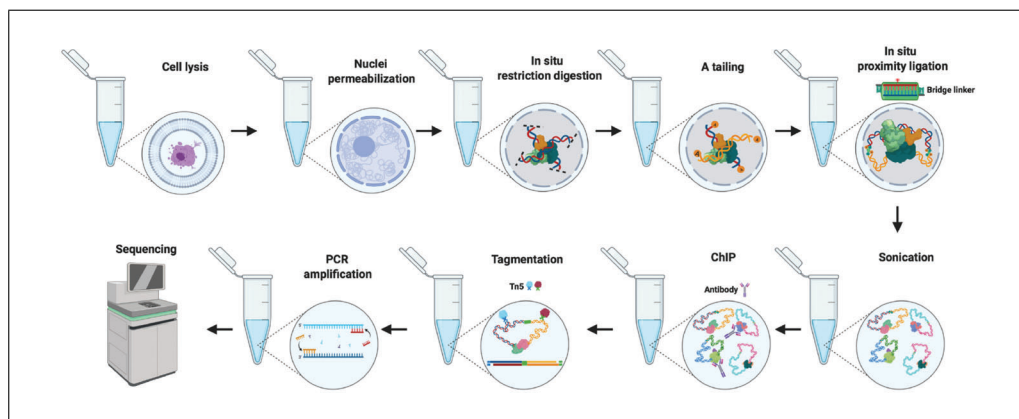


Figure 2 Schematic overview of Basic Protocol 1. The nuclei isolated from dual-crosslinked cells are permeabilized, and *AluI* (a four-base-pair cutter of restriction enzyme) is added to the reaction for *in situ* digestion. Proximity ligation is performed *in situ*, with a bridge linker after the A-tailing step. Chromatin immunoprecipitation (ChIP) against a specific protein factor is carried out after sonication of the DNA. The de-crosslinked ChIP DNA is tagged by Tn5 transposase, and the ChIA-PET library is then amplified by PCR and is ready for sequencing analysis.

of long-range chromatin contacts through protein-DNA and protein-protein interactions, we employ a dual crosslinking method, which we describe here for both suspension cells and adherent cells. The dual-crosslinking process uses formaldehyde (FA) and ethylene glycol bis(succinimidylsuccinate) (EGS) to maximize the capture of chromatin interactions that involve DNA-protein and protein-protein contacts participating in chromatin interactions. Formaldehyde forms covalent bonds between DNA and associated proteins, and thus captures DNA-protein interactions, whereas EGS preserves protein-protein interactions by forming relatively weak but long-arm bonds. Although we describe the procedure for cultured cells, the protocols can be adapted for primary cells with minor modification. Once crosslinked, the nuclear conformation is preserved, and the cellular samples can be stored at -80°C for months before proceeding to library construction. In Part 2, we describe detailed steps required for ChIA-PET library construction, which involve cell lysis, nuclei permeabilization, *in situ* digestion, *in situ* proximity ligation, chromatin sonication, ChIP enrichment, DNA tagmentation, and PCR amplification, which result in the final *in situ* ChIA-PET library, ready for high-throughput DNA sequencing.

Materials

Cultured suspension or adhesion cells (1-10 million cells/reaction) and appropriate culture medium
1% FA-DPBS solution (see recipe)
2 mM EGS-DPBS solution (see recipe)
Dulbecco's phosphate-buffered saline (DPBS), calcium- and magnesium-free (Gibco, cat. no. 14190-250)
2.5 M glycine (see recipe)
Formaldehyde (FA; 36% v/v; Sigma-Aldrich, cat. no. 47608-250ML-F)
0.1% SDS cell lysis buffer (see recipe)
0.55% SDS solution (see recipe)
cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail (PI; Roche, cat. no. 11836170001)
10% Triton X-100, molecular biology grade (Sigma-Aldrich, cat. no. 648464)
10× CutSmart buffer (New England Biolabs)
AluI restriction enzyme (NEB, cat. no. R0137L)
TE buffer, pH 8.0, RNase-free (Thermo Fisher Scientific, cat. no. AM9858)
Proteinase K solution (Thermo Fisher Scientific, cat. no. AM2548)
QIAquick PCR Purification Kit (Qiagen, cat. no. 28106)
10 mM dATP solution (see recipe)
Agilent DNA High-Sensitivity Kit (Agilent Technologies, cat. no. 5067-4626)
Bovine serum albumin (BSA, molecular-biology grade, 20 mg/ml; NEB, cat. no. B9000S)
Nuclease-free water (Thermo Fisher Scientific, cat. no. AM9932)
DNA Polymerase I, large (Klenow) fragment (NEB, cat. no. M0210L)
NEBNext® Quick Ligation Reaction Buffer, 5× (NEB, cat. no. B6058S)
Bridge linker, 200 ng/μl (see Support Protocol)
T4 DNA ligase (NEB, cat. no. M0202L)
Dynabeads Protein G beads for immunoprecipitation (Thermo Fisher Scientific, cat. no. 10009D)
1× PBST buffer (see recipe), cold
Antibody against protein of interest, e.g., monoclonal antibody against RNA Polymerase II (8WG16) (BioLegend, cat. no. 664912)
LiCl buffer (see recipe)
ChIP elution buffer (see recipe)
High-salt buffer (see recipe)
Buffer EB (Qiagen, cat. no. 19086)
Qubit® dsDNA HS Assay Kit (Thermo Fisher Scientific, cat. no. Q32854)

PCR primers (see Strategic Planning)
LightCycler[®] 480 SYBR Green I Master (Roche, 0-4707516001)
Nextera XT Index Kit v2 Set A (Illumina, cat. no. FC-131-2001)
Illumina Tagment DNA Enzyme and Buffer Large Kit (Illumina, cat. no. 20034198)
Dynabeads M-280 Streptavidin (Thermo Fisher Scientific, cat. no.11205D)
2× Binding & Wash buffer (see recipe)
iBlock buffer (see recipe)
Sheared genomic DNA mixture (see recipe)
2× SSC/0.5% (w/v) SDS (see recipe)
NEBNext[®] High-Fidelity 2× PCR Master Mix (NEB, cat. no. M0541S)
AmPure XP beads (60 ml; Beckman, cat. no. A63881)
80% ethanol
BluePippin Cassette Kit (Sage Science, cat. no. BDF2010)

75 cm² (T-75) cell culture flasks (Thermo Fisher Scientific, cat. no. 430641U)
Falcon 50-ml conical centrifuge tubes (Fisher Scientific, cat. no. 14-959-49A)
Centrifuge (Eppendorf 5810R, cat. no. 22628180)
Centrifuge (Eppendorf 5424R, cat. no. 5404000332)
DNA LoBind Tubes (1.5 ml, Eppendorf, cat. no. 022431021)
Eppendorf ThermoMixer[®] F1.5 (Eppendorf, cat. no. 5384000020)
500-cm² TC-treated culture dishes (Corning, cat. no. 431110)
Stuart Orbital Shaker (Stuart, cat. no. SSL1)
Cell scraper (Corning, cat. no. 3011)
Falcon 15-ml conical centrifuge tubes (Fisher Scientific, cat. no. 14-959-53A)
RM-2M Intelli-Mixer, Medium (ELMI, cat. no. IMIX-02)
Bench-top incubator (Thermo Scientific Heratherm, cat. no. IGS60 51028063)
Agilent 2100 Bioanalyzer (Agilent Technologies, cat. no. G2940CA)
Vertical Multi-function Rotator (Grant Instruments, cat. no. PTR-35)
DynaMag-2 Magnet (magnetic stand; Thermo Fisher Scientific, cat. no. 12321D)
Sonicator (SONICS, cat. no. VCX 130)
Qubit 2.0 Fluorometer (Thermo Fisher Scientific, cat. no. Q32866)
Qubit[®] Assay tubes (Thermo Fisher Scientific, cat. no. Q32856)
LightCycler[®] 480 Multiwell Plate 384 (Roche, cat. no. 04729749001)
LightCycler[®] 480 Sealing Foil (Roche, cat. no. 04729757001)
LightCycler[®] 480 System (Roche)
0.2-ml PCR tubes PCR machine (BioRad, C1000 Touch Thermal Cycler)
DNA Clean & Concentrator-5 kit (Zymo Research, cat. no. D4014)
BluePippin instrument (Sage Science, BluePippin)
Illumina NovaSeq 6000 high-throughput sequencing machine (Illumina)

Additional reagents and equipment for cell culture techniques including counting viable cells with trypan blue and a hemocytometer and trypsinization of cells (see Current Protocols article: Phelan & May, 2015), and ChIP qPCR (Read, 2017)

Part 1: Dual crosslinking of cells

In this section, we describe the optimized double-crosslinking procedures for both cultured suspension and adhesion cells, to highlight the differences in chromatin sample preparation for different cell types, which can be adapted for primary cells with minor modifications.

Crosslinking procedure for suspension cells

- 1a. Grow cells to log phase (10⁶ cells/ml maximum density) in a T-75 flask. Measure viability before harvesting by staining an aliquot of the cells with trypan blue and

counting cells with a hemocytometer (see Current Protocols article: Phelan & May, 2015). If viability, as measured by exclusion of the blue dye, is >90%, proceed with the next steps. Otherwise, discard the cells and start over with new cells.

- 2a. Freshly prepare the FA-DPBS and EGS-DPBS solution and keep the EGS-DPBS solution in the 37°C incubator to facilitate dissolving of the EGS. Also prepare another bottle of DPBS stored at the 4°C refrigerator for later use.
- 3a. After cell counting (step 1a), transfer cells and medium from the flasks into either 50-ml conical tubes or centrifuge bottles. Spin down the cells by centrifuging 10 min at $1761 \times g$ (3000 rpm in Eppendorf 5810R), room temperature, remove all supernatant, and add DPBS to resuspend the cell pellet to a final concentration of 10^7 /ml. Transfer 10 million cells to one 1.5 ml Eppendorf DNA LoBind tube and record the cell number.

Use low-binding tubes to prevent cell loss.

- 4a. Spin down the cells 10 min at $1761 \times g$ (3000 rpm in Eppendorf 5810R), room temperature, and remove all supernatant.
- 5a. Resuspend the cells in 1 ml of 1% FA-DPBS solution and mix well using a pipette tip. Incubate for 20 min at room temperature with rotation on Intelli-Mixer (F1, 12 rpm).
- 6a. Add 2.5 M glycine to a final concentration of 0.2 M (87 μ l for 1 ml 1% FA-DBPS solution). Incubate for 10 min at room temperature with rotation on Intelli-Mixer (F1, 12 rpm).
- 7a. Centrifuge the samples 10 min at $1223 \times g$ (2500 rpm in Eppendorf 5810R), room temperature. Remove supernatant and resuspend cells in 1 ml DPBS.
- 8a. Centrifuge the samples 10 min at $1223 \times g$ (2500 rpm in Eppendorf 5810R), room temperature. Remove supernatant and resuspend cells in 1 ml of 2 mM pre-warmed EGS-DPBS solution (pre-warmed solution in the 37°C incubator, step 1a). Incubate for 45 min at room temperature with rotation on Intelli-Mixer (F1, 12 rpm).
- 9a. Add 2.5 M glycine to a final concentration of 0.2 M (87 μ l for 1 ml EGS-DPBS solution). Incubate for 10 min at room temperature with rotation on Intelli-Mixer (F1, 12 rpm).
- 10a. Centrifuge the samples 10 min at $1223 \times g$ (2500 rpm in Eppendorf 5810R), room temperature. Remove supernatant and resuspend cells in 1 ml DPBS. Count the cells with an hemocytometer (see Current Protocols article: Phelan & May, 2015) to ensure that no significant loss of cells happened during the crosslinking steps.
- 11a. Centrifuge the samples 10 min at $1223 \times g$ (2500 rpm in Eppendorf 5810R), room temperature. Remove supernatant.
- 12a. Store the cell pellets at -80°C until ready to proceed to Part 2.

The crosslinked cells can be stored for months

Crosslinking procedure for adherent cells

- 1b. Grow on tissue culture dishes rather than in flasks to facilitate cell harvesting, and prepare an additional dish for counting the cells. For example, grow cells to a final cell density of $6-10 \times 10^7$ per 500 cm^2 in a TC-treated culture dish. Do not grow cells to confluency. Trypsinize the additional dish (see Current Protocols article: Phelan & May, 2015), and test for viability by staining with trypan blue and counting cells with a hemocytometer (see Current Protocols article: Phelan & May, 2015). If viability, as measured by exclusion of the blue dye, is >90%, proceed to next steps with the original dishes. Meanwhile, freshly prepare the FA-DPBS and

EGS-DPBS solutions, and keep the EGS-DPBS solution in the 37°C incubator to facilitate dissolving of the EGS in solution (same as described in 2a). Prepare another bottle of DPBS and store it at 4°C.

- 2b. Remove the culture dishes from the incubator and place at room temperature on the bench.
- 3b. Remove the medium and rinse the monolayer twice with ~30 ml of room temperature DPBS.
- 4b. Add 50 ml of room temperature DPBS to each dish, and then add 1.429 ml of 36% formaldehyde (FA) to each dish. Shake the dishes on a Stuart Orbital Shaker for 20 min at room temperature at 90 rpm.
- 5b. Add 4.47 ml of 2.5 M glycine (glycine final concentration: 0.2 M) to each dish. Shake the dishes on a Stuart Orbital Shaker for 10 min at room temperature at 90 rpm.
- 6b. Remove the FA crosslinking solution and wash once with ~30 ml room temperature DPBS. Then, as much of the buffer as possible with a pipette.
- 7b. Add 50 ml of 2 mM EGS-DPBS solution (pre-warmed solution in the 37°C incubator; see step 1b) to each dish and shake for 45 min at room temperature at 90 rpm on a Stuart Orbital Shaker.
- 8b. Add 4.47 ml of 2.5 M glycine (glycine final concentration: 0.2 M) to each dish. Shake the dishes on a Stuart Orbital Shaker for 10 min at room temperature at 90 rpm.
- 9b. Remove the crosslinking solution and rinse the cells twice with 30 ml room temperature DPBS.
- 10b. Add 30 ml DPBS and store the dishes in the cold room (4°C) until ready to harvest.
- 11b. Bring dishes to the bench. Discard the 30 ml of DPBS.
- 12b. Add 5 ml of pre-chilled (4°C) DPBS (see step 1b), scrape cells with a large scraper, and, for each dish, label a 50-ml collecting tube with date, passage number, and cell type. Collect cells in the 50-ml tubes. Repeat the scraping step twice with 10 ml DPBS to rinse the dishes, and collect these remaining cells into the same 50-ml tube to avoid cell loss.
- 13b. Centrifuge the tubes 10 min at $782 \times g$ (2000 rpm in Eppendorf 5810R), 4°C.
- 14b. Remove supernatant and store the cell pellet at –80°C until ready to proceed to Part 2.

Part 2: In situ ChIA-PET library construction

This section describes the procedures for *in situ* ChIA-PET library construction (Fig. 2). It takes around 5 days to generate one *in situ* ChIA-PET library, and we have divided the protocol steps by day, for convenience. Usually, it will be more efficient and economical to prepare multiple libraries simultaneously. At each major stop point, quality control (QC) measurements are included to evaluate the quality and quantity of chromatin DNA materials (Fig. 3).

Day 1: Cell lysis, nuclei permeabilization, and in situ restriction digestion

15. Take out a 10 million dual-crosslinked cell pellet tube from the –80°C freezer (step 12a or 14b), leave it on ice for 20 min to thaw the cells, centrifuge the tube 5 min at $2500 \times g$, 4°C, and discard the DPBS buffer.

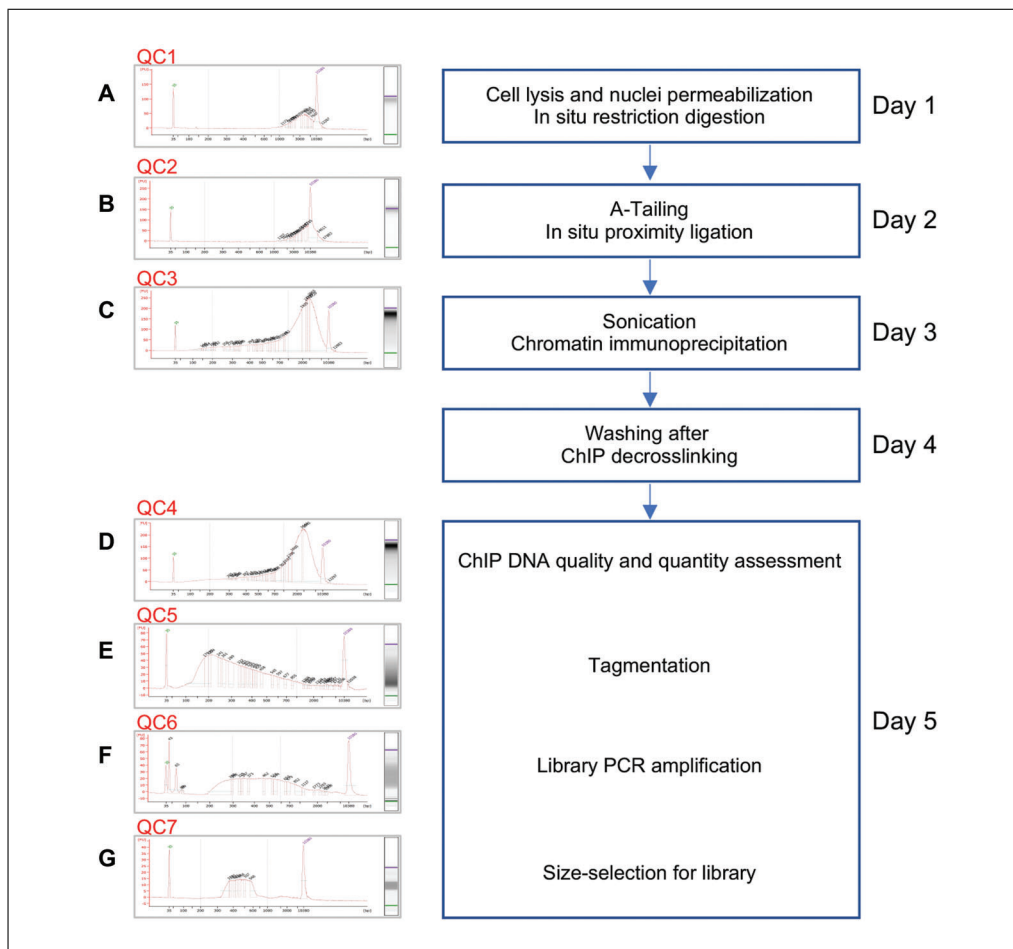


Figure 3 Quality control steps for Basic Protocol 1. Critical quality controls for each step are evaluated via Bioanalyzer. **(A)** QC #1. DNA size distribution profiles of the chromatin DNA after *in situ* *AluI* digestion. The distribution usually peaks at around 5-8 kb, as determined by Bioanalyzer profiling. **(B)** QC #2. DNA size distribution profiles after proximity ligation of the *AluI*-digested chromatin sample. It is expected that after ligation the size of chromatin DNA fragments would increase accordingly. Indeed, as shown in this sample data, the DNA size observed in QC #2 is approximately 2-5 kb larger than the DNA fragments in QC #1, indicating the proximity ligation was successful. **(C)** QC #3. DNA size distribution profiles of sonication-sheared chromatin DNA fragments, showing an obvious shift of peak from > 10 kb to short chromatin fragments in the range of 2-3 kb. **(D)** QC #4. After ChIP enrichment, the enriched chromatin DNA is analyzed by Bioanalyzer for quality and quantity. The profile looks the same as in QC #3, as expected. **(E)** QC #5. DNA size distribution profiles of tagged DNA fragments after tagmentation by the Tn5 transposase step, showing the desired shifting of DNA fragments to the size range of 200-700 bp. **(F)** QC #6. DNA size distribution profiles of PCR-amplified DNA fragments, showing that the majority of amplicons are in the 300-700 bp range. **(G)** QC #7. DNA size distribution profiles of the final *in situ* ChIA-PET library after size selection for the 350-600 bp range, which is the ideal size range for Illumina paired-end-tag sequencing analysis.

16. Prepare 10 ml of 0.1% SDS cell lysis buffer and 10 ml of 0.55% SDS solution in two separate 15-ml tubes. Add one cComplete™, Mini, EDTA-free Protease Inhibitor (PI) Cocktail tablet to each of the tubes, and dissolve completely. Leave the 0.1% SDS cell lysis buffer tube on ice and the 0.55% SDS solution tube at room temperature.
17. Resuspend the cell pellet (step 15) with 1 ml of the 0.1% SDS cell lysis buffer (plus protease inhibitors) on ice. Incubate the tube for 1 hr at 4°C with rotation.
18. Centrifuge the tube 5 min at 2500 × g, 4°C, and then carefully discard the supernatant.

19. Resuspend the cell pellet with 100 μ l of the 0.55% SDS solution (plus protease inhibitors).
20. Incubate the tube at room temperature for 10 min and then at 62°C in a ThermoMixer for 10 min, followed by an incubation at 37°C for 10 min.

The purpose of this step is to permeabilize the nuclear membranes, which will facilitate the entry of enzymes and reagents into the nuclear space for the following in situ digestion and ligation steps. The incubation temperature and duration can be modified according to the cell type used. These suggested parameters have been tested for Drosophila cells (S2 and KC167 cell) and a number of human cell lines for the ENCODE phase 4 Project (e.g., GM12878, MCF7, among others).

21. Add 270 μ l of double-distilled water (ddH₂O) and 50 μ l of 10% Triton X-100 solution to the tube and mix well. Incubate the tube at 37°C for 15 min to quench SDS.
22. Add 50 μ l of 10 \times NEB CutSmart buffer and 30 μ l of *AluI* restriction enzyme.

The tube will now have a final volume of 500 μ l.

23. Incubate the tube overnight with shaking on an Intelli-Mixer (UU, 35 rpm mode) in a 37°C bench-top incubator.

Day 2: A-tailing and proximity ligation

24. Perform the first quality control test (QC #1):
 - a. After the overnight *AluI* digestion (step 23), transfer 10 μ l of the digestion sample to a new tube.
 - b. To that, add 90 μ l of TE buffer, pH 8.0, and 5 μ l of proteinase K solution, mix, and incubate for 1 hr at 65°C.
 - c. Purify the DNA with QIAquick PCR Purification Kit (following the manufacturer's instructions) and check the DNA profile with a Bioanalyzer 2100 HS DNA chip (see QC1, Fig. 3A).
 - d. Proceed to the next step only if suitable digestion fragments (range around 5-8 kb) are visualized in the Bioanalyzer DNA profile.

25. Set up an A-tailing reaction on ice. Add the reagents in the order shown below:

490 μ l restriction-digested sample (from step 23)
11 μ l BSA (20 mg/ml)
11 μ l 10 mM dATP
4 μ l 10 \times CutSmart buffer
3 μ l nuclease-free water
11 μ l Klenow large fragment (3'→5' exo-)
Total: 530 μ l.

Incubate the tube for 1 hr, with rotation, at 37°C.

26. Leave the sample tube at room temperature. Add proximity ligation reagents in the order shown below:

530 μ l A-tailed sample (from step 25)
257 μ l nuclease-free water
200 μ l NEBNext 5 \times Quick Ligation Buffer
3 μ l bridge linker (200 ng/ μ l, see Support Protocol)
10 μ l T4 DNA ligase
Total: 1000 μ l.

Mix well and incubate for 1 hr at room temperature. Transfer the tube to the Intelli-Mixer for overnight ligation (F1 rotation mode, 12 rpm) in a 16°C incubator.

Thaw the pre-prepared bridge linker solution gently on ice, and mix well gently. Be sure to add bridge linker and ligase last. Additional details for the design of bridge linker nucleotide sequences and method for linker preparation are described in the Support Protocol.

27. In the meantime, to coat the antibody with Dynabeads protein G beads, thoroughly resuspend the protein G beads in the bottle and aliquot 100 μ l of protein G beads into a 1.5-ml microcentrifuge tube for the chromatin immunoprecipitation (ChIP) step.
28. Wash the beads three times, each time with 500 μ l of cold PBST buffer as follows. Place the microcentrifuge tube on the magnet rack, wait for 1-2 min, and remove the supernatant. Resuspend the beads in 500 μ l of cold PBST buffer by vortexing briefly, and repeat washing twice. Finally, re-suspend the beads in 500 μ l of cold PBST buffer.
29. Add 20 μ g of antibody to the washed protein G beads from step 28, transfer the tube to the Intelli-Mixer, and incubate the tube at 4°C overnight, with rotation at 12 rpm.

Prepare the antibody bound to protein G beads ~ 6-24 hr before the ChIP experiment.

Day 3: Sonication and ChIP

30. *Perform QC step #2:* After the overnight chromatin proximity ligation step (step 26), transfer 10 μ l from the sample to a new tube and add 90 μ l of TE buffer, pH 8.0. Then, add 5 μ l of proteinase K solution, mix, and incubate for 1 hr at 65°C. Purify the chromatin DNA with the QIAquick PCR Purification Kit (following the manufacturer's instructions) and check the DNA profile with a Bioanalyzer 2100 HS DNA chip (see QC2, Fig. 3B). Proceed to the next step only if suitable ligation fragments are visualized in the DNA profile.

The fragment length of ligated DNA should be approximately 2-5 kb larger than the digested DNA fragments in QC #1 (step 24), which is a good indication of a successful ligation.

31. Spin the sample from step 26 10 min at 5500 \times g, 4°C, remove the supernatant, and gently resuspend the cell pellet in 0.5 ml of 0.1% SDS cell lysis buffer (+PI), for sonication.
32. Shear the chromatin via sonication on ice.

The optimal sonication parameters for each individual cell type should be empirically determined (see Strategic Planning).

33. Centrifuge the sonicated sample 10 min at 6500 \times g, 4°C. Transfer the supernatant to a new tube.
34. *Perform QC step #3:* Transfer 5 μ l of the sample from step 33 to a new tube and add 90 μ l of TE buffer, pH 8.0. Then, add 5 μ l proteinase K, mix, and incubate at 65°C for 1 hr. Purify the chromatin DNA with the QIAquick PCR Purification Kit (following the manufacturer's instructions) and check the DNA profile with a Bioanalyzer 2100 HS DNA chip (see QC3, Fig. 3C). Proceed to the next step only if ideal sonication fragment sizes (2-3 kb) are visualized in the DNA profile.
35. Take 100 μ l of protein G beads per sample, transfer the beads to a microcentrifuge tube on the magnet rack, wait for 1-2 min, and remove supernatant. Then add 500 μ l of cold PBST and vortex briefly to mix well. Place the microcentrifuge tube on the magnet rack again, wait for 1-2 min, and remove supernatant. Add 500 μ l of cold PBST buffer and repeat the washing step twice.

36. Transfer the rest of the supernatant from step 33 to the washed Dynabeads Protein G beads from step 35 and incubate at 4°C for at least 1 hr with rotation on Intelli-Mixer at 12 rpm.

This step is for pre-clearing chromatin to reduce nonspecific binding of chromatin materials to the magnetic beads.

37. Take out the tube of antibody-coated beads from step 29, place the microcentrifuge tubes on the magnet rack, wait for 1-2 min, and remove the supernatant. Add 500 μ l of cold PBST and vortex briefly to mix well. Place the microcentrifuge tubes on the magnet rack again, wait for 1-2 min, and remove the supernatant. Add 500 μ l of cold PBST buffer and repeat the washing step twice.
38. Incubate the antibody-coated beads (from step 37) with the pre-cleared chromatin materials (from step 36) overnight with rotation at 4°C.

Day 4: Wash chromatin DNA on beads after ChIP and reverse crosslinking

39. Place the tube from step 38 in the magnetic stand and remove supernatant.
40. Wash the beads three times with 1 ml of 0.1% SDS cell lysis buffer. For each wash, nutate for 5 min Multi-function Rotator at 4°C and centrifuge 1 min at $60 \times g$ (800 rpm in Eppendorf 5424R), 4°C.
41. Wash the beads twice with 1 ml of high-salt buffer. For each wash, nutate for 5 min on Multi-function Rotator at 4°C and spin at $60 \times g$ (800 rpm in Eppendorf 5424R) for 1 min at 4°C.
42. Wash the beads once with 1 ml of LiCl buffer. Nutate for 5 min on Multi-function Rotator at 4°C and centrifuge 1 min at $60 \times g$ (800 rpm in Eppendorf 5424R), 4°C.
43. Wash the beads twice with 1 ml of TE buffer, pH 8. For each wash, nutate for 5 min on Multi-function Rotator at 4°C and centrifuge 1 min at $60 \times g$ (800 rpm in Eppendorf 5424R), 4°C.
44. Elute DNA from the beads. For this, add 200 μ l of ChIP elution buffer and incubate the tube in the ThermoMixer at 65°C for 30 min with agitation at 900 rpm.
45. Place the microcentrifuge tube (from step 44) on the magnet rack, wait for 1-2 min, and transfer the supernatant (200 μ l) to a new tube (supernatant tube).
46. Remove the tube with the beads from the magnetic rack and add 100 μ l of buffer EB to re-suspend the beads. Mix well. Place the microcentrifuge tube back on the magnet rack, wait for 1-2 min, and transfer the supernatant (100 μ l of EB buffer) to the same “supernatant tube” as in step 45, which has the 200 μ l of supernatant solution, for a final volume of \sim 300 μ l.
47. Add 10 μ l of proteinase K (20 mg/ml) to the mixed solution and proceed to DNA de-crosslinking by incubating the tube in the ThermoMixer at 65°C overnight.

Day 5: ChIP DNA quantity and quality assessment, Tn5 tagmentation, biotin enrichment, and Illumina library preparation

48. Purify the DNA from step 47 using the QIAquick PCR purification kit (following manufacturer’s instructions) and elute DNA twice with 11 μ l of buffer EB (final volume \sim 20 μ l). Take 1 μ l of the DNA to measure DNA concentration with Qubit (following the manufacturer’s instructions) and calculate the yield of ChIP-DNA.

Routinely, we obtain ChIP-DNA in the range of 50-100 ng, depending on the nature of the protein factor and the robustness of the antibody used. In the case of RNAPII ChIA-PET, the yield of ChIP-DNA is usually >50 ng, whereas for CTCF ChIA-PET, the yield of ChIP-DNA is >100 ng.

Stopping point: ChIP DNA can be stored at -20°C for 2 weeks.

Table 1 Example of qPCR Results from QC #4^a

	qPCR Samples	Replicate 1	Replicate 2	Replicate Average	Δ Ct	Final $\Delta\Delta$ Ct
TestPrimer Set1	Input	23.76	23.53	23.645		
TestPrimer Set1	ChIP	18.53	18.75	18.64	-5.005	-7.645
TestPrimer Set2	Input	23.11	23.04	23.075		
TestPrimer Set2	ChIP	18.74	18.53	18.635	-4.44	-7.08
Negative Primer Set	Input	23.26	23.55	23.405		
Negative Primer Set	ChIP	26.16	25.93	26.045	2.64	

^aNumbers in Replicates are Ct values of PCR results for each test primer set in ChIP-DNA or input samples. Δ Ct: differential Ct value between two Ct numbers (ChIP vs. input or Test primer vs. negative primer). $\Delta\Delta$ Ct: the final differential Ct values for the fold of ChIP enrichment as determined by ChIP-qPCR analysis after calibration with negative primer. According to our experience, a final $\Delta\Delta$ Ct above 5 indicates good target enrichment during the ChIP process, directly reflected by sharp binding peaks of the protein in the genome.

49. *Perform QC #4:* First, check the ChIP DNA profile of the sample from step 48 with a Bioanalyzer 2100 HS DNA chip (QC4, Fig. 3D). The profile of the ChIP DNA should look similar to that of the DNA before ChIP enrichment (QC3).

Second, perform qPCR to check for ChIP enrichment of a selected specific DNA region (binding region of the protein of interest) in ChIP-DNA. To do so, two pairs of PCR Test Primers for two known target binding sites and one Negative Primer pairs based on a non-target locus are used for ChIP-qPCR. Take 0.1 ng of the ChIP DNA for each of the 3 PCR primer pairs to set up the qPCR reactions (Read, 2017) with at least 2 replicates, and use the same amount of non-enriched DNA from step 34 (QC3) as total input control, for a total of 12 qPCR reactions (2 template samples [ChIP-DNA and input DNA] \times 3 primer pairs [two for 2 Test primers + one for negative primers] \times 2 technical replicates = 12).

After the qPCR reactions are completed, calculate the fold ChIP-enrichment based on qPCR cycle threshold value (Ct) as follows:

- Calculate the average Ct for each primer pair with 2 replicates.
- Calculate the average differential Ct (Δ Ct) for each primer pair using the formula below:

$$\Delta \text{ Ct} = \text{ChIP} - \text{DNA replicate average Ct} - \text{Input replicate average Ct}$$

- Calculate the Final $\Delta\Delta$ Ct for each Test Primer ChIP-DNA using formula:

$$\text{Final } \Delta\Delta \text{ Ct Primer Set} = \Delta\text{Ct Test; Primer} - \Delta\text{Ct Negative Primer}$$

A sample with a final $\Delta\Delta$ Ct value > 5 is considered a successful ChIP enrichment for protein factors of interest.

Sample data can be found in Table 1.

50. *Proceed with tagmentation with ChIP-DNA:* Add the following reagents to a new 0.2 ml PCR tube, and mix well with a pipette after each addition:

- ___ μ l ChIP DNA (add 50 ng)
- 25 μ l 2 \times NEXTERA Tagmentation Buffer
- ___ μ l NEXTERA Transposase enzyme (TDE)
- ___ μ l nuclease-free water to adjust the final volume
- 50 μ l total volume.

Add enzyme last and mix well. Adjust the amount of Tn5 transposase enzyme according to the amount of ChIP DNA. In this step, all ChIP-DNA will be used for tagmentation. Usually, a maximum of 50 ng ChIP-DNA is used for each reaction, and 8.5 μ l of enzyme is added. The ratio of enzyme and DNA is around 0.17. If the

amount of ChIP DNA is less than 50 ng, adjust the enzyme amount to preserve the 0.17 ratio.

51. Centrifuge briefly to bring solution to the bottom of the tube, incubate the PCR tube at 55°C for 5 min and then at 10°C for 10 min in thermal cycler.
52. Purify tagmented ChIP DNA from step 51 by using the DNA Clean & Concentrator-5 kit (following the manufacturer's instructions) and elute the DNA with 10 µl of EB buffer.

Stopping point: DNA can be stored at -20°C for 2 weeks.

53. *Perform QC #5:* Measure DNA concentration with Qubit and check the DNA profile of the sample from step 52 with a Bioanalyzer 2100 HS DNA chip (QC5, Fig. 3E).

The expected size range of tagmented DNA fragments should be larger than 150 bp and less than 700 bp. Otherwise, the ratio of enzyme to DNA should be empirically adjusted.

In the event of ChIP-DNA yield (step 48) more than 50 ng, the remaining ChIP-DNA can also be processed for additional tagmentation reactions according to the abovementioned ratio, by repeating steps 50 to 52. Combine all purified tagmentation ChIP-DNA together in preparation for the biotin-DNA enrichment step.

54. Equilibrate the M-280 streptavidin Dynabeads to room temperature for 30 min. Then, fully resuspend and transfer 30 µl of the suspended Dynabeads into a new 1.5-ml tube.
55. Place the tube with the beads on the magnetic stand and wait for 1-2 min until the solution is clear. Discard the supernatant and wash twice with 150 µl 2× Binding & Wash buffer.
56. Resuspend beads in 100 µl iBlock buffer, mix, and incubate at room temperature for 45 min on a rotating Intelli-Mixer (UU, 50 rpm).
57. Microcentrifuge briefly to bring the solution to the bottom of the tube, place it on the magnet rack, discard the iBlock buffer, and then wash the beads twice with 200 µl of 1× Binding & Wash buffer.
58. Discard the washing buffer, then add 100 µl of sheared genomic DNA mixture (consisting of 500 ng of genomic DNA in 50 µl nuclease-free water + 50 µl 2× Binding & Washing buffer; see Reagents and Solutions) to the beads. Mix well and incubate on the Intelli-Mixer with rotation for 30 min (UU, 50 rpm) at room temperature.

The purpose of this step is to use total genomic DNA fragments to block nonspecific affinity to M-280 beads, thereby increasing the efficiency of specific streptavidin-biotin binding to enrich the tagmented DNA fragments containing the bridge linker with biotin and reducing the nonspecific DNA (non-biotin DNA) binding to the M-280 beads.

59. Place the tube with the beads and genomic DNA on the magnetic stand and wait for 1-2 min until the solution is clear. Discard the solution with the blocking DNA mixture, taking care not to lose the beads. Add 200 µl of 1× Binding & Wash buffer to the beads and mix well. Discard supernatant and repeat the washing step with 200 µl of 1× Binding & Wash buffer.
60. Pool together all of the tagmented ChIP DNA products from step 52 and bring the volume to 50 µl with nuclease-free water. Then, add an equal volume of 2× Binding & Wash buffer, mix well with the beads from step 59, and incubate at room temperature for 45 min using the Intelli-Mixer (UU, 50 rpm).

61. Microcentrifuge briefly to bring the solution to the bottom of the tube, place tube on magnetic stand, and wait for 1-2 min until the solution are clear. Discard the supernatant, add 500 μ l of 0.5% SDS/2 \times SSC buffer, and mix well. Place the tube with the beads on the magnetic stand and wait for 1-2 min until the solution is clear. Discard the supernatant and repeat the washing of the beads four times as described above, each time with 500 μ l 0.5% SDS/2 \times SSC buffer.
62. Microcentrifuge briefly to bring the solution to the bottom of the tube, place tube on magnetic stand, and wait for 1-2 min until the solution is clear. Use a pipette tip to carefully remove and discard the supernatant. Wash the beads by adding 500 μ l of 1 \times Binding & Wash buffer, resuspending the beads, placing the tube with the beads on the magnetic stand, waiting for 1-2 min until the solution is clear, and then carefully removing and discarding the supernatant. Repeat this washing with 500 μ l of 1 \times Binding & Washing buffer once more. Resuspend the washed beads with the immobilized DNA in 30 μ l of buffer EB.
63. Prepare an Illumina library PCR reaction as follows:
 - 10 μ l DNA-coated beads (from step 62)
 - 5 μ l nuclease-free water
 - 25 μ l NEBNext[®] High-Fidelity 2 \times PCR Master Mix
 - 5 μ l Index Primer 1 (i5, from Nextera XT Index Kit)
 - 5 μ l Index Primer 2 (i7, from Nextera XT Index Kit)
 - Total: 50 μ l.
64. Set up the following PCR run:
 - Initial denaturation at 72 $^{\circ}$ C for 3 min, then at 98 $^{\circ}$ C 30 s
 - 12 cycles of 10 s 98 $^{\circ}$ C \rightarrow 30 s 63 $^{\circ}$ C \rightarrow 40 s 72 $^{\circ}$ C
 - Final extension at 72 $^{\circ}$ C for 5 min
 - Hold at 4 $^{\circ}$ C.

In this step, tagmented ChIP DNA fragments immobilized on beads are directly amplified with the specific primer pair i5 and i7 (Nextera XT Index Kit v2 Set A, see Materials) by PCR into a library that is compatible with Illumina sequencing. Note that only 1/3 of the tagmented ChIP DNA from step 51 is used for PCR here.

Here Nextera transposase enzyme was used for library preparation. In this way, we have to do the extension at 72 $^{\circ}$ C first to fill in the 9-bp DNA gap caused by the enzyme tagmentation step before the denaturation step at 98 $^{\circ}$ C.
65. While the PCR is running, equilibrate AMPure XP beads to room temperature for 30 min before using.
66. After the PCR reaction done, transfer the 50 μ l of the PCR product to a new tube and place it in the magnetic rack. Fully resuspend the AMPure XP beads by vortexing. Add an equal volume (1:1 ratio) of AMPure beads to the reaction tube in the rack (containing the PCR amplicons) and mix well by pipetting up and down \sim 10 times. Incubate the mixture at room temperature for 5 min using the Intelli-Mixer (F8, 30 rpm mode).
67. Microcentrifuge briefly to bring all suspension to the bottom of the tube, put the tube on the magnetic stand, and wait for the solution to clear (\sim 3-5 min). Discard the supernatant, and then add 200 μ l of freshly prepared 80% ethanol to resuspend the beads. Spin down the tube briefly, put the tube back on the magnetic stand, and wait for the solution to clear (\sim 1-2 min). Remove ethanol and repeat the washing step with 200 μ l of 80% ethanol. Let the tube sit on the magnetic stand for 1 min to allow any disturbed beads to settle, and carefully remove any residual ethanol. Leave the tubes open on the magnetic stand to air dry on the bench top for 10 min.

68. Add 10 μ l buffer EB to the tube, vortex, and mix well. Spin down the tube briefly, put the tube on the magnetic stand, and wait for the solution to clear (\sim 3-5 min). Transfer the clear solution to a new 1.5-ml microcentrifuge tube.
69. Measure the concentration of the PCR product from step 68 with Qubit (following the manufacturer's instructions) and check the DNA profile with a Bioanalyzer 2100 HS DNA chip (see QC6, Fig. 3F).
70. Based on the QC6 results (DNA yield and size profile of PCR product), adjust the number of PCR cycles (12) used in step 64 to as low as 9 cycles or up to 13 cycles, and repeat steps 64-69 to amplify the remaining 20 μ l of DNA beads. Pool the purified PCR products together in final volume of 30 μ l for at least 100 ng of amplicons before library size selection.

Stopping point: DNA can be stored at -20°C for 2 weeks.

The purpose of this step is to prepare a library with enough amplified ChIP DNA for library sequencing analysis but also avoid over-amplification of ChIP DNA that could lead to high redundancy of DNA fragments in a library. In general, lower cycle numbers (9-10) usually result in high library complexity, but higher cycle numbers (above 14) would result in a low-complexity library and high redundancy in sequencing reads.

71. Add 10 μ l of V1 maker (from BluePippin Cassette Kit) to the library and use BluePippin for DNA size-selection in the range of 300 bp to 600 bp using a 2% agarose gel cassette.

Refer to the BluePippin's manual for detailed operation instructions.

Usually, the yield of the final size-selected DNA is around 40-60 ng from the 100 ng input DNA (step 68). An alternate approach for DNA size selection for short DNA fragments (300-600 bp), if BluePippin is not available, is to perform AMPure XP beads purification twice for the PCR product.

72. Perform the final QC step #7: Measure the DNA concentration of the final library using Qubit and check the DNA size profile with a Bioanalyzer 2100 HS DNA chip (see QC, Fig. 3G).

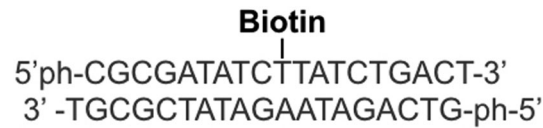
*At this point, the construction of an *in situ* ChIA-PET library is complete, and the library is ready for high-throughput sequencing.*

73. Perform paired-end sequencing of an *in situ* ChIA-PET library with 2×150 bp module using an Illumina NovaSeq 6000 sequencer according to the manufacturer's instructions.

*Most of the DNA templates (300-600 bp) for sequencing within the library are composed of a bridge linker in the middle and two short DNA fragments, one on each side, derived from interacting chromatin loci. Paired-end sequencing is done preferably using a 2×150 bp module with the 300-cycle sequencing kit to read the paired-end tags with the bridge linker, to capture chromatin interactions. The sequencing reads are processed by ChIA-PIPE, a fully automated pipeline for comprehensive ChIA-PET data analysis and visualization (Lee et al., 2020; see Basic Protocol 2). For each library, we routinely perform a pilot small-scale sequencing of about 1-10 million reads as a final confirmation of the quality of the *in situ* ChIA-PET library. The initial pilot sequencing data reveals if most of the reads possess the correct DNA template composition (tag-linker-tag), the degree of library complexity, and whether mapping data for ChIP-enrichment (peaks) and chromatin interactions (loops) are detected at expected genomic loci. Once high quality of the library is confirmed, it can be further sequenced more deeply with the Illumina platform 2×150 bp module for at least 200 million reads, for final data production.*

BRIDGE LINKER PREPARATION

In an *in situ* ChIA-PET experiment, a DNA oligo linker, called the bridge linker, is used to connect the two interacting chromatin fragments during proximity ligation. The bridge linker is composed of a double-stranded DNA oligonucleotide (18 bp) with a “T” overhang at the 3’ end and phosphorylation at the 5’ end of both strands, and a modified “T” nucleotide with a biotin group in the middle of the top strand.



During proximity ligation, each bridge linker will be ligated to two chromatin fragments, each with an “A” overhang added via the A-tailing reaction after the restriction digestion during chromatin sample preparation. Therefore, during proximity ligation, only the bridge linker and chromatin fragments will be compatible for ligation, whereas self-ligation of bridge linkers and the ligation between chromatin fragments will be prevented. The bridge linkers should be prepared beforehand and can be stored at -20°C for several months.

Materials

DNA oligonucleotides (ordered from IDT or similar, HPLC-purified):

Top oligo: 5’-/5Phos/CGA TAT C/iBIODT/T ATC TGA CT-3’

Bottom oligo: 5’-/5Phos/GT CAG ATA AGA TAT CGC GT-3’

TE buffer, pH 8.0 (Thermo Fisher Scientific, cat. no. AM9849)

Novex™ TBE Gels, 4%-20%, 10 well (Thermo Fisher Scientific, cat. no. EC6225BOX)

25-bp DNA ladder (Thermo Fisher Scientific, cat. no. 10597-011)

TBE buffer (10×), RNase-free (Thermo Fisher Scientific, cat. no. AM9865)

0.2-ml RNase-free PCR tubes (Thermo Fisher Scientific, cat. no. AM12225)

PCR machine (BioRad, C1000 Touch Thermal cycler)

NanoDrop™ 8000 Spectrophotometer (Thermo Scientific)

SureLock™ Tandem Midi Gel Tank (Thermo Fisher Scientific, cat. no. STM1001)

Additional reagents and equipment for SDS-PAGE (see Current Protocols article: Gallagher, 2012)

1. Order the oligos from IDT (or similar) at 250-nmol scale, HPLC purified in desalted form.
2. Add TE buffer to dissolve the oligos to a final concentration of 100 μM .
3. Vortex to mix well, then place the tubes at 4°C overnight (without vortexing) to allow oligos to resuspend completely.
4. Prepare five different ratios of top oligo:bottom oligo (1:1, 1.5:1, 2:1, 1:1.5, 1:2) to test for the best annealing efficiency. Ultimately, the selected ratio will be used for large scale annealing of bridge linkers.

Example: for 1.5:1, mix together 7.5 μl of top oligo (100 μM) and 5 μl of bottom oligo (100 μM).

5. Run on a PCR machine using the following program:

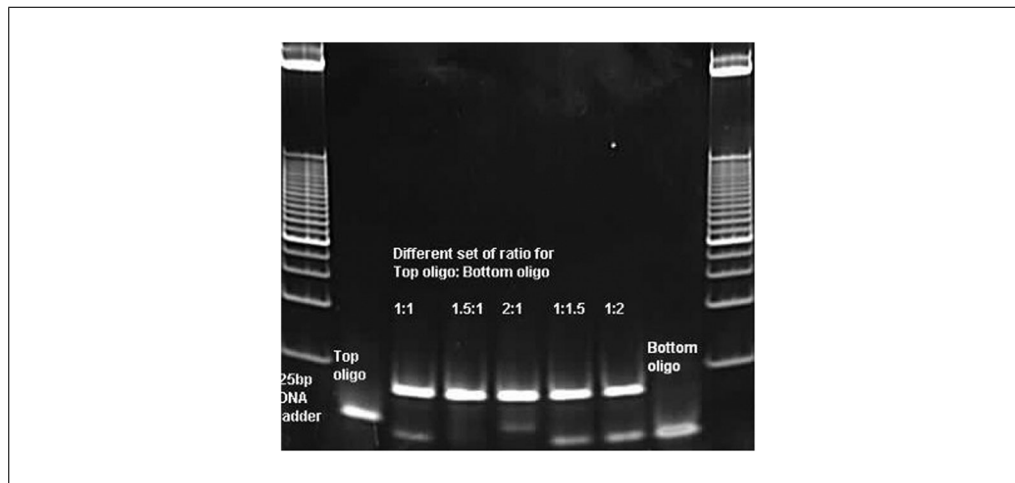


Figure 4 Bridge linker preparation described in the Support Protocol. A gel picture of electrophoresis for the bridge linker DNA oligos. Sample lanes 2 and 8 (by the ladders) are for the single-strand top and bottom oligo, respectively. Lanes 3-7 are for the annealed bridge linkers under five different annealing ratio of top-over-bottom oligos, 1:1, 1.5:1, 2:1, 1:1.5, and 1:2, respectively. Lanes 1 and 9 show the molecular markers of a 25-bp ladder. As shown in this example, the ratio of top oligo:bottom oligo at 1.5:1 has the best result to get the maximum double stranded linkers. All other ratios showed remaining single-strand oligos that were not annealed into double-stranded linkers.

95°C	2 min
Ramp 95°C to 75°C (rate: 0.1°C/s)	Hold at 75°C 2 min
Ramp 75°C to 65°C (rate: 0.1°C/s)	Hold at 65°C 2 min
Ramp 65°C to 50°C (Rate of 0.1°C/s)	Hold at 50°C 2 min
Ramp 50°C to 37°C (rate of 0.1°C/s)	Hold at 37°C 2 min
Ramp 37°C to 20°C (rate of 0.1°C/s)	Hold at 20°C 2 min
Ramp 20°C to 4°C (rate of 0.1°C/s)	Hold at 4°C Indefinitely until collection

6. Measure the concentration of the annealed bridge linkers using Nanodrop™ 8000 Spectrophotometer.
7. Dilute annealed bridge linkers to 200 ng/μl with TE buffer.

Important! Keep all annealed bridge linkers cold on ice.
8. Run 200 ng of each single-stranded oligos and 200 ng of the annealed bridge linkers from each of the five different annealing conditions on the same 4%-20% TBE gel (see Current Protocols article: Gallagher, 2012).

Usually, the best ratio of top oligo versus bottom oligo is the one that shows the maximal amount of annealed double-strand linker with the minimal top or bottom oligo left on the gel. For example, in Figure 4, the best ratio of top versus bottom oligo is 1.5 to 1.
9. Perform a large-scale annealing reaction with the remaining top and bottom oligos according to the optimal ratio. For this, repeat step 5.
10. Perform Nanodrop quantification and dilute the annealed bridge linker to 200 ng/μl with TE buffer for use in ChIA-PET protocol (Basic Protocol 1). Store at -20°C.

DATA PROCESSING AND VISUALIZATION

The workflow for sample preparation and library generation described in Basic Protocol 1 results in paired-end raw sequencing reads that need to be further processed to gain biological insights. Here, we provide a step-by-step guide to convert raw sequencing data into processed files that can be visualized and analyzed. By running ChIA-PIPE

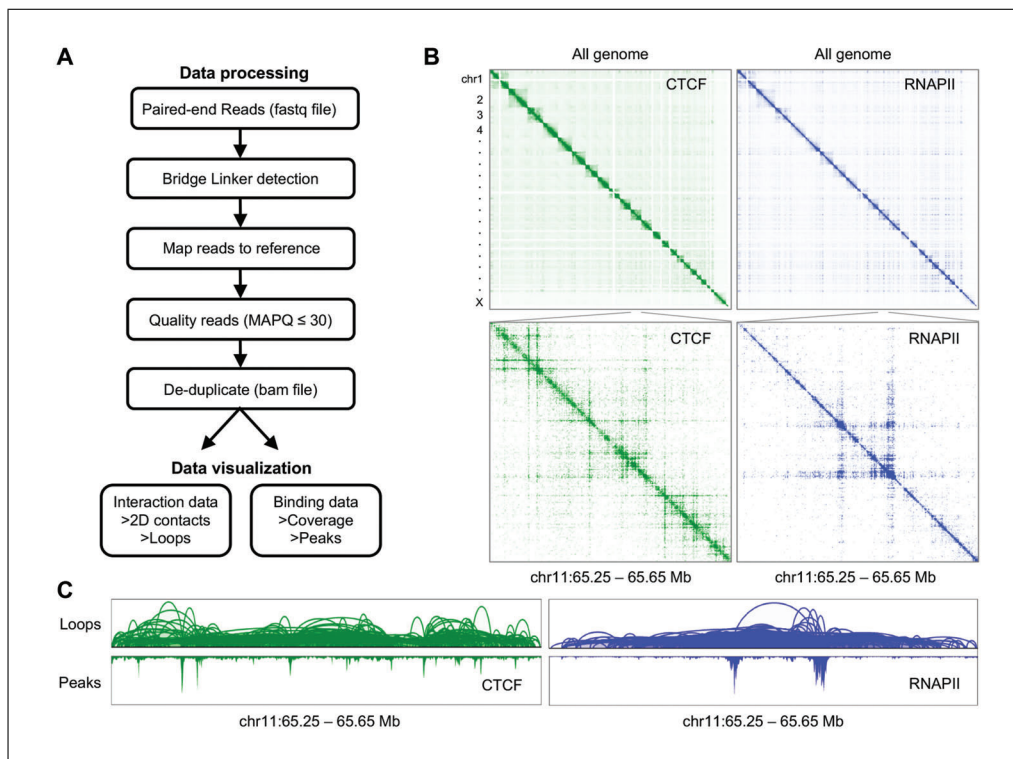


Figure 5 Overview of *in situ* ChIA-PET data processing and visualization. **(A)** Data processing flowchart. **(B)** Juicebox 2D contact map visualization for the whole genome (top) and a zoom-in genomic region in chr 11 (bottom) of the GM10248 cell line for CTCF *in situ* ChIA-PET (green) and RNAPII *in situ* ChIA-PET (blue) data. **(C)** BASIC browser view of *in situ* ChIA-PET loops and binding peak coverage in the region of chr11: 65.25-65.65 Mbp for the same proteins.

(Lee et al., 2020), our fully automated ChIA-PET data processing pipeline, in a high-performance computing environment, the user should expect to obtain loops and protein-binding profiles. The accompanying web browsers BASIC and Juicebox allow users to visualize the output files from ChIA-PIPE. An overview of the process is shown in Figure 5A. Here, we exemplify the protocol using a small test dataset.

Materials

High-performance computing environment, with an ability to allocate up to 60 GB of Random Access Memory (RAM) and 24 hr of runtime.

A desktop computer or a laptop with internet access for data visualization.

ChIA-PIPE package

(<https://github.com/TheJacksonLaboratory/ChIA-PIPE.git>), v1.0

Sequencing results (steps below are shown for a small test dataset, available at Zenodo, doi: 10.5281/zenodo.4706038)

1. Make a test directory in the home directory of the high-performance computing environment.

```
$ mkdir -p testing_chia_pipe
$ cd testing_chia_pipe
```

2. Clone the ChIA-PIPE package (v1.0) from the github repository. If git is installed, type:

```
$ git clone
https://github.com/TheJacksonLaboratory/ChIA-PIPE.git
```

If git is not installed, download ChIA-PIPE directly:

- ```
$ wget
git@github.com:TheJacksonLaboratory/chia_pipe.zip
```
3. Install the dependencies for ChIA-PIPE:

```
$ dep_dir="dep_dir"
$ bash local_install_chia_pipe_dependencies.sh -i
${dep_dir}
```
  4. Download test data from Zenodo: <https://zenodo.org/record/4706038#.YIAx2R0pCHs>.

```
$ mkdir -p fastq
$ cp LDK0004-ds_*.fastq.gz fastq
```
  5. Review the config file in `chia_pipe-master/example_config_file.sh` and ensure that `bin_dir` is specified according to the directory where ChIA-PIPE has been installed.
  6. Launch ChIA-PIPE.

```
$ qsub -F "--conf
chia_pipe-master/example_config_file.sh"
chia_pipe-master/0.chia_pipe_hpc.pbs
```
  7. Wait for the pipeline to finish running. Note that this is expected to take 5-10 hr depending on the user's computing environment. After the run, there should be a `4.LDK0004-ds.extract_summary_stats.o` and a `LDK0004-ds.final_stats.tsv` file in the `/LDK0004-ds/` directory.
  8. Transfer the 7 key processed files in the `/LDK0004-ds/` directory from HPC environment to the local desktop for downstream visualization and analyses.
    - a. `LDK0004-ds.final_stats.tsv` (similar examples with the same data format as provided in Table 2)

*A summary statistics table including the total read pairs, uniquely mapped read pairs, number of peaks, and number of PET clusters (loops).*
    - b. `LDK0004-ds.e500.clusters.cis.gz`

*A list of intra-chromosomal loops in bedpe format, with 7<sup>th</sup> columns denoting the number of PETs contributing to a particular loop (7<sup>th</sup> column also referred to as PET count). In other words, this file can be considered as a table with 7 columns: chrom1, start1, end1, chrom2, start2, end2, PETcount.*
    - c. `LDK0004-ds.e500.clusters.cis.BE3`

*A subset of `LDK0004-ds.e500.clusters.cis.gz` with 7<sup>th</sup> column  $\geq 3$ .*
    - d. `LDK0004-ds.e500.clusters.trans.gz`

*A list of inter-chromosomal loops in bedpe format, with 7<sup>th</sup> column denoting PET count.*
    - e. `LDK0004-ds.for.BROWSER.sorted.bedgraph`

*The protein binding coverage file in a standard 4-column bedgraph format.*
    - f. `LDK0004-ds.no_input_all_peaks.narrowPeak`

*A list of peaks called by MACS2 in a bed format.*
    - g. `ChIA-PET_dm3_Kc167_RNAPII_LDK0004-ds_miseq_pairs.hic`

*A binary file that can be visualized through Juicebox (Durand et al., 2018)*
  9. Visualize intra-chromosomal loops, binding coverage, and peaks.
    - a. Download and install the dockerized version of the BASIC Browser: <https://github.com/TheJacksonLaboratory/basic-browser>.



**Table 2** An Example of Two *in situ* ChIA-PET Library Statistics Outputs<sup>a</sup>

| Raw       | QC matrix                                    | In-house Library ID: LHG0047H | In-house Library ID:LHG0118V | Notes                                                                                                     |
|-----------|----------------------------------------------|-------------------------------|------------------------------|-----------------------------------------------------------------------------------------------------------|
| 1         | Seq method (2 × 150)                         | Hiseq                         | Novaseq                      |                                                                                                           |
| 2         | Reference genome                             | hg38                          | hg38                         |                                                                                                           |
| 3         | Cell type                                    | GM10248                       | GM10248                      |                                                                                                           |
| 4         | Factor                                       | CTCF                          | RNAPII                       |                                                                                                           |
| <b>5</b>  | <b>Total read pair</b>                       | <b>340,840,469</b>            | <b>316,510,593</b>           | Library sequencing depth and quality<br>Higher fraction indicates higher quality of the library           |
| <b>6</b>  | <b>Read pair with bridge linker</b>          | <b>320,834,185</b>            | <b>286,944,905</b>           |                                                                                                           |
| <b>7</b>  | <b>Fraction of read pairs with linker</b>    | <b>0.94</b>                   | <b>0.91</b>                  |                                                                                                           |
| 8         | Quality non-redundant tag                    | 441,205,365                   | 74,642,122                   | Tags used for peak calling<br># of peaks varies depending on factors                                      |
| 9         | Protein factor binding peak                  | 929,221                       | 222,016                      |                                                                                                           |
| 10        | Quality paired-end-tag (PET)                 | 218,869,947                   | 170,109,710                  | These PETs are used for interaction analysis<br>Library complexity                                        |
| 11        | Uniquely mapped PET                          | 162,534,265                   | 127,491,848                  |                                                                                                           |
| <b>12</b> | <b>Non-redundant PET</b>                     | <b>142,662,939</b>            | <b>16,107,717</b>            |                                                                                                           |
| <b>13</b> | <b>PET redundancy</b>                        | <b>0.12</b>                   | <b>0.87</b>                  |                                                                                                           |
| 14        | Self-ligation PET (<8 kb)                    | 27,534,605                    | 10,575,144                   | Chromatin fragment                                                                                        |
| 15        | Inter-ligation PET (> 8 kb)                  | 115,128,334                   | 5,532,573                    | Long-range interaction                                                                                    |
| <b>16</b> | <b>Intra-chr PET</b>                         | <b>83,450,947</b>             | <b>3,997,431</b>             | Most useful data are intra-chromosomal<br>Higher ratio (>1) indicates higher quality of interaction data. |
| 17        | Inter-chr PET                                | 31,677,387                    | 1,535,142                    |                                                                                                           |
| <b>18</b> | <b>Ratio of intra/inter PET</b>              | <b>2.63</b>                   | <b>2.6</b>                   |                                                                                                           |
| 19        | Singleton                                    | 101,710,672                   | 4,135,031                    |                                                                                                           |
| <b>20</b> | <b>Intra-chr singleton</b>                   | <b>70,457,789</b>             | <b>2,909,661</b>             |                                                                                                           |
| 21        | Inter-chr singleton                          | 31,252,883                    | 1,225,370                    |                                                                                                           |
| 22        | PET cluster                                  | 4,453,925                     | 498,026                      |                                                                                                           |
| <b>23</b> | <b>Intra-chr PET cluster (≥2)</b>            | <b>4,245,707</b>              | <b>379,818</b>               | Indicator for good data                                                                                   |
| 24        | Inter-chr PET cluster (≥2)                   | 208,218                       | 118,208                      | Indicator for noise                                                                                       |
| <b>25</b> | <b>Ratio of intra/inter PET cluster (≥2)</b> | <b>20.39</b>                  | <b>3.21</b>                  | Higher ratio means higher quality of the data with more frequently recurrent PET cluster                  |
| <b>26</b> | <b>Ratio of intra/inter PET cluster (≥5)</b> | <b>1914.56</b>                | <b>3.82</b>                  |                                                                                                           |

(Continued)

**Table 2** An Example of Two *in situ* ChIA-PET Library Statistics Outputs<sup>a</sup>, *continued*

| Raw | QC matrix                                                      | In-house Library ID: LHG0047H | In-house Library ID:LHG0118V | Notes |
|-----|----------------------------------------------------------------|-------------------------------|------------------------------|-------|
| 27  | <b>Ratio of intra/inter PET cluster (<math>\geq 10</math>)</b> | <b>11409.57</b>               | <b>9.01</b>                  |       |
| 28  | Intra-chr PET cluster                                          |                               |                              |       |
| 29  | PET number_2                                                   | 3,097,854                     | 263,388                      |       |
| 30  | PET number_3                                                   | 606,301                       | 54,569                       |       |
| 31  | PET number_4                                                   | 214,162                       | 24,180                       |       |
| 32  | PET number_5                                                   | 101,160                       | 13,720                       |       |
| 33  | PET number_6                                                   | 56,565                        | 8343                         |       |
| 34  | PET number_7                                                   | 35,610                        | 5186                         |       |
| 35  | PET number_8                                                   | 23,968                        | 3290                         |       |
| 36  | PET number_9                                                   | 17,347                        | 2180                         |       |
| 37  | PET number_10                                                  | 12,873                        | 1298                         |       |
| 38  | PET number >10                                                 | 79,867                        | 3664                         |       |
| 39  | Inter-chr PET cluster                                          |                               |                              |       |
| 40  | PET number_2                                                   | 201,273                       | 86,018                       |       |
| 41  | PET number_3                                                   | 6234                          | 15,427                       |       |
| 42  | PET number_4                                                   | 540                           | 6887                         |       |
| 43  | PET number_5                                                   | 99                            | 3942                         |       |
| 44  | PET number_6                                                   | 45                            | 2354                         |       |
| 45  | PET number_7                                                   | 11                            | 1483                         |       |
| 46  | PET number_8                                                   | 7                             | 871                          |       |
| 47  | PET number_9                                                   | 2                             | 514                          |       |
| 48  | PET number_10                                                  | 0                             | 308                          |       |
| 49  | PET number >10                                                 | 7                             | 404                          |       |

<sup>a</sup> These two examples show a summarized statistical report for a ChIA-PET run for RNAPII or CTCF in GM10248 cells. The key parameters are in **bold**.

b. Follow the github instructions to upload LDK0004-ds.e500.clusters.cis.BE3 (loops), LDK0004-ds.for.BROWSER.sorted.bedgraph (binding coverage), and LDK0004-ds.no\_input\_all\_peaks.narrowPeak (peaks).

*Alternatively, the data can be visualized via the WashU Epigenome browser: <https://epigenomegateway.wustl.edu>.*

10. Visualize inter- and intra-chromosomal interactions through 2D contact maps.

a. Transfer \*.hic file from the high-performance computing environment to the local drive (e.g., Desktop).

b. On a web browser, visit <https://aidenlab.org/juicebox/>.

c. Click on “Load Map”, “Local File” and locate the ChIA-PET\_dm3\_Kc167\_RNAPII\_LDK0004-ds\_miseq\_pairs.hic file.

*Alternatively, 2D contact maps can be visualized via higlass (<https://higlass.io>) and the 3D genome browser (<http://3dgenome.fsm.northwestern.edu>).*

*A summary of the computational data processing procedure is shown in Figure 5A. A 2D contact map visualization is shown in Figure 5B, and BASIC Browser visualization in*

Figure 5C. Our data processing pipeline generates a statistical output (Table 2), which allows quality assessment of the *in situ* ChIA-PET library data. As an example, Table 2 shows a summary report for a ChIA-PET run for RNAPII and CTCF in GM10248 cells.

## REAGENTS AND SOLUTIONS

### ***Binding & Wash buffer, 1 ×***

5 mM Tris-HCl pH 7.5 (Thermo Fisher Scientific, cat. no. 15567027)

0.5 mM EDTA (Thermo Fisher Scientific, cat. no. 9261)

1 M NaCl

Store at room temperature for several months

### ***Binding & Wash buffer, 2 ×***

5 mM Tris-HCl pH 7.5 (Thermo Fisher Scientific, cat. no. 15567027)

1 mM EDTA (Thermo Fisher Scientific, cat. no. 9261)

1 M NaCl

Store at room temperature for several months

### ***ChIP elution buffer***

50 mM Tris-HCl pH 7.5 (Thermo Fisher Scientific, cat. no. 15567027)

10 mM EDTA (Thermo Fisher Scientific, cat. no. 9261)

1% SDS (Thermo Fisher Scientific, cat. no. AM9822)

Prepare immediately before use

### ***dATP solution, 10 mM***

Add 20  $\mu$ l of dATP (100 mM; NEB, cat. no. N0440S) to 180  $\mu$ l of nuclease-free water (Thermo Fisher Scientific, cat. no. AM9932). Store in  $-20^{\circ}\text{C}$  freezer.

### ***EGS-DPBS solution, 2 mM***

Take EGS [ethylene glycol bis(succinimidyl succinate; Thermo Fisher Scientific, cat. no. 21565)] out of the  $4^{\circ}\text{C}$  refrigerator and equilibrate to room temperature for at least 1 hr. Dissolve 45.63 mg of EGS in 250  $\mu$ l of DMSO with vortexing. Add the EGS-DMSO solution to 50 ml pre-warmed ( $37^{\circ}\text{C}$ ) DPBS. Prepare immediately before use.

### ***FA-DPBS solution, 1%***

Add 1429  $\mu$ l of 36% formaldehyde Sigma-Aldrich, cat. no. 47608-250ML-F) to 50 ml DPBS (Gibco, cat. no. 14190-250). Prepare immediately before use.

### ***Glycine solution, 2.5 M***

Add 27.89 g of glycine (Sigma, cat. no. G7126) to 80 ml of ddH<sub>2</sub>O and mix well until the solution is clear. Bring the final solution volume to 100 ml. Store at room temperature for several months.

### ***High-salt buffer***

5 mM Tris-HCl pH 7.5 (Thermo Fisher Scientific, cat. no. 15567027)

0.5 mM EDTA (Thermo Fisher Scientific, cat. no. 9261)

1 M NaCl

Store at  $4^{\circ}\text{C}$  for several months

### ***iBlock buffer***

Dissolve 2 g of iBlock Protein-Based Blocking Reagent (Thermo Fisher Scientific, cat. no. T2015) in 90 ml of ddH<sub>2</sub>O in a  $65^{\circ}\text{C}$  water bath, add 5 ml of 10% (w/v) SDS (Thermo Fisher Scientific, cat. no. AM9822), and bring volume to 100 ml with ddH<sub>2</sub>O. Stored at room temperature for several months.

### ***LiCl buffer***

10 mM Tris-HCl pH 8.0 (Thermo Fisher Scientific, cat. no. AM9856)  
250 mM LiCl (Sigma-Aldrich, cat. no. L7026)  
1 mM EDTA (Thermo Fisher Scientific, cat. no. 9261)  
0.5% Nonidet P-40 (Sigma-Aldrich, cat. no. 11754599001)  
0.5% sodium deoxycholate  
Store at 4°C for several months

### ***PBST buffer, 1×***

Add 89.9 ml of ddH<sub>2</sub>O to 10 ml of 10× PBS (Life Technologies, cat. no. AM9937) and 100 μl of Tween 20 (molecular biology grade; Sigma-Aldrich, cat. no. P9416), and mix them well. Store at 4°C for several months.

### ***SDS cell lysis buffer, 0.1%***

50 mM HEPES-KOH pH 7.5 (Fisher Scientific, cat. no. BP299-1)  
150 mM NaCl  
1 mM EDTA (Thermo Fisher Scientific, cat. no. 9261)  
1% (w/v) Triton X-100 (Sigma-Aldrich, cat. no. 648464)  
0.1% (w/v) sodium deoxycholate  
0.1% (w/v) sodium dodecyl sulfate (SDS) solution (Thermo Fisher Scientific, cat. no. AM9822)  
Store at 4°C for several months

### ***SDS solution, 0.55%***

Add 55 μl of 10% (w/v) sodium dodecyl sulfate (SDS) solution (Thermo Fisher Scientific, cat. no. AM9822) to 10 ml of nuclease-free water (Thermo Fisher Scientific, cat. no. AM9932). Prepare before use.

### ***Sheared genomic DNA mixture***

The sheared genomic DNA can be prepared from any species. Usually, genomic DNA is sheared to an approximate size range of 200-1000 kb using Covaris or Bioruptor. Measure the sheared DNA concentration with Qubit assay and store it at –20°C for several months. Use 500 ng of sheared DNA for each ChIA-PET reaction.

### ***SSC, 2×/SDS, 0.5% (w/v)***

Combine 85 ml of ddH<sub>2</sub>O, 10 ml of 20× SSC solution (Thermo Fisher Scientific, cat. no. AM9763), and 5 ml of 10% (w/v) SDS solution (Thermo Fisher Scientific, cat. no. AM9822) and mix them well. Store at room temperature for several months.

## **COMMENTARY**

### **Background Information**

The original ChIA-PET method (Fullwood et al., 2009) was developed based on ChIP-PET (Wei et al., 2006), which enabled the genome-wide mapping of transcription factor (TF) occupancy. ChIA-PET added a proximity ligation step to the ChIPed chromatin material, thereby allowing users to observe the connectivity between chromatin fragments tethered together by protein factors. It employs dual linkers in proximity ligation to distinguish nonspecific ligation, and a type IIS restriction enzyme digestion (*MmeI*) to generate a 20-bp tag of chromatin DNA on each side of the dual-linker ligation prod-

ucts. The paired-end-tag (PET, tag-linker-tag) templates are subjected to high-throughput paired-end sequencing and mapping, to define long-range chromatin interactions. To increase PET mapping efficiency, we employed random digestion by Tn5 transposase of the proximity ligation products, to generate longer templates (tagmentation) for 2 × 150 bp PET sequencing (Li et al., 2017). This improved approach, called long-read ChIA-PET, significantly increases mapping efficiency and allows for the detection of allele-specific chromatin interactions at single-base-pair resolution (Li et al., 2017; Tang et al., 2015)

One drawback of the long-read ChIA-PET protocol, however, is that it requires 100 million cells per experiment, thereby limiting its application in the study of primary cells or clinical samples, in which cells of interest are scarce and tissue sources limited. Inspired by the *in situ* Hi-C protocol (Rao et al., 2014), we developed an *in situ* ChIA-PET protocol, in which chromatin materials are first fragmented by *in situ* restriction digestion, followed by *in situ* proximity ligation, chromatin immunoprecipitation (ChIP) against a protein factor of interest, and library preparation using Tn5 transposase. The *in situ* ChIA-PET procedure described in this article allows for a substantial reduction in the required number of cells to as few as 1 million cells per *in situ* ChIA-PET experiment, thereby making this mapping technique widely applicable to precious samples with low numbers of cells of interest. The key update in this new method is the improved efficiency of proximity ligation in the intact nuclei compared to ligation in solution. The *in situ* ChIA-PET data shows significant reduction of nonspecific proximity ligations and increased intra- to inter-chromosomal interaction ratios compared to previous versions of the ChIA-PET protocols (Goh et al., 2012; Li et al., 2017). As a result, the *in situ* ChIA-PET library cost per sample (including library preparation and sequencing costs for 300-400 million reads) is approximately 1000-1500 USD, which is substantially lower than the previous library cost per sample of approximately 2500-3000 USD.

As part of the ENCODE phase 4 project, the *in situ* ChIA-PET data generated from various cell lines and primary cells (e.g., human hematopoietic stem cells and primary immune cells) has provided high-resolution 3D chromatin topology on a global scale and helped to comprehensively map functional elements in the mammalian genome (<https://www.encodeproject.org/>).

Note that two other protocols, HiChIP (Mumbach et al., 2016) and PLAC-seq (Fang et al., 2016), share the same strategy as ChIA-PET for mapping chromatin interactions mediated by protein factors and have procedures for *in situ* digestion/ligation similar to *in situ* ChIA-PET, but exhibit minor technical differences.

## Critical Parameters

### Cell number

*In situ* ChIA-PET is most robust with 10 million cells; it is also feasible, however, with

fewer cells, as few as 1 million, with the trade-off of reduced library complexity and data quality.

### Cell crosslinking conditions

Crosslinking is critical to capture chromatin interactions. However, there is a fine balance: too little formaldehyde (FA) could result in loss of expected chromatin interactions, and too much could lead to high background noise. The optimal concentration of FA for crosslinking should be empirically determined. We have used up to 2% FA with successful results. In addition, EGS, as a protein-protein crosslinker, will strengthen the robustness of chromatin complex mediated by proteins indirectly bound to DNA.

### Nuclei permeabilization

Nuclei permeabilization is critically important for successful *in situ* digestion and ligation, which ensures capturing bona fide chromatin contacts and avoids nonspecific ligation events. Equally important is the ChIP enrichment for specific chromatin interactions mediated by a particular protein factor under study.

### Antibody quality and ChIP

A ChIP-grade antibody of interest is required for *in situ* ChIA-PET. Usually, different assays are recommended to comprehensively characterize and assess the quality of an antibody, such as western blot, immunoprecipitation (IP), and ChIP-seq, with multiple cell lines. Those antibodies that exhibit the highest IP pull-down efficiency and cleanest background should be selected for *in situ* ChIA-PET experiments.

### Quality controls

It is important to implement early QC steps in a ChIA-PET experiment. We have included several QC steps for the wet-lab part of the protocol, and data statistics and quality-assessment metrics after data processing. During the experiment, we specifically check for the efficiency of enzyme digestion, ligation, DNA size distribution after sonication, ChIP, tagmentation, PCR, and library size-selection, to ensure that each step is done correctly and generates the expected results (Fig. 3).

### Troubleshooting

There are a number of problems that are commonly encountered in ChIA-PET experiments, as revealed by QC analysis during library construction and after library sequencing. In Table 3, we list some of these commonly observed problems, their possible

**Table 3** Troubleshooting Guide for In Situ ChIA-PET

| Problem                                                                                                                                | Possible reason                                                                                                                                                                                                                                                                                                                        | Solution                                                                                                                                                                                                                                                                                                                                                      |
|----------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Troubleshooting during sample preparation and data generation (Basic Protocol 1)</b>                                                |                                                                                                                                                                                                                                                                                                                                        |                                                                                                                                                                                                                                                                                                                                                               |
| Large DNA fragments (>9-10 kb) at step 24                                                                                              | This indicates inefficient restriction digestion. Two possible reasons: <ol style="list-style-type: none"> <li>The restriction enzyme did not function well.</li> <li>Incomplete nuclei permeabilization, so that the enzyme did not enter the nuclei for <i>in situ</i> digestion.</li> </ol>                                         | <ol style="list-style-type: none"> <li>Check if the correct enzyme was used and its expiration date.</li> <li>Optimize the conditions for nuclei permeabilization by increasing SDS concentration, temperature (up to 62°C), and incubation time. Check the dynamic changes of nuclear morphology under microscopy to find out optimal conditions.</li> </ol> |
| No significant size shift of DNA fragments (at least 2 kb) after the ligation step (step 30), indicating inefficient ligation reaction | <ol style="list-style-type: none"> <li>Ligase did not work properly.</li> <li>Nuclei permeabilization might be too strong.</li> </ol>                                                                                                                                                                                                  | <ol style="list-style-type: none"> <li>Check and make sure you are using the ligase properly.</li> <li>Try gentler conditions for nuclei permeabilization. Examine the nuclei under the microscope, and make sure that most nuclei are still intact after the permeabilization step.</li> </ol>                                                               |
| Low $\Delta\Delta Ct$ (<5) in step 49, indicating poor ChIP enrichment by the antibody used in ChIA-PET                                | <ol style="list-style-type: none"> <li>Low-quality (not ChIP-grade) antibody.</li> <li>Crosslinking might not be strong enough, so that the target protein and DNA were not crosslinked.</li> <li>Permeabilization conditions were too strong, so that they disrupted the protein-DNA interaction.</li> </ol>                          | <ol style="list-style-type: none"> <li>Use fully characterized antibody (ChIP grade), test ChIP efficiency, and identify the best batch of antibody for further use.</li> <li>Increase the concentration of FA in the crosslinking step.</li> <li>Try gentler conditions for nuclei permeabilization.</li> </ol>                                              |
| Too many large DNA fragments ( $\geq 1$ kb) in step 53, indicating improper tagmentation                                               | The ratio of DNA versus Tn5 transposase was suboptimal; too much DNA and not enough Tn5 transposase in the tagmentation reaction                                                                                                                                                                                                       | Increase the amount of Tn5 transposase and perform a titration test to determine the best ratio in tagmentation for the optimal size of DNA fragment (300-800 bp).                                                                                                                                                                                            |
| Low DNA yield of PCR product for the final ChIA-PET library in step 69 (<30 ng after 13 cycles OF PCR amplification)                   | <ol style="list-style-type: none"> <li>Not enough cells as starting material.</li> <li>Poor ligation efficiency due to suboptimal conditions for nuclei permeabilization.</li> <li>Poor ligation efficiency due to improper preparation of the bridge linker.</li> </ol>                                                               | <ol style="list-style-type: none"> <li>Increase cell numbers.</li> <li>Try gentler conditions for permeabilization.</li> <li>Change to a new prep of bridge linker.</li> </ol>                                                                                                                                                                                |
| <b>Troubleshooting after DNA sequencing analysis based on ChIA-PET library statistical output (Table 2)</b>                            |                                                                                                                                                                                                                                                                                                                                        |                                                                                                                                                                                                                                                                                                                                                               |
| Low fraction of read pairs with bridge linker (<70%) in raw 7 (fraction of read pairs with linker)                                     | <ol style="list-style-type: none"> <li>The size of DNA templates in the final library was too big (&gt;0.8 kb), so that the sequencing may not read through to the linker.</li> <li>Inefficient washing of streptavidin beads for the enrichment of ligated DNA fragments carrying the biotin-linker for ligation products.</li> </ol> | <ol style="list-style-type: none"> <li>Increase the amount of Tn5 transposase to optimize tagmentation reaction, to ensure that most DNA fragments in the final library are in the size range of 300-600 bp.</li> <li>Use freshly made iBlock buffer and wash buffer to ensure the enrichment of ligation products (step 56).</li> </ol>                      |



**Table 3** Troubleshooting Guide for In Situ ChIA-PET, *continued*

| Problem                                                                                                                                                                                                               | Possible reason                                                                                                                                                                                                | Solution                                                                                                                                                      |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Final library with high redundancy or low complexity in raw 13 (PET redundancy)                                                                                                                                       | a. Not enough starting material.<br>b. Too much PCR amplification (more than 13 PCR cycles).<br>c. The library was over sequenced.                                                                             | a. Increase the number of cells for starting material.<br>b. Make sure you use the minimum number of PCR cycles (<13 PCR cycles).<br>c. Do not over sequence. |
| Low ratio of intra-chr/inter-chr PET (<1) in raw 18 (ratio of intra/inter PET), indicating more nonspecific interaction data than expected; most inter-chr data were nonspecific due to technical or biological noise | The nuclei permeabilization conditions might be too disruptive, leading to the rupture of nuclei, resulting in most ligation happening in nuclei-free solution, instead of <i>in situ</i> inside intact nuclei | Use gentler conditions for nuclei permeabilization, including decreased concentration of SDS in solution and reduced temperature of incubation                |

causes, and potential solutions based on our experience.

### Understanding Results

*In situ* ChIA-PET yields genome-wide chromatin contacts mediated by a specific protein (e.g., CTCF or RNAPII), along with the protein binding coverage. By running our data processing pipeline, ChIA-PIPE, with default parameters (Lee et al., 2020), detailed summary statistics of the *in situ* ChIA-PET library will be generated (example in Table 2, see below), and chromatin contacts are saved in a bedpe file format and are visualized in our BASIC Browser as arcs between two genomic loci, with the loop height denoting the interaction frequency (Fig. 5C). The protein binding profile (akin to that from a ChIP-seq experiment) is saved as bedgraph and bigwig files, which can be visualized in any genome browser. Thus, ChIA-PET offers two layers of information: chromatin contacts associated with a specific protein, and that protein's binding coverage. ChIA-PIPE also produces a 2D contact map file in `.hic` file format to be visualized on Juicebox (Fig. 5B), and a peak file (`.bed` file format) generated by either SPP (Kharchenko, Tolstorukov, & Park, 2008) or MACS2 (Zhang et al., 2008) peak callers.

As mentioned, after the library is sequenced, ChIA-PIPE automatically processes the sequencing reads and generates a summary statistics report for quality assessment (Table 2). Key quality metric values are: (1) fraction of read pairs with linker, indicating the percentage of sequencing reads with bridge linkers; (2) redundancy, which quantifies the library complexity; and (3) number of intra-

chromosomal and inter-chromosomal PETs. A high-quality library should have a high fraction of read pairs with linker, low redundancy, a low number of inter-chromosomal PETs, and a high number of intra-chromosomal PETs (Lee et al., 2020). We recommend performing a shallow QC sequencing (5-10 million sequencing reads) for quality assessment before a large-scale sequencing experiment (~200 million sequencing reads).

Below we comment on some of the parameters included in the summary statistics report and their meaning:

**Total read pair:** For decent coverage, we routinely aim for 200-500 million paired end reads ( $2 \times 150$  bp) for each library dataset.

**Read pair with bridge linker:** we use a bridge linker to connect two chromatin fragments in the ChIA-PET protocol. The higher the fraction of read pairs containing the linker, the higher the number of chromatin interactions that will be captured. Usually, the fraction value is around 0.7-0.9.

**Non-redundant PET:** The most useful data for chromatin interaction is the uniquely mapped non-redundant PET. The more the better. Usually >10 million non-redundant PET per dataset will provide robust interactions data.

**PET redundancy:** The PET redundancy reflects the complexity of a library. Lower PET redundancy means higher library complexity. Most ChIA-PET libraries would reach plateau after total reads >500 million.

**Intra-chr PET:** The intra-chromosome PETs are the most useful data for discovery of *cis*-acting elements involved in chromatin interactions. The number of PETs varies in the range of millions to 100s of millions,



depending on library sequencing depth and protein factors tested.

**Ratio of intra/inter PET:** This is a very useful value to assess the noise level of proximity ligation. Most inter-chromosomal PETs are derived from random nonspecific ligation, except for possible translocation events in cancer cells and likely rare specific inter-chromosomal contacts, which will be analyzed separately according to your research interests (not included in this protocol). In our experience, most of the high-quality ChIA-PET library datasets have a ratio value higher than 2, meaning more intra-chromosome PETs than inter-chromosome PETs.

**Intra-chr PET cluster:** The cluster of PETs is a group of PETs that are overlapped in mapping alignment to the reference genome, indicating that these PETs are derived from chromatin interactions of the same loci and repeatedly captured in different cells. The minimal PET count of a cluster is 2, and could be up into the 1000s. The number of a PET cluster measures the relative frequency of chromatin interactions between two loci in a cell population. A higher PET count in a cluster suggests higher contact frequency and higher confidence in the interaction data.

**Ratio of intra/inter PET cluster:** The PET clusters derived from inter-chromosomal PET data can be used as random background noise, and the ratio of intra/inter PET cluster (at least >1) provides the best indicator of the quality of a ChIA-PET library.

Based on this, and as shown in Table 2, the quality of the CTCF ChIA-PET library data is higher, but that of the RNAPII ChIA-PET library is also high for capturing RNAPII-associated chromatin interactions.

### Time Considerations

As shown in Figure 3, the experimental protocol can be completed in 5 days. The protocol recommends overnight restriction digestions and proximity ligations to achieve better results, even though incubation times can be as short as 2 hr at 37°C for enzyme digestion and 4 hr at room temperature for ligation. Several stopping points are possible after the DNA is purified, and are highlighted in the protocol. The data processing time generally takes 1-3 hr for 10 million reads and 1-2 days for 300 million reads using a multi-thread high-performance computer (Threads = 10 – 20, Memory = 20 – 60 GB, Cluster OS = CentOS 6.5, CPU = Intel Xeon E5-2670 @ 2.60 GHz).

### Acknowledgments

This work is mainly supported by ENCODE Phase 4 project (UM1-HG009409).

### Author Contributions

**Ping Wang:** conceptualization, methodology, project administration, writing original draft; **Yijun Ruan:** conceptualization, methodology, supervision, funding acquisition, project administration, writing review and editing; **Yuliang Feng:** methodology; **Kun Zhu:** methodology; **Haoxi Cai:** methodology; **Ya-ting Chang:** methodology; **Xiaofei Yang:** methodology; **Xiyuan Liu:** methodology; **Chen Shen:** methodology; **Eva Gega:** methodology; **Byoungkoo Lee:** software visualization; **Minji Kim:** software, visualization; **Xiaoan Ruan:** methodology, supervision.

### Conflict of Interest

The authors declare no conflict of interest.

### Data Availability Statement

For research transparency, we make available the in-house ChIA-PET pipeline and BASIC browser scripts in the following Github repository:

<https://github.com/>

*TheJacksonLaboratory/ChIA-PIPE.git*

<https://github.com/>

*TheJacksonLaboratory/basic-browser*

The data used on Basic Protocol 2 is openly available in Zenodo, at <http://doi.org/10.5281/zenodo.4706038>.

### Literature Cited

- Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., ... Cavalli, G. (2017). Multiscale 3D genome rewiring during mouse neural development. *Cell*, 171, 557–572.e24.
- Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., ... Ren, B. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature*, 518, 331–336.
- Fang, R., Yu, M., Li, G., Chee, S., Liu, T., Schmitt, A. D., & Ren, B. (2016). Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Research*, 26, 1345–1348.
- Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Han, X., Mohamed, Y. B., ... Ruan, Y. (2009). An oestrogen receptor  $\alpha$ -bound human chromatin interactome. *Nature*, 462, 58–64.
- Gallagher, S. R. (2012). One-dimensional SDS gel electrophoresis of proteins. *Current Protocols in Molecular Biology*, 97, 10.2A.1–10.2A.64. doi: 10.1002/0471142727.mb1002as97
- Goh, Y., Fullwood, M. J., Poh, H. M., Peh, S. Q., Ong, C. T., Zhang, J., ... Ruan, Y. (2012).

- Chromatin Interaction Analysis with Paired-End Tag Sequencing (ChIA-PET) for mapping chromatin interactions and understanding transcription regulation. *Journal of Visualized Experiments*, 3770.
- Ji, X., Dadon, D. B., Powell, B. E., Fan, Z. P., Borges-Rivera, D., Shachar, S., ... Young, R. A. (2016). 3D chromosome regulatory landscape of human pluripotent cells. *Cell Stem Cell*, 18, 262–275.
- Kharchenko, P. V., Tolstorukov, M. Y., & Park, P. J. (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature Biotechnology*, 26, 1351–1359.
- Lee, B., Wang, J., Cai, L., Kim, M., Namburi, S., & Tjong, H. ... Li, S. (2020). ChIA-PIPE: A fully automated pipeline for comprehensive ChIA-PET data analysis and visualization. *Science Advances*, 6, eaay2078.
- Li, G., Fullwood, M. J., Xu, H., Mulawadi, F. H., Velkov, S., & Vega, V., ... Sung, W.-K. (2010). ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biology*, 11, R22.
- Li, G., Sun, T., Chang, H., Cai, L., Hong, P., & Zhou, Q. (2019). Chromatin interaction analysis with updated ChIA-PET Tool (V3). *Genes*, 10, 554.
- Li, X., Luo, O. J., Wang, P., Zheng, M., Wang, D., Picuch, E., & Ruan, Y. (2017). Long-read ChIA-PET for base-pair-resolution mapping of haplotype-specific chromatin interactions. *Nature Protocols*, 12, 899–915.
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., ... Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326, 289–293.
- Mukhopadhyay, A., Deplancke, B., Walhout, A. J. M., & Tissenbaum, H. A. (2008). Chromatin immunoprecipitation (ChIP) coupled to detection by quantitative real-time PCR to study transcription factor binding to DNA in *Caenorhabditis elegans*. *Nature Protocols*, 3, 698–709.
- Mumbach, M. R., Rubin, A. J., Flynn, R. A., Dai, C., Khavari, P. A., Greenleaf, W. J., & Chang, H. Y. (2016). HiChIP: Efficient and sensitive analysis of protein-directed genome architecture. *Nature Methods*, 13, 919–922.
- Phanstiel, D. H., Boyle, A. P., Heidari, N., & Snyder, M. P. (2015). Mango: A bias-correcting ChIA-PET analysis pipeline. *Bioinformatics*, 31, 3092–3098.
- Phelan, K., & May, K. M. (2015). Basic techniques in mammalian subculture. *Current Protocols in Cell Biology*, 66, 1.1.1–1.1.22. doi: 10.1002/0471143030.cb0101s66
- Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., ... Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159, 1665–1680.
- Read, J. E. (2017). Chromatin immunoprecipitation and quantitative real-time PCR to assess binding of a protein of interest to identified predicted binding sites within a promoter. In D. Gould (Ed.) *Mammalian synthetic promoters* (pp. 23–32). New York, NY: Springer New York. Available at: [https://doi.org/10.1007/978-1-4939-7223-4\\_3](https://doi.org/10.1007/978-1-4939-7223-4_3)
- Sima, J., Chakraborty, A., Dileep, V., Michalski, M., Klein, K. N., Holcomb, N. P., ... Gilbert, D. M. (2019). Identifying cis elements for spatiotemporal control of mammalian DNA replication. *Cell*, 176, 816–830.e18.
- Tang, Z., Luo, O. J., Li, X., Zheng, M., Zhu, J. J., Szalaj, P., ... Ruan, Y. (2015). CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell*, 163, 1611–1627.
- Wang, P., Tang, Z., Lee, B., Zhu, J. J., Cai, L., Szalaj, P., ... Ruan, Y. (2020). Chromatin topology reorganization and transcription repression by PML-RAR $\alpha$  in acute promyeloid leukemia. *Genome Biology*, 21, 110.
- Wei, C.-L., Wu, Q., Vega, V. B., Chiu, K. P., Ng, P., Zhang, T., ... Ruan, Y. (2006). A Global Map of p53 transcription-factor binding sites in the human genome. *Cell*, 124, 207–219.
- Weintraub, A. S., Li, C. H., Zamudio, A. V., Sigova, A. A., Hannett, N. M., Day, D. S., ... Young, R. A. (2017). YY1 is a structural regulator of enhancer-promoter loops. *Cell*, 171, 1573–1588.e28.
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoutte, J., Johnson, D. S., Bernstein, B. E., ... Liu, X. S. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, 9, R137.