

Contents lists available at [SciVerse ScienceDirect](#)

Methods

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

Review Article

A genome-wide 3C-method for characterizing the three-dimensional architectures of genomes

Zhijun Duan^{a,b,*}, Mirela Andronescu^c, Kevin Schutz^d, Choli Lee^c, Jay Shendure^c, Stanley Fields^{b,c,e}, William S. Noble^{c,*}, C. Anthony Blau^{a,b,c,*}^aInstitute for Stem Cell and Regenerative Medicine, University of Washington, USA^bDepartment of Medicine, University of Washington, USA^cDepartment of Genome Sciences, University of Washington, USA^dGraduate Program in Molecular and Cellular Biology, University of Washington, USA^eHoward Hughes Medical Institute Seattle, WA 98195-8056, USA

ARTICLE INFO

Article history:

Available online xxxxx

Keywords:

Chromatin

Chromosome

Chromosome conformation capture (3C)

Chromosome conformation capture on chip (4C)

Genome architecture

Three-dimensional (3D) organization

ABSTRACT

Accumulating evidence demonstrates that the three-dimensional (3D) organization of chromosomes within the eukaryotic nucleus reflects and influences genomic activities, including transcription, DNA replication, recombination and DNA repair. In order to uncover structure-function relationships, it is necessary first to understand the principles underlying the folding and the 3D arrangement of chromosomes. Chromosome conformation capture (3C) provides a powerful tool for detecting interactions within and between chromosomes. A high throughput derivative of 3C, chromosome conformation capture on chip (4C), executes a genome-wide interrogation of interaction partners for a given locus. We recently developed a new method, a derivative of 3C and 4C, which, similar to Hi-C, is capable of comprehensively identifying long-range chromosome interactions throughout a genome in an unbiased fashion. Hence, our method can be applied to decipher the 3D architectures of genomes. Here, we provide a detailed protocol for this method.

Published by Elsevier Inc.

1. Introduction

Genomes carry both genetic and epigenetic information and serve as a scaffold for reading and transmitting both types of inheritable information. Insight into the 3D organization of the genome, including the multilevel folding and positioning of chromosomes within the nucleus, is essential for understanding various genomic functions [1]. To date, two types of tools have been used to dissect chromosome structure: microscopy-based imaging technologies and more recently developed molecular and biochemical tools. DNA imaging technologies are based on electron microscopy and light microscopy (reviewed in [2–4]). Electron microscopes have been typically employed in studies using cell-free systems, whereas light microscopy-based techniques, such as DNA fluorescence *in situ* hybridization (FISH) [5] and live-cell imaging [6] have been applied to visualize the organization of chromosomes in the nuclei of single cells *in situ*. Although microscopy has provided important insights into the 3D architecture of chromosomes, including their dynamic nature and non-random organization, lim-

itations in resolution and throughput have reduced microscopy's utility in understanding genome structure-function relationships.

During the past decade, several biochemical methods have been developed for characterizing genome architecture (reviewed in [7,8]). By measuring spatial proximity, these new techniques offer detailed molecular views of chromosome structure beyond the resolution limits of microscopy. One subset of techniques includes chromatin immunoprecipitation (ChIP) and DamID methods, which probe physical contacts between genomic loci of interest and nuclear landmarks such as the nuclear envelope or nucleolus, yielding important information about the position of genomic loci in nuclear space [9–12]. Another set of molecular tools, including RNA-TRAP [13] and 3C-based methods [14], are able to measure the relative spatial proximity between individual genomic loci, providing insight into the local or global folding of chromosomes and into the relative positioning of individual chromosomes in relationship to one another.

The relative simplicity of 3C has led to its widespread adoption in studies of long-range chromatin interactions, making it and its derivatives the most commonly used tools for characterizing chromosome structure [15–25]. 3C is based on the principle of proximity ligation. Briefly, under conditions of very low DNA concentrations (usually less than 0.8 µg/µl), ligation between two

* Corresponding authors. Fax: +1 206 685 1357.

E-mail addresses: zjduan@uw.edu (Z. Duan), william-noble@uw.edu (W.S. Noble), tblau@uw.edu (C. Anthony Blau).

cross-linked chromatin fragments is strongly favored over random inter-molecular ligation between two unassociated chromatin fragments [14]. All of the restriction enzyme digestion-based 3C techniques share four experimental steps: (1) cells are fixed with formaldehyde, which cross-links chromatin interactions; (2) the cross-linked chromatin is digested with a restriction enzyme (RE1); (3) DNA ends are ligated under conditions that favor intra-molecular ligation (proximity ligation); and (4) cross-links are reversed and DNA is recovered. However, the various 3C derivatives differ in their downstream steps for detecting chromatin interactions.

We recently developed a genomic method for mapping all the chromatin interactions that occur within a genome in an unbiased manner [26]. Briefly, the method starts with construction of a 3C library, followed by digestion of the library with a second restriction enzyme (RE2). As in the 4C protocol, the resulting DNA fragments are circularized to form small DNA circles. The circular DNA is subsequently digested again with the primary 3C RE1 to linearize the DNA. The re-opened RE1 sites serve as anchoring sites for the interacting DNA fragments and are ligated with an adapter containing an EcoP15I restriction site. The anchoring sites are then marked with biotin through DNA circularization, and the DNA circles are cut by the enzyme EcoP15I to produce biotin-labeled paired-end tags of 25–27 bp. The resulting biotin-labeled paired-end tags, representing the interacting DNA fragments, are pulled down with streptavidin beads, and paired-end sequencing enables the detection of ligation junctions (Fig. 1).

Chromatin interaction libraries generated with our method consist of DNA molecules with uniform structure and size (Fig. 1B), unlike those constructed with other recently developed similar methods such as Hi-C [19,21] and TCC [18]. This unique feature of our method provides a straightforward way to calculate the interaction frequency of each individual chromatin interaction. Therefore, our method can be very useful for characterizing the 3D architectures of relatively simple genomes at unprecedented

resolution (kb) as well as for the identification of functionally relevant (statistically significant) long-range chromatin interactions between distant genomic elements (such as promoter-enhancer interactions) on a whole-genome scale. In principle, this method is applicable to all genomes. Here, we describe the step-by-step protocol for the haploid budding yeast genome.

2. Our method

2.1. The experimental procedure

2.1.1. Cross-linking of yeast cells with formaldehyde

To capture dynamic chromosomal contacts, it is necessary to covalently link the interacting protein–protein or protein–DNA partners together. There are several cross-linking agents available. Among them, formaldehyde is the most widely used because (1) formaldehyde is cell-permeable; (2) the cross-linking reaction mediated by formaldehyde is very efficient and readily controllable (usually temperature and reaction time are the two adjustable parameters); (3) formaldehyde-mediated cross-linking can be conveniently reversed; and (4) formaldehyde is readily commercially available. Hence, we used formaldehyde to cross-link yeast cells. Unlike mammalian cells, yeast cells are protected by a cell wall. To achieve efficient restriction enzyme digestion of the chromosomes in the yeast nuclei, it is necessary to disrupt the cell wall and to isolate spheroplasts. There are two alternative strategies to prepare cross-linked yeast spheroplasts: either isolating spheroplasts before carrying out cross-linking or cross-linking the cells first before isolating spheroplasts. However, it is possible that the 3D organization of the yeast genomes might be disturbed during the processing of spheroplast isolation before cross-linking. We therefore chose the latter strategy. Another general consideration before starting a 3C-based experiment is related to the probabilistic nature of the results from 3C-based techniques. 3C methods are unable to describe the topological variability between individual

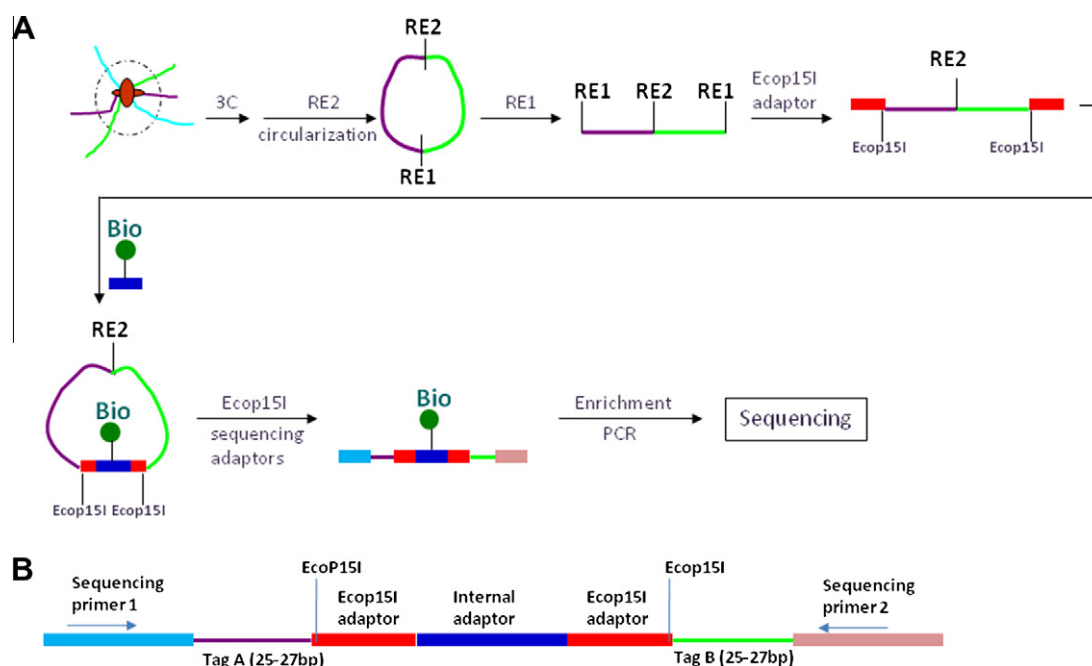


Fig. 1. Overview of the method. (A) Schematic outline of the method. The method starts with the construction of a 3C library. The library is then digested with a second restriction enzyme (RE2) and followed by DNA circularization. The resulting circular DNA is digested again with the primary 3C RE1. The re-opened RE1 sites are ligated to an adaptor containing an EcoP15I restriction site and marked with biotin through DNA circularization, and the DNA circles are cut by the enzyme EcoP15I to produce biotin-labeled paired-end tags of 25–27 bp, which are isolated, PCR amplified and finally sequenced on a next generation sequencing platform. (B) Schematic of the structure of the DNA molecules containing the paired-end tags in the chromatin interaction library constructed using our method.

cells in a given population, whereas genome topologies can be quite different between cells at different cell cycle stages. Therefore, in certain experiments when cell homogeneity is a necessity, yeast cells may first be synchronized at a particular cell cycle stage before subjected to cross-linking. Finally, it is critical to optimize cross-linking conditions, since insufficient cross-linking will result in missing some chromatin interactions, while over-cross-linking will lead to inefficient RE digestion and higher noise levels due to random chromatin collisions. As shown in Fig. 2A, we found that cross-linking with 1% formaldehyde at room temperature for 10 min is an appropriate condition for yeast cells.

Experimental steps

1. Yeast cells such as the *Saccharomyces cerevisiae* strain BY4741 (genotype: *Mata his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 bar1::KanMX*) are cultured at 30° C by shaking overnight in 50 ml of YEP media plus 2% glucose.
2. Cultured cells are diluted the next morning to an OD600 = 0.2 in one liter of YEP plus 2% glucose. Cells are incubated with shaking at 30° C until reaching an OD600 = 1.0 (about 3–4 h). Since the 3D architectures of yeast genomes are related to cell state, it is important to keep the yeast cells healthy and not over-grown.

3. Cells are treated with 27.7 ml of 37% formaldehyde (final concentration = 1%) for 10 min at room temperature with constant stirring.
4. Fixation is quenched with 52.6 ml of 2.5 M glycine (final concentration = 0.125 M) for 15 min at room temperature with constant stirring.
5. Fixed cells are collected via centrifugation (1500xg - 5 min) and re-suspended in 50 ml of spheroplast buffer plus 30 mM dithiothreitol (DTT).
6. Fixed cells are recollected via centrifugation (1500g - 5 min) and re-suspended in 50 ml of spheroplast buffer plus 1 mM DTT.
7. Fixed cells are converted to spheroplasts with Zymolyase 20T (MP Biomedicals LLC.) (Final concentration = 0.66 g/L) treatment at 30° C with gentle rotation. Conversion to spheroplasts is confirmed by microscopy.

! Troubleshooting

8. Fixed Spheroplasts are collected via centrifugation at 4° C (1500g - 5 min).
9. Fixed spheroplasts are washed twice in 50 ml spheroplast buffer and collected via centrifugation at 4° C (1500g - 5 min).

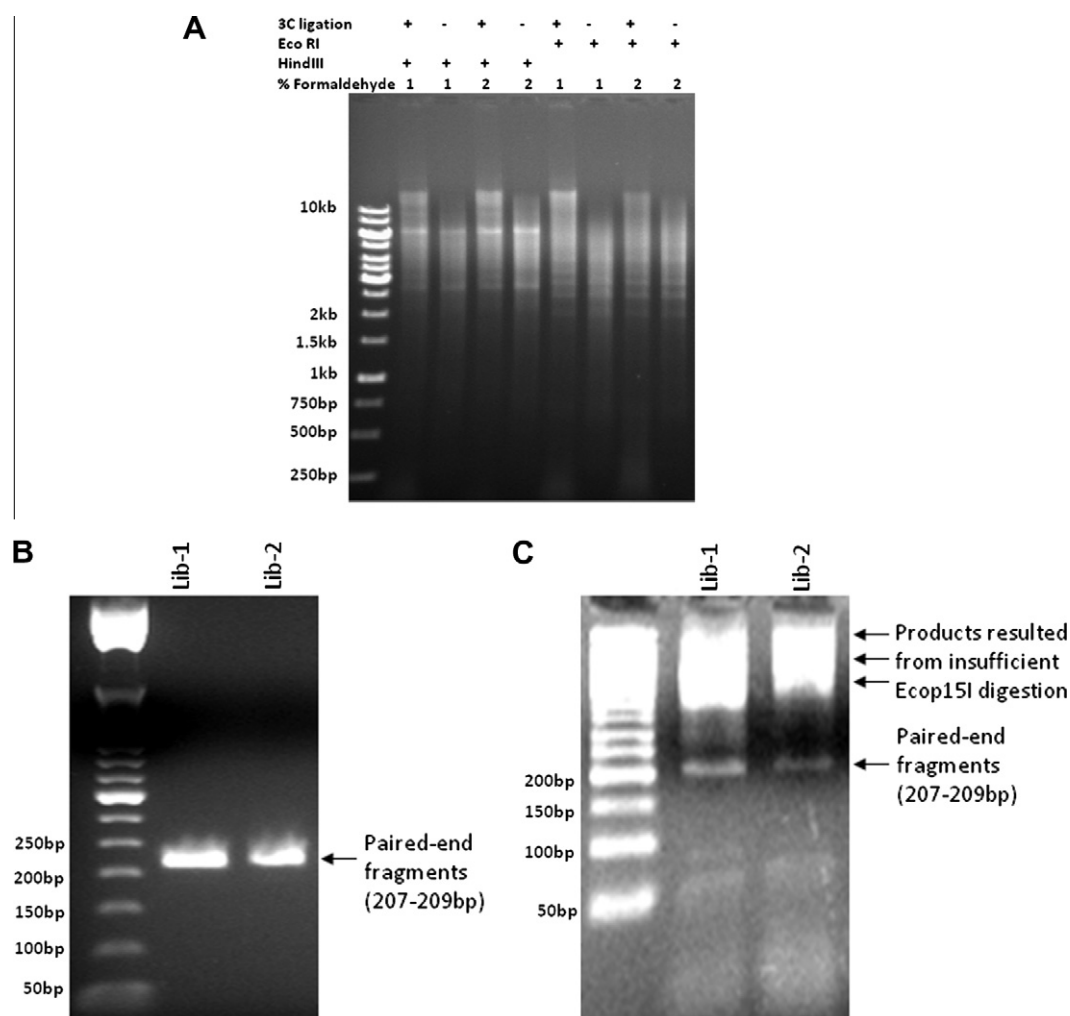


Fig. 2. Representative results of quality control assays. (A) Examining the efficiency of RE1 digestion and 3C ligation by DNA gel electrophoresis. DNA samples were prepared from RE1(Hind III or EcoRI)-digested cross-linked (with 1% or 2% formaldehyde) yeast cells before or after 3C ligation as described in the main text. Samples were then resolved in a 0.5% agarose gel. The 1 kb DNA ladder (Fermentas) was used for estimating the size of the DNA fragments. (B) Examples of DNA gel electrophoresis of the PCR products of the interaction libraries constructed using our method. Two libraries were shown here. The 50 bp DNA ladder (NEB) was used. (C) Results of DNA gel electrophoresis of the PCR products of two bad libraries resulted from the inefficient digestion of EcoP15I (step 41).

10. Fixed spheroplasts are re-suspended in 50 ml of the appropriate 1× restriction enzyme buffer, depending upon the restriction enzyme to be used. For HindIII or EcoRI, use 1× NEBuffer 2 (B7002S) and aliquot into 50 1.7-ml microcentrifuge tubes (1 ml per tube, containing about $1\text{--}2 \times 10^9$ spheroplasts). Collect spheroplasts via centrifugation (2000g – 5 min) in a refrigerated desktop centrifuge. Spheroplasts can be rapidly frozen in liquid nitrogen and stored at -80°C until use (up to 1 year).

2.1.2. Digestion of cross-linked cells with the first enzyme (RE1, e.g. Hind III or EcoRI)

Efficient RE1 digestion of chromatin is critical for the successful construction of a 3C library. Criteria for RE1 choice are discussed in detail in Section 2.3.1. Before RE1 digestion, the cross-linked spheroplasts need to be treated with the anionic surfactant, sodium dodecyl sulfate (SDS), to remove non-cross-linked proteins and to make the chromatin accessible for RE1 digestion. However, SDS treatment at 65°C can reverse formaldehyde-mediated cross-linking. Hence, it is important to optimize the incubation time and temperature for SDS treatment. We found that treating cross-linked spheroplasts with 0.6% SDS at 65°C for 20 min followed by 1 h at 37°C with shaking can lead to good outcomes. After chromatin solubilization, the non-ionic detergent, Triton X-100, should be added to the reaction to neutralize the SDS before carrying out RE1 digestion.

Experimental steps:

11. Add 1 ml 1× NEBuffer 2 to each of four tubes of fixed spheroplasts. Mix well and aliquot each tube of spheroplasts into eight micro-tubes (250 µl per tube), i.e. 32 micro-tubes in total.
12. Add 15 µl 10% SDS (final concentration = 0.6%) to each tube and incubate at 65°C for 20 min followed by 1 h at 37°C with shaking.
13. Add 250 µl 1× NEBuffer 2, 6 µl 10× NEBuffer 2 and 50 µl 20% Triton X-100 (final concentration = 2%) to each tube, mix carefully and incubate at 37°C for 1 h with shaking.
14. Add 800 U of RE1 (Hind III) per tube, mix well, and incubate reaction overnight at 37°C with shaking.
15. Add 112 µl 10% SDS (final concentration = 1.8%) to each tube and incubate at 65°C for 20 min. Two tubes of samples can be used for reversing cross-linking and DNA purification (Section 2.1.4) to measure the amount of DNA in each tube (usually should be less than 5 µg DNA in each tube) and to assess the RE1 digestion efficiency by real-time PCR. It is also useful to check RE1 digestion efficiency by DNA electrophoresis (Fig. 2A)

! TROUBLESHOOTING

2.1.3. In-situ (3C) ligation

All 3C methods are based on proximity ligation, the principle of which is that extremely low DNA concentrations strongly favor ligations within a single molecule over ligations between two molecules. Linked DNA fragments within the same chromatin complex behave as a single molecule. Hence, low DNA concentrations will reduce the noise from random inter-molecular ligations. However very low DNA concentrations require large reaction volumes, increasing cost, and complicating the recovery of DNA. We chose to carry out the 3C ligation reaction in a volume of 25 ml with a DNA concentration of 0.3 or 0.5 ng/µl.

Experimental steps:

16. Combine every three-microtubes of RE1 digested samples into a 50-ml disposable conical tube (i.e. distribute the 30 micro-tubes of RE1 digested samples into ten 50-ml tubes), add the following to each tube and incubate for 1 h at 37°C :
18.65 ml distilled water (molecular biology grade)
2400 µl 10× ligation buffer
1035 µl 20% Triton X-100
240 µl BSA
17. Add 250 U of T4 DNA ligase (Fermentas, 5u/µl) to each tube and incubate at 16°C for 4 h s and 25°C for 1 h.

2.1.4. Reversing cross-linking and DNA (3C library) purification

After 3C ligation, proteins in the chromatin complexes are digested by proteinase K in the presence of SDS. DNA is then recovered by isopropanol-mediated precipitation.

Experimental steps:

18. Add 200 µl of 20 mg/ml proteinase K and 1 ml 10% SDS to each ligation mixture and incubate overnight at 65°C .
19. The following day, add 50 µl 20 mg/ml proteinase K to each tube and incubate at 55°C for 2 h.
20. Add 20 µl GlycoBlue (Ambion), 1/10 volume of 3 M sodium acetate (pH 5.2) and an equal volume of isopropanol to precipitate DNA at -80°C for 2 h. Centrifuge for 2 h at 2987g (Sorvall 75006441 swinging bucket rotor) at 4°C .
21. Dissolve the DNA pellet of each 50-ml tube in 1.5 ml 1×TE buffer and transfer each sample into two 1.7 ml tubes. Add 4 µl 1 mg/ml RNase A to each tube and incubate at 37°C for 30 min.
22. Extract DNA with an equal volume of phenol: chloroform (Invitrogen) and repeat three times.
23. Following the third extraction, precipitate DNA with isopropanol as described in step 20 and wash the pellet three times with 70% ethanol.
24. Air-dry the pellets for 10 min and add 60 µl water. Pool the entire DNA sample (the 3C library) and determine its concentration with a Nanodrop-1000 spectrophotometer (Thermo Scientific). The yield of purified DNA from each 25 ml ligation reaction should be between 8–13 µg. It is also useful to check 3C ligation efficiency by DNA electrophoresis (Fig. 2A)

2.1.5. Digestion of the 3C library with the second restriction enzyme (RE2)

At this point, the construction of a typical 3C library is completed. However, DNA samples of a 3C library are usually not suitable for high throughput sequencing on current next generation sequencing platforms due to their large range of size distribution. For example, the size distribution of a HindIII-mediated 3C library of the haploid yeast genome ranges from less than 1 kb to bigger than 10 kb (Fig. 2A). Hence, all the downstream experimental steps aim at coupling the 3C technique with the next generation sequencing technologies to achieve comprehensive identification of chromatin interactions on a whole-genome scale. In the 3C library constructed above, each pair of interacting DNA fragments is joined together at a RE1 site. Hence, these RE1 sites can serve as markers for the chromatin interactions. To label these RE1 sites with biotin for further isolation, it is necessary to re-open these sites. To release these RE1 sites without losing the chromatin pairing information, it is necessary to find a second anchor point for each pair of interacting DNA fragments. The more frequently occurring RE2 sites can serve as the second anchor points.

Experimental steps:

25. Digest about 30 µg of the 3C library DNA overnight (37 °C) at a concentration of 10 ng/µl with 1000 U of RE2 (e.g. MspI or MseI, NEB) in the appropriate RE2 buffer. Following overnight incubation, add an additional 100 Units of RE2 to the reaction mixture and incubate at 37 °C for two more hours.
26. Precipitate DNA with isopropanol as described in step 20 and further clean up the sample with the Qiaquick PCR purification kit (Qiagen) according to the manufacturer's instruction. Determine the DNA concentration with a Nanodrop-1000 spectrophotometer.

2.1.6. DNA circularization, circular DNA purification and RE1 re-linearization

After RE2 digestion, DNA fragments in the 3C library are circularized to form DNA circles. Hence, each pair of the interacting DNA fragments is now connected via an RE1 site at one end and an RE2 site at the other end (Fig. 1B). DNA fragments that failed to form circles can be degraded by the ATP-dependent plasmid-safe DNase. Since the purified circular DNA is not stable, it must be linearized immediately.

Experimental steps:

27. For each library, circularize about 12 µg of RE2-digested DNA with 200 U T4 DNA ligase (Fermentas, 5 U/µl) in a 24 ml reaction volume (in a 50 ml tube) overnight at 16 °C.
28. Precipitate the circularized DNA with isopropanol, and wash the pellet three times with 70% ethanol as described in previous steps. Air-dry the pellet and dissolve it in 657.5 µl water.
29. Degrade linear DNA with 100 U ATP-dependent plasmid-safe DNase (10 U/µl, Epicentre) in 750 µl 1× ATP-dependent plasmid-safe DNase buffer containing 1 mM ATP overnight at 37 °C.
30. Precipitate the remaining circular DNA with isopropanol as described in step 20 and dissolve the pellet in 200 µl water. Further clean up the DNA sample with the Qiaquick PCR kit according to the manufacture's instruction. Elute DNA in 160 µl water.
31. Linearize the purified circular DNA immediately (circular DNA is not very stable) with RE1 by adding 200 U RE1 enzyme (e.g. FastDigest® HindIII or EcoRI, Fermentas, 10 u/µl) and 20 µl 10× FastDigest® buffer, and incubate at 37 °C for 30 min.
32. Purify the linearized DNA with the Qiaquick PCR purification kit (Qiagen) according to the manufacture's instruction. Elute DNA in 85 µl water. Determine the DNA concentration with a Nanodrop-1000 spectrophotometer.

! TROUBLESHOOTING

2.1.7. EcoP15I methylation, EcoP15I adaptor ligation, and Biotin labeling

EcoP15I is a type III restriction-modification enzyme. In the presence of ATP, EcoP15I functions as a site-specific DNA endonuclease whose cutting site is 25–27 bp downstream of its recognition site, whereas in the presence of S-adenosylmethionine and absence of ATP, EcoP15I can also function as a DNA methyltransferase. In this method, we first use its DNA methyltransferase activity to methylate the EcoP15I sites in the yeast genome to protect them from being cleaved by EcoP15I. We then take advantage of its unique restriction enzyme activity to produce paired-end tags, which represent the pairs of interacting DNA fragments.

Experimental steps:

33. Add 15 µl 10× NEBuffer 3, 1.5 µl BSA, 2.5 µl EcoP15I (10 u/µl, NEB), 1.8 µl S-adenosylmethionine (final concentration = 380 µM), and 50 µl water to 80 µl RE1 re-linearized DNA (2–3 µg), and incubate overnight at 37 °C. Purify the EcoP15I methylated DNA fragments with the Qiaquick PCR kit and elute with 70 µl water. Determine the DNA concentration with a Nanodrop-1000 spectrophotometer.
34. Ligate the corresponding EcoP15I adaptor (with one end compatible to the RE1 site) to the methylated DNA at 25 °C for 30 min in 100 µl 1× Fast ligation buffer (Fermentas) containing 400 pmol of the EcoP15I adaptor and 25 U T4 DNA ligase (5 U/µl, Fermentas).
35. Isolate the adaptor-ligated DNA fragments from the excess free adaptors via agarose gel electrophoresis and recover the DNA with the Qiaquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions. Determine the DNA concentration with a Nanodrop-1000 spectrophotometer.
36. After ligation with the corresponding EcoP15I adaptor, the RE1 site in each DNA fragment is disrupted – a feature which is designed to eliminate the chimerical DNA resulting from random ligation during the above EcoP15I adaptor ligation reaction (step 34). Therefore, digest the purified DNA again with RE1 (HindIII or EcoRI) at 37 °C for 2 h to eliminate the chimeric DNA. Purify the DNA sample with the Qiaquick PCR kit and elute with 300 µl water. Determine the DNA concentration with a Nanodrop-1000 spectrophotometer.
37. Add 150 µl adaptor-ligated DNA (0.7 µg), 75 µl 10× ligation buffer, 2 µl 1 µM biotin labeled internal adaptor, 6 µl T4 DNA ligase (5 U/µl, Fermentas) and water to a total 750 µl reaction volume and incubate overnight at 16 °C.
38. Precipitate the DNA with isopropanol as described in step 20 and purify with the Qiaquick PCR purification kit. Elute DNA in 90 µl water and determine the DNA concentration with a Nanodrop-1000 spectrophotometer.
39. Digest the DNA sample overnight at 37 °C with 50 U ATP-dependent plasmid-safe DNase (10 U/µl, Epicentre) in 100 µl 1× ATP-dependent plasmid-safe DNase buffer containing 1 mM ATP.
40. Purify the remaining circular DNA with the Qiaquick PCR kit. Elute DNA in 90 µl water and determine the DNA concentration with a Nanodrop-1000 spectrophotometer.

! TROUBLESHOOTING

2.1.8. EcoP15I digestion, end-repair and ligation of sequencing adaptors

All subsequent steps may be carried out in DNA LoBind microtubes (Eppendorf).

The paired-end tags produced by EcoP15I digestion need to be end-repaired and ligated to appropriate sequencing adaptors before the library can be amplified and subjected to high throughput sequencing on a selected sequencing platform. In step 45, we describe how we carried out the ligation reaction with our custom sequencing adaptors for the Illumina sequencing platform, which provides an example. Indeed, it should be more efficient to use the commercially available Illumina “Y-shape” PE adaptor.

Experimental steps:

41. Digest the circular DNA (84 µl from step 39) overnight at 37 °C with 20 U EcoP15I (10 U/µl, NEB) in 100 µl 1× NEBuffer 3 containing 1 µg/µl BSA, 2 mM ATP, and 100 uM Sinefungin.

42. Add an additional 0.5 μ l EcoP15I (10 U/ μ l), 0.5 μ l 100 mM ATP, and 1 μ l 10 mM Sinefungin to the reaction the next day and incubate for another 2 h.
43. Stop the reaction by incubating at 65 °C for 20 min, and then place the tube on ice for 5 min.
44. Add 1.5 μ l 25 mM dNTPs (Invitrogen) and 1 μ l Klenow (5 U/ μ l, NEB) to the reaction and incubate at 25 °C for 30 min. Stop the reaction by incubating at 65 °C for 20 min and on ice for 5 min.
45. Add 1 μ l 1 M $MgCl_2$, 2 μ l 100 mM ATP, 60 μ l 25% PEG-8000, 2 μ l each of appropriate sequencing adaptors (e.g. for the Illumina sequencing platform, 2 μ l each of 40 uM Illumina-PE-Adaptor-A and 40 uM Illumina-PE-Adaptor-B), and 5 μ l Quick Ligase (NEB) to the reaction and incubate at 25 °C for 30 min. Stop the reaction by incubating at 65 °C for 20 min and on ice for 5 min. Add 100 μ l water to the tube to bring the total volume to 300 μ l.

2.1.9. Biotin pull-down and nick repair

The biotin-labeled DNA fragments can be isolated using commercially available streptavidin Dynabeads, and the two nicks in each DNA fragment resulting from the ligation reaction (step 45) need to be repaired by DNA polymerase I.

Experimental steps:

46. Immobilize the biotin-labeled, paired-end adaptor-ligated DNA sample to 15 μ l Dynabeads M-280 Streptavidin beads (Invitrogen) according to the manufacturer's instruction.
47. Resuspend the DNA-bound M-280 beads in 200 μ l 1 \times B&W buffer and transfer to a new tube. Wash the beads twice with 200 μ l 1 \times B&W buffer and once with 200 μ l 1 \times NEBuffer 2.
48. Resuspend the beads in 34.5 μ l 1 \times NEBuffer 2, and add 4 μ l 25 mM dNTPs mix (Invitrogen) and 1.5 μ l DNA polymerase I (10 U/ μ l, NEB), mix well and incubate at 16 °C for 30 min with shaking. Wash the beads once with 200 μ l EB (Qiagen) and resuspend in 40 μ l EB.

2.1.10. Library amplification, purification and sequencing

It is important to ensure linear amplification of the library.

Experimental steps:

49. To determine the number of PCR cycles necessary to generate enough PCR products for sequencing, set up five trial PCR reactions with 12, 15, 18, 21, or 24 cycles. (For details of the PCR amplification, please see reference [27]). Determine the optimal PCR cycle number by running the PCR products on a 3% agarose gel. The expected product size is 207–209 bp (Fig. 2C).

! TROUBLESHOOTING

Table 1
Troubleshooting.

Step	Problem	Solution
7	The conversion rate of spheroplasts is low or spheroplasts is of poor quality	It is important to monitor the conversion of spheroplasts by microscopy. If the conversion rate is low, increase the incubation time until most cells are converted to spheroplasts; if the majority spheroplasts are not intact, it might be due to over-digestion by Zymolyase 20T. Optimize the enzyme activity of Zymolyase 20T and re-start from step 1
15	The efficiency of RE1 digestion is low	Choose the appropriate RE1 (see Section 2.3.1); reduce the amount of chromatin complexes in each reaction; increase the amount of RE1 in each reaction
32	Too much DNA was recovered	It is possible due to inefficient digestion of ATP-dependent plasmid-safe Dnase at step 29 (see Section 2.3.3.2). Increase the amount of Dnase or increase the incubation time.
40	Too much DNA was recovered	It is possible due to inefficient digestion of ATP-dependent plasmid-safe Dnase at step 39 (see Section 2.3.3.2). Increase the amount of Dnase or increase the incubation time.
49	Low or no specific PCR products, while a lot of big size DNA smear	It is possible due to inefficient digestion of EcoP15I at step 41 (see Section 2.3.3.3). Carefully calculate the amount of the DNA template and keep the molar ratio between the EcoP15I and the template as close as possible; increase the incubation time.

50. To amplify the remainder of the library-bound beads in a large-scale PCR with the optimal number of cycles, set up twelve 50- μ l reactions (2 μ l beads per reaction) with Phusion High-Fidelity DNA polymerase (NEB) and the appropriate library amplification primers (e.g. Illumina-lib-PCR-A and Illumina-lib-PCR-B primer pair). Pool the PCR products from the individual reactions and resolve the products on a 3% low melting agarose gel. Recover the paired end tags of the desired size range (207–209 bp) with the Qiaquick Gel Extract Kit (Qiagen). Determine the DNA concentration with a Nanodrop-1000 spectrophotometer.
51. Sequence the library with an appropriate next generation sequencing platform. In principle, the library generated with our method can be sequenced using any currently available platform. However, since the length of the paired-end tags in the library is only 25–27 bp, the Illumina and Solid platforms are the preferred choices.

2.2. Troubleshooting

Troubleshooting advice can be found in Table 1.

2.3. Quality control and discussion

Since our method is an extension of 4C, all of the general considerations for 3C [28,29] and 4C [7,30] experiments also apply to this protocol. Here, we highlight some issues which, based on our experience, are critical to achieving the best outcomes.

2.3.1. Restriction enzyme selection

There are three general considerations regarding the choice of the restriction enzymes for the first (RE1) and second (RE2) digestion: (1) cutting frequency – usually RE1 should be a 6-base cutter, whereas RE2 should be a 4-base cutter; (2) sensitivity to DNA methylation – neither RE1 nor RE2 should be sensitive to DNA methylation; and (3) digestion efficiency, which is especially important for RE1, because not all restriction enzymes are able to digest cross-linked chromatin efficiently. To date, a few enzymes, such as HindIII, EcoRI, BglII, and Nco I, have shown good performance. We also suggest an added consideration specific to the choice of RE2. Since the recognition sites of any given enzyme are not evenly distributed throughout a genome, to obtain higher resolution from a 3C library generated by an RE1, we suggest that two different RE2 combinations be used to generate two sub-libraries for sequencing. On the one hand, since these two sub-libraries are from the same 3C library, they should overlap significantly. On the other hand, since the genome-wide distributions of the recognition sites of the two RE2 enzymes are different, the two sub-libraries should complement one another. Hence, to maximize their complementarity, the greater the difference be-

tween the recognition sequences of the two RE2, such as MspI (CCGG) and MseI (TTAA), the better.

2.3.2. Background noise control

There are at least five types of noise associated with 3C experiments: (1) cross-linking-captured random collisions occurring within and between chromosomes; (2) incomplete digestion products from the restriction enzyme-mediated chromatin fragmentation step; (3) ligation of restriction sites at either end of a DNA fragment (self-ligation); (4) re-ligation of immediately adjacent DNA fragments (adjacent ligation); and (5) random inter-molecular ligation of non-adjacent DNA fragments due to Brownian motion of DNA fragments in solution during proximity ligation. Although the last four types of noise can be eliminated at the data analysis stage by employing appropriate bioinformatic tools, the inefficiencies they introduce impede large scale sequencing efforts. Furthermore, bioinformatic tools may not completely eliminate noise arising from random collisions because the frequency of such non-specific interactions is related to the genomic distance between the two interrogated sites [31]. Hence, it is important to optimize the respective experimental conditions to reduce the various types of noise during library construction. For example, the frequency of random inter-molecular ligation during proximity ligation is significantly influenced by DNA concentration. However, as mentioned in Section 2.1.3, although low DNA concentrations will reduce the noise from random inter-molecular ligation, they will also require large volumes for the ligation reaction. To estimate the DNA concentration that is sufficient to limit the frequency of random inter-molecular ligation to an acceptable level during 3C ligation, we previously constructed two independent sets of experimental libraries differing by DNA concentration (~0.5 µg/ml or ~0.3 µg/ml) at the 3C ligation step [26]. We observed that both conditions yield similar results (see discussion in Section 3.3 below), indicating that a DNA concentration of ~0.5 µg/ml is sufficient to yield good outcomes.

2.3.3. Quality control

This is a lengthy protocol and we suggest carrying out quality control analysis for several critical steps.

2.3.3.1. RE1 digestion and 3C ligation. The efficiency of RE1 digestion and 3C ligation can be qualitatively assessed by DNA gel electrophoresis (Fig. 2A). For example, after Hind III or Eco RI digestion of cross-linked yeast cells, genomic DNA should run as a smear with the majority of the bands smaller than 10 kb, whereas, after subsequent 3C ligation, there should be an apparent shift of the smear to a larger size range (Fig. 2A). The efficiency of RE1 digestion can also be quantitatively assessed using real-time PCR assays. In our previous studies, by surveying multiple randomly picked RE1 sites throughout the genome, we observed that the digestion efficiency of cross-linked yeast chromosome by Hind III or Eco RI was usually around 90% [26].

2.3.3.2. DNA circularization and purification of circular DNA (steps 27–32). Based on our experience, about 30% of RE2 digested DNA template will form circular DNA during the circularization step (step 27), and at the end of step 32, only 10% of the DNA will remain (i.e. 12 µg RE2 digested DNA template will yield about 1.2 µg RE1 linearized DNA after step 32). Too little remaining DNA indicates inefficient DNA circularization, whereas too much remaining DNA indicates inefficient digestion of ATP-dependent plasmid-safe DNase.

2.3.3.3. EcoP15I digestion. It is important to keep the molar ratio between the EcoP15I enzyme and the DNA template as close as

possible to 1:1, as both excess and insufficient amounts of enzyme will result in inefficient digestion. The PCR products of the library generated with this protocol should run in a tight band (207–209 bp, when using the Illumina paired end adaptors) on a 3% agarose gel (Fig. 2B), whereas inefficient EcoP15I digestion will result in an apparent decrease in the intensity of the specific band and the presence of a background smear (Fig. 2C).

3. Data analysis, interpretation and expected results

To characterize the genome topology and to uncover its potential functional implications, the completion of library construction and high throughput sequencing is just the first step of a long journey – biological insights cannot be obtained without sophisticated computational analysis. Here we outline the basic data analysis pipeline we have implemented for characterizing the haploid budding yeast genome [26].

3.1. Alignment of sequence reads to the reference genome

In principle, the sequence reads of the libraries constructed by our method can be mapped to the appropriate reference genome with any short sequence alignment algorithm (Maq, BWA, SOAP, Bowtie, etc.). For libraries constructed with budding yeast, due to the small size and relative simplicity of the *S. cerevisiae* genome, sequence reads can be directly mapped to the reference genome using the Maq tool (<http://maq.sourceforge.net/>) with default parameters [26]. For mammalian genomes such as the human genome, which are characterized by their immense size and complex repetitive sequences, one can take advantage of the unique feature of our libraries, i.e. each paired-end genomic tag is 25–27 bp in length with a RE1 site (HindIII or EcoRI) at its end. Instead of directly mapping to the human reference genome, a custom reference genome can be built by extracting the 60 bp genomic sequence flanking each RE1 site (HindIII or EcoRI, 30 bp upstream and 30 bp downstream), which is not only much smaller in size than the original reference genome but also drastically reduces the sequence complexity.

3.2. Identification of statistically confident chromatin interactions

The pipeline for identifying statistically confident chromatin interactions is summarized in Fig. 3. Briefly, self-ligations and ligations between adjacent restriction fragments are eliminated, and the existence of an RE1 site (e.g. HindIII or EcoRI) in each of the remaining ligation products should be confirmed. Then the mappability (i.e., uniqueness) of each RE1 (e.g. HindIII or EcoRI) fragment in the yeast genome is calculated, and all the ligation products that contained at least one unmappable fragment are discarded. To estimate a false discovery rate (FDR), the remaining intra-chromosomal interactions are subdivided into 5 kb bins as measured by the genomic distance between the midpoints of the two ligated fragments in order to account for the strong influence of genomic proximity on ligation frequency. All the remaining inter-chromosomal interactions can be placed into a separate bin. In each bin, the interactions are ranked according to their sequencing frequency and a p value relative to all other possible interactions in the same bin is assigned. Lastly, the p value of each interaction is converted into a q value (defined as the minimal FDR threshold at which the interaction is deemed significant), upon which the chromosomal interactions are ranked library-wide. This approach allows us to distinguish a propensity for relatively short-range interactions, which arise from the polymer-like behavior of chromosomes, from interactions due to higher order chromosome folding.

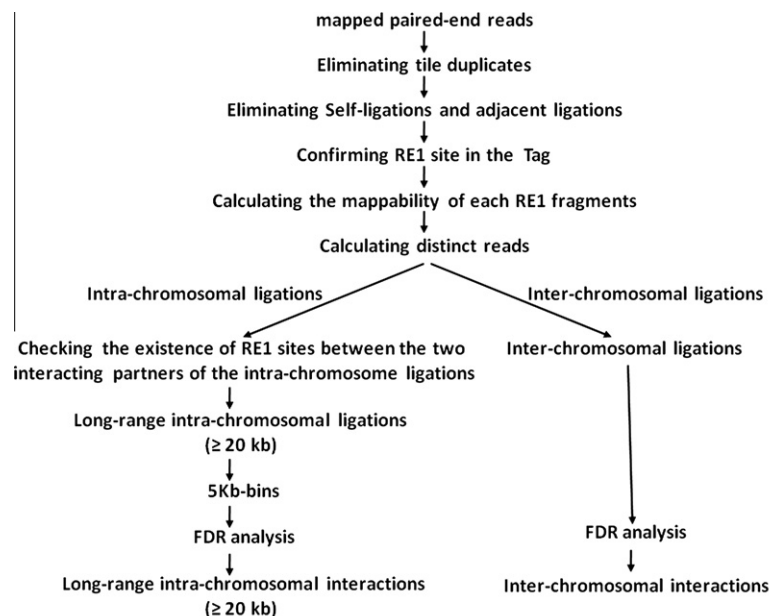


Fig. 3. Outline of the pipeline for identifying statistically confident chromatin interactions.

3.3. Statistical validation of the method

To assess whether an experiment is successful, the results from several types of statistical analysis can be used as indicators. First, due to the polymer properties of chromatin fibers and the fact that cross-linking is able to retain chromosome conformation, experimental libraries constructed with cross-linked cells should exhibit a strong inverse correlation between the frequency of intra-chromosomal ligations of RE1 (e.g. HindIII or EcoRI) fragments and their genomic distance [19,26]. In contrast, this polymer-like behavior should not be observed in control libraries constructed with uncross-linked cells or purified DNA. Second, for the same reason, the percentile of long-range (defined as ≥ 20 kb, non-adjacent) intra-chromosomal ligations should be significantly higher in the experimental libraries than in the control libraries [26]. Moreover, the observed ratio of the percentage of long-range intra-chromosomal ligations (≥ 20 kb) to that of the inter-chromosomal

ligations should be significantly higher in the experimental libraries than in the control libraries and also higher than the expected ratio of RE1 fragments (e.g. HindIII or EcoRI) (Fig. 4). Third, for all possible combinations of the RE1 (e.g. HindIII or EcoRI) fragments in a genome, high correlations between captures from the 5' ends and those from the 3' ends should be observed [26]. Fourth, high reproducibility between biological replicates should be observed.

3.4. Experimental validation of the identified chromatin interactions

Individual chromatin interactions identified by our method can be validated by DNA FISH and/or 3C experiments. Detailed descriptions of 3C confirmation experiments can be found in the 3C literature [29,32]. When using real-time PCR to quantify 3C products, it is important to confirm the specificity of the PCR primers both by DNA gel electrophoresis and melting curve analysis. FISH

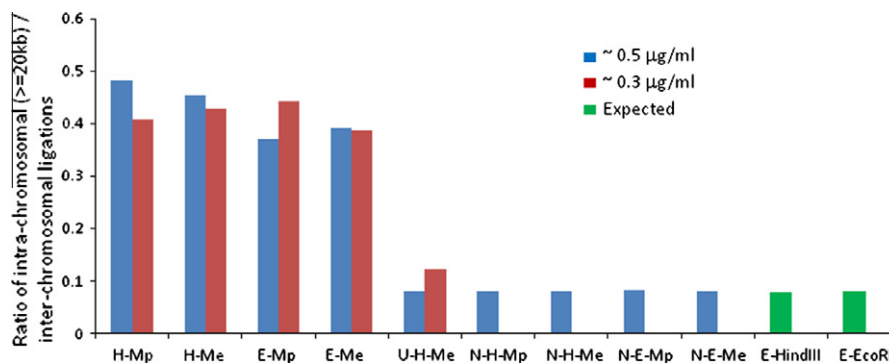


Fig. 4. Ratio of the long-range intra-chromosomal (≥ 20 kb) versus inter-chromosomal ligations in the libraries constructed by using our method. The ratio of the long-range intra-chromosomal (≥ 20 kb) versus inter-chromosomal ligations in each experimental library (H-Mp, H-Me, E-Mp, or E-Me) constructed with the cross-linked cells at both 3C ligation conditions (DNA concentration of ~ 0.5 μ g/ml or ~ 0.3 μ g/ml) is much higher than in the control libraries (U-H-Me, N-H-Mp, N-H-Me, N-E-Mp, and N-E-Me). While the ratio in each of the control libraries is close to the expected ratio (intra (≥ 20 kb)/inter) of all possible random ligation products of RE1 fragments (HindIII or EcoRI) in the yeast genome. The expected ratio of HindIII (E-HinIII) or EcoRI (E-EcoRI) fragments is 0.07877308 or 0.08018724, respectively. Hence, the ratio of the long-range intra-chromosomal (≥ 20 kb) versus inter-chromosomal ligations can serve as an indicator for estimating the quality of an experimental library constructed by our method. H-Mp, H-Me, E-Mp, and E-Me, experimental libraries with the RE1-RE2 combination of HindIII-MspI, HindIII-MseI, E-Mp, and EcoRI-MspI, respectively; U-H-Me, control library generated with uncross-linked yeast cells with the RE1-RE2 combination of HindIII-MspI, HindIII-MseI, E-Mp, and EcoRI-MspI, respectively.

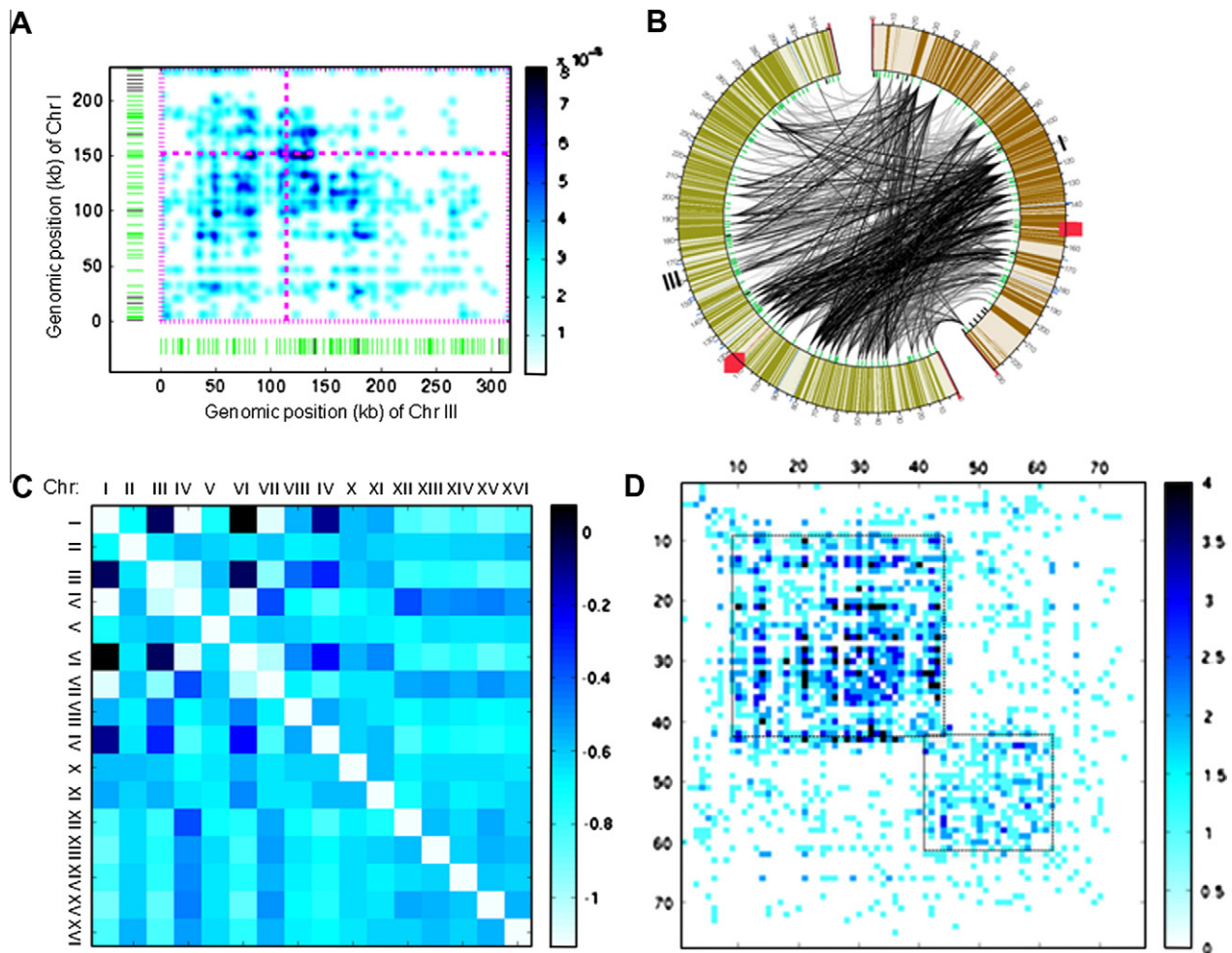


Fig. 5. Data visualization and analysis. (A) A 2D heat map showing interactions between Chromosomes I and III, generated using the intra-chromosomal interactions identified from the HindIII libraries at an FDR threshold of 1%. The chromosomal positions of telomeres (pink hatches), centromeres (dashed pink lines), mappable (green hatches) and un-mappable (black hatches) HindIII fragments are indicated. (B) A Circos diagram showing the same interactions as in (A). Each arc connects two HindIII fragments and represents a distinct interaction. The shade of each arc, from very light grey to black, is proportional to the negative log of the p-value of the interaction. The chromosomal positions of telomeres (red colored area), centromeres (red hexagon), tRNA genes (blue outer hatches), mappable (green inner hatches) and un-mappable (black inner hatches) HindIII fragments are indicated. Black outer hatches and numbers mark genomic positions. (C) A heat map showing the whole chromosome interaction patterns (enrichment log ratios) among all 16 chromosomes in yeast. Note that the strong interactions are among the short chromosomes (I, III, VI, and IX). The color scale corresponds to the natural log of the ratio of the observed versus expected number of interactions. (D) Clusters of the early-firing replication origins (Cib5-independent) revealed by using a hierarchical average-link clustering algorithm. To achieve better visualization, dashed lines were drawn to define the borders of the clusters.

experiments always serve as the “gold standard” for validating chromatin interactions identified by 3C-based methods. Recent studies have provided good examples of how to validate Hi-C data with FISH techniques [33,34].

3.5. Data visualization and interpretation

Once the chromatin interactions have been identified and validated, genome topology can be visualized and analyzed by using a variety of computational tools. For example, both folding patterns of individual chromosomes (intra-chromosomal interactions) and the interaction patterns between different chromosomes (inter-chromosomal interactions) can be visualized using either 2D heat maps (Fig. 5A) or Circos diagrams (Fig. 5B). As an example, the interaction pattern between budding yeast chromosomes I and III is shown as a 2D heat map in Fig. 5A and a Circos diagram in Fig. 5B. In both figures, the enrichment of inter-chromosome interactions around the two centromeres is apparent.

2D heat maps can also be used to visualize the chromosome-chromosome spatial relationship. As shown in Fig. 5C, by analyzing

each chromosome pair in terms of the ratio of observed over expected interactions, we found that the smaller chromosomes (I, III, VI, and IX) exhibit a higher probability to contact each other, while only three pairs of larger chromosomes (IV and VII, IV and XII, and IV and XV) displayed relatively high contact probabilities.

Clustering and receiver operating curve (ROC) analysis are two tools that can be employed to study structural features of chromosomes based on the information of chromatin interactions [26]. For example, by applying a hierarchical average-link clustering algorithm, we observed that early-firing DNA replication origins clustered into at least two discrete regions (Fig. 5D), which was recently demonstrated to be mediated by Forkhead transcription factors [35].

In summary, by comprehensively mapping the chromatin interactions using our method, we were able to generate a map at kilobase resolution of the haploid budding yeast genome, which recapitulates the well-known organizational features of the genome, including the Rab1 configuration, centromere clustering, telomere pairing, and clustering of the tRNA genes. We also revealed some new structural features of the yeast genome, such as the

unique conformation of chromosome XII and clustering of the early-firing DNA replication origins [26].

4. Limitations and alternative methods

Despite the successful application of our method to characterize the budding yeast genome, transitioning to diploid mammalian genomes requires several technical issues to be considered. First, in a diploid mammalian genome, each chromosome has a homologous partner, and it is not clear whether the two homologous chromosomes interact with other chromosomes in the same way or not. Hence, it might be important to distinguish the chromatin interactions involving any given chromosome from those of its homologous partner. This issue might be addressed by using single-nucleotide polymorphism analysis to distinguish homologous chromosomes. Second, the genomes of mammals are much larger than in yeast, and much more sequencing will be required to obtain sufficient long-range chromosomal interactions to build a high-resolution map. Using our current method, the long-range interactions (non-adjacent intra-chromosomal (≥ 20 kb) and inter-chromosomal ligations) only account for about 20% of the total ligation products in a library. The percentage of the short-range interactions (non-adjacent intra-chromosomal (<20 kb)) is also about 20%, while various types of background noise (see Section 2.3.2) account for up to 60% of the total ligation products in a

library. Therefore, achieving a high signal to noise ratio minimizes an already enormous sequencing effort. Third, in our current method the paired-end tags, representing the two interacting genomic fragments, are produced by EcoP15I digestion and are only 25–27 bp in length. Since the genome size and the size of repetitive regions are much larger in mammalian cells, the mapping efficiency of sequence reads, in particular, the mappability for repetitive regions could be relatively low. The strategy of building a reduced custom reference genome as described in Section 3.1 will dramatically improve the mapping speed but might not be sufficient to significantly improve the mappability for repetitive genomic regions. Hence, more sophisticated computational approaches, such as the approach employed by the CHIA-PET technology (in which the paired end tag is only 18 bp [16,36,37]), might be required. Fourth, like all other 3C-based methods, our method is associated with various types of experimental biases. For example, we observed restriction enzyme site-dependent differences in ligation efficiency and mappability differences of RE1 fragments [26]. In a mammalian genome, the various biases will become even more severe [38]. Hence, mammalian chromatin interaction data produced by our method may also be normalized as described in Yaffe et al [38]. Fifth, our method contains more experimental steps than Hi-C [19,21] and TCC [18], although it contains fewer experimental steps involving chromatin fragments, which are much more difficult to deal with than DNA fragments. In these regards, our method

Table 2
reagents.

Reagents name	Company	Catalogue number	Comments
37% (vol/vol) formaldehyde	Sigma-aldrich	F8775	Step 5
Glycine	Sigma-aldrich	50046	Step 6
NeBuffer 2	NEB	B7002S	Step 10
10% (wt/vol) SDS	Cellgro	46-040-CI	Step 12
10% (vol/vol) Triton X-100	Sigma-aldrich	T9284	Steps 13, 16
High concentration EcoRI	NEB	R0101M	Step 14
High concentration Hind III	NEB	R0104M	Step 14
1 kb DNA ladder	Fermentas	SM0311	Steps 15, 24, 50, 51
Agarose	Invitrogen	15510-027	Steps 15, 24, 50, 51
10 mg/ml BSA	NEB	B9001S	Steps 16, 33
T4 DNA ligase	Fermentas	EL0011	Steps 16, 27, 34
Proteinase K	Fermentas	EO0492	Step 18
3 M sodium acetate (pH 5.2)	Cellgro	46-033-CI	Steps 20, 26, 28, 30
Ethanol	Decon labs, Inc	2716	Steps 20, 26, 28, 30
GlycoBlue	Ambion	AM9516	Steps 20, 26, 28, 30
Iso-propanol	Sigma-aldrich	437522	Steps 20, 26, 28, 30
RNase A (DNase-free)	Sigma-aldrich	R6513	Step 21
UltraPure™ Phenol:Chloroform:Isoamyl Alcohol	Invitrogen	15593-031	Step 22
MseI	NEB	R0525M	Step 25
MspI	NEB	R0106M	Step 25
Qiaquick PCR purification kit	Qiagen	28106	Steps 26, 30, 32, 33, 35, 36, 38, 40
Plasmid-Safe™ ATP-Dependent DNase	Epicentre	E3105K	Steps 29, 39
FastDigest® Eco RI	Fermentas	FD0274	Step 31
FastDigest® Hind III	Fermentas	FD0504	Step 31
10× NEBuffer 3	NEB	B7003S	Steps 33, 41
EcoP15I	NEB	R0646L	Steps 33, 41
S-adenosylmethionine	NEB	B9003S	Step 33
Sinefungin	Sigma-aldrich	S8559-10MG	Step 41
100 mM dNTP Set	Invitrogen	10297-018	Steps 42, 48
Klenow	NEB	M0210	Step 42
PEG-8000	Sigma-aldrich	89510	Step 45
Quick Ligation™ Kit	NEB	M2200S	Step 45
Dynabeads® M-280 Streptavidin	Invitrogen	112-05D	Step 46
DNA Polymerase I (E. coli)	NEB	M0209S	Step 48
Phusion HF DNA polymerase	NEB	M0530S	Steps 49, 50
100 bp DNA ladder	Fermentas	SM0243	Steps 49, 50
50 bp DNA ladder	NEB	N3236S	Steps 49, 50
Qiaquick Gel Extract Kit	Qiagen	28706	Step 50
Sorbitol	Sigma-aldrich	S6021	Spheroplast buffer
ATP	Sigma-aldrich	A9187	T4 DNA ligase buffer
DTT	Bio-rad	161-0610	T4 DNA ligase Buffer
0.5 M EDTA	Cellgro	46-034-CI	TE Buffer

is best suited for characterizing the topologies of relatively simple genomes at high resolution (kb) and for the identification of functionally relevant (statistically significant) long-range chromatin interactions between distant genomic elements (such as promoter-enhancer interactions) on a whole-genome scale.

5. Appendixes

5.1. Material

Reagents can be found in Table 2.

5.2. Buffer

Spheroplast buffer
1 M sorbitol
100 mM potassium phosphate (PH 7.5)

10× T4 DNA ligase Buffer

500 mM Tris-HCl, PH7.5
100 mM DTT
100 mM MgCl₂
10 mM ATP
TE Buffer (PH 8.0)

10 mM Tris-HCl (PH 8.0)
1 mM EDTA

5.3. Primer

HindIII-ECOP 15I-Adaptor-F 5'/5Phos/AGC TCT GCT GTA C 3'
HindIII-ECOP 15I-Adaptor-R 5'/5Phos/ACA GCA G 3'
Eco RI-ECOP 15I-Adaptor-F 5'/5Phos/AAT TTC TGC TGT AC 3'
Eco RI-ECOP 15I-Adaptor-R 5'/5Phos/ACA GCA GA 3'
Biotin-internal adaptor-F 5'/5Phos/CGTACAT(Bio)CCGCCTTGG
CCGT 3'
Biotin-internal adaptor-R 5'/5Phos/GGCCAAGGCGGATGTACGGT
3'

Illumina-PE-Adaptor-A-F
5' AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA
CAC GAC GCT CTT CCG ATC T 3'

Illumina PE-Adaptor-A-R
5' AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT AGA TCT
CGG TGG TCG CCG TAT CAT TTT 3'

Illumina-PE-Adaptor-B-F
5' CAA GCA GAA GAC GGC ATA CGA GAT CGG TCT CGG CAT TCC
TGC TGA ACC GCT CTT CCG ATC T 3'

Illumina-PE-Adaptor-B-R
5' AGA TCG GAA GAG CGG TTC AGC AGG AAT GCC GAG ACC GAT
CTC GTA TGC CGT CTT CTG CTT GTT 3'

Illumina-lib-PCR-F 5' AAT GAT ACG GCG ACC ACC GAG ATC TAC
3'

Illumina-lib-PCR-R 5' CAA GCA GAA GAC GGC ATA CGA GAT
CGG 3'

5.4. Equipment

Gel Doc™ XR + System (Bio-Rad)
Microcentrifuge, for example, Sorvall Legend micro17R (for
1.5 ml tubes)

Centrifuge for 15 and 50 ml Falcon tubes, for example Sorvall
Legend RT

Gene Amp PCR system 9700 (ABI)

7900 HT Fast Real-time PCR system (ABI)

Nanodrop-1000 spectrophotometer (Thermo Scientific).

Acknowledgements

Supported by NIH grants P01GM081619 (CBA), P41RR0011823 (WSN), and the Howard Hughes Medical Institute (SF). We thank Ferhat Ay for his assistance in preparing Fig. 4.

References

- [1] T. Misteli, Cell 128 (2007) 787–800.
- [2] J.R. Daban, Micron 42 (2011) 733–750.
- [3] B. Huang, H. Babcock, X. Zhuang, Cell 143 (2010) 1047–1058.
- [4] L.M. Rapkin, D.R. Ansel, R. Li, D.P. Bazett-Jones, Micron 43 (2012) 150–158.
- [5] A. Bolzer, G. Kreth, I. Solovei, D. Koehler, K. Saracoglu, C. Fauth, S. Muller, R. Eils, C. Cremer, M.R. Speicher, T. Cremer, PLoS Biol. 3 (2005) e157.
- [6] T. Tsukamoto, N. Hashiguchi, S.M. Janicki, T. Tumber, A.S. Belmont, D.L. Spector, Nat. Cell Biol. 2 (2000) 871–878.
- [7] E. de Wit, W. de Laat, Gene Dev. 26 (2012) 11–24.
- [8] B. van Steensel, J. Dekker, Nat. Biotechnol. 28 (2010) 1089–1095.
- [9] M. Capelson, Y. Liang, R. Schulte, W. Mair, U. Wagner, M.W. Hetzer, Cell 140 (2010) 372–383.
- [10] L. Guelen, L. Pagie, E. Brasset, W. Meuleman, M.B. Faza, W. Talhout, B.H. Eussen, A. de Klein, L. Wessels, W. de Laat, B. van Steensel, Nature 453 (2008) 948–951.
- [11] B. Kalverda, H. Pickersgill, V.V. Shloma, M. Fornerod, Cell 140 (2010) 360–371.
- [12] D. Peric-Hupkes, W. Meuleman, L. Pagie, S.W. Bruggeman, I. Solovei, W. Brugman, S. Graf, P. Flicke, R.M. Kerkhoven, M. van Lohuizen, M. Reinders, L. Wessels, B. van Steensel, Mol. Cell 38 (2010) 603–613.
- [13] D. Carter, L. Chakalova, C.S. Osborne, Y.F. Dai, P. Fraser, Nat. Genet. 32 (2002) 623–626.
- [14] J. Dekker, K. Rippe, M. Dekker, N. Kleckner, Science 295 (2002) 1306–1311.
- [15] J. Dostie, T.A. Richmond, R.A. Arnaout, R.R. Selzer, W.L. Lee, T.A. Honan, E.D. Rubio, A. Krumm, J. Lamb, C. Nusbaum, R.D. Green, J. Dekker, Genome Res. 16 (2006) 1299–1309.
- [16] M.J. Fullwood, M.H. Liu, Y.F. Pan, J. Liu, H. Xu, Y.B. Mohamed, Y.L. Orlov, S. Velkov, A. Ho, P.H. Mei, E.G. Chew, P.Y. Huang, W.J. Welboren, Y. Han, H.S. Ooi, P.N. Ariyaratne, V.B. Vega, Y. Luo, P.Y. Tan, P.Y. Choy, K.D. Wansa, B. Zhao, K.S. Lim, S.C. Leow, J.S. Yow, R. Joseph, H. Li, K.V. Desai, J.S. Thomsen, Y.K. Lee, R.K. Karuturi, T. Herve, G. Bourque, H.G. Stunnenberg, X. Ruan, V. Cacheux-Rataboul, W.K. Sung, E.T. Liu, C.L. Wei, E. Cheung, Y. Ruan, Nature 462 (2009) 58–64.
- [17] S. Horike, S. Cai, M. Miyano, J.F. Cheng, T. Kohwi-Shigematsu, Nat. Genet. 37 (2005) 31–40.
- [18] R. Kalhor, H. Tjong, N. Jayatilaka, F. Alber, L. Chen, Nat. Biotechnol. 30 (2012) 90–98.
- [19] E. Lieberman-Aiden, N.L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B.R. Lajoie, P.J. Sabo, M.O. Dorschner, R. Sandstrom, B. Bernstein, M.A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L.A. Mirny, E.S. Lander, J. Dekker, Science 326 (2009) 289–293.
- [20] S. Schoenfelder, T. Sexton, L. Chakalova, N.F. Cope, A. Horton, S. Andrews, S. Kurukuti, J.A. Mitchell, D. Umlauf, D.S. Dimitrova, C.H. Eski, Y. Luo, C.L. Wei, Y. Ruan, J.J. Bieker, P. Fraser, Nat. Genet. 42 (2010) 53–61.
- [21] T. Sexton, E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc, M. Hoichman, H. Parrinello, A. Tanay, G. Cavalli, Cell 148 (2012) 458–472.
- [22] M. Simonis, P. Klous, E. Splinter, Y. Moshkin, R. Willemsen, E. de Wit, B. van Steensel, W. de Laat, Nat. Genet. 38 (2006) 1348–1354.
- [23] H. Tanizawa, O. Iwasaki, A. Tanaka, J.R. Capizzi, P. Wickramasinghe, M. Lee, Z. Fu, K. Noma, Nucleic Acids Res. 38 (2010) 8164–8177.
- [24] V.K. Tiwari, L. Cope, K.M. McGarvey, J.E. Ohm, S.B. Baylin, Genome Res. 18 (2008) 1171–1179.
- [25] Z. Zhao, G. Tavoosidana, M. Sjölinder, A. Gondor, P. Mariano, S. Wang, C. Kanduri, M. Lezcano, K.S. Sandhu, U. Singh, V. Pant, V. Tiwari, S. Kurukuti, R. Ohlsson, Nat. Genet. 38 (2006) 1341–1347.
- [26] Z. Duan, M. Andronescu, K. Schutz, S. McIlwain, Y.J. Kim, C. Lee, J. Shendure, S. Fields, C.A. Blau, W.S. Noble, Nature 465 (2010) 363–367.
- [27] I. Maccallum, D. Przybylski, S. Gnerre, J. Burton, I. Shlyakhter, A. Gnirke, J. Malek, K. McKernan, S. Ranade, T.P. Shea, L. Williams, S. Young, C. Nusbaum, D.B. Jaffe, Genome Biol. 10 (2009) R103.
- [28] J. Dekker, Nat. Methods 3 (2006) 17–21.
- [29] A. Miele, N. Gheldof, T.M. Tabuchi, J. Dostie, J. Dekker, Curr Protoc Mol Biol, Chapter 21 (2006) Unit 21 11.
- [30] M. Simonis, J. Kooren, W. de Laat, Nat. Methods 4 (2007) 895–901.
- [31] M.J. Fullwood, Y. Ruan, J. Cell. Biochem. 107 (2009) 30–39.
- [32] H. Hagege, P. Klous, C. Braem, E. Splinter, J. Dekker, G. Cathala, W. de Laat, T. Forne, Nat. Protoc. 2 (2007) 1722–1733.
- [33] J.R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J.S. Liu, B. Ren, Nature 485 (2012) 376–380.

- [34] E.P. Nora, B.R. Lajoie, E.G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N.L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Bluthgen, J. Dekker, E. Heard, *Nature* 485 (2012) 381–385.
- [35] S.R. Knott, J.M. Peace, A.Z. Ostrow, Y. Gan, A.E. Rex, C.J. Viggiani, S. Tavare, O.M. Aparicio, *Cell* 148 (2012) 99–111.
- [36] L. Handoko, H. Xu, G. Li, C.Y. Ngan, E. Chew, M. Schnapp, C.W. Lee, C. Ye, J.L. Ping, F. Mulawadi, E. Wong, J. Sheng, Y. Zhang, T. Poh, C.S. Chan, G. Kunarso, A. Shahab, G. Bourque, V. Cacheux-Rataboul, W.K. Sung, Y. Ruan, C.L. Wei, *Nat. Genet.* 43 (2011) 630–638.
- [37] G. Li, X. Ruan, R.K. Auerbach, K.S. Sandhu, M. Zheng, P. Wang, H.M. Poh, Y. Goh, J. Lim, J. Zhang, H.S. Sim, S.Q. Peh, F.H. Mulawadi, C.T. Ong, Y.L. Orlov, S. Hong, Z. Zhang, S. Landt, D. Raha, G. Euskirchen, C.L. Wei, W. Ge, H. Wang, C. Davis, K.I. Fisher-Aylor, A. Mortazavi, M. Gerstein, T. Gingeras, B. Wold, Y. Sun, M.J. Fullwood, E. Cheung, E. Liu, W.K. Sung, M. Snyder, Y. Ruan, *Cell* 148 (2012) 84–98.
- [38] E. Yaffe, A. Tanay, *Nat. Genet.* 43 (2011) 1059–1065.