



**HAL**  
open science

# Determination of High-Resolution 3D Chromatin Organization Using Circular Chromosome Conformation Capture (4C-seq)

Mélo<sup>dy</sup> Matelot, Daan Noordermeer

► **To cite this version:**

Mélo<sup>dy</sup> Matelot, Daan Noordermeer. Determination of High-Resolution 3D Chromatin Organization Using Circular Chromosome Conformation Capture (4C-seq). Polycomb Group Proteins: Methods and Protocols, 1480, Springer Science+Business Media New York, pp.223-241, 2016, 9781493963782. 10.1007/978-1-4939-6380-5\_20 . hal-01412298

**HAL Id: hal-01412298**

**<https://hal.science/hal-01412298>**

Submitted on 8 Dec 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Determination of High-Resolution 3D Chromatin Organization Using Circular Chromosome Conformation Capture (4C-seq)

Mélody Matelot and Daan Noordermeer\*

## Abstract

3D chromatin organization is essential for many aspects of transcriptional regulation. Circular Chromosome Conformation Capture followed by Illumina sequencing (4C-seq) is among the most powerful techniques to determine 3D chromatin organization. 4C-seq, like other modifications of the original 3C technique, uses the principle of “proximity ligation” to identify and quantify ten thousands of genomic interactions at a kilobase scale in a single experiment for predefined loci in the genome.

In this chapter we focus on the experimental steps in the 4C-seq protocol, providing detailed descriptions on the preparation of cells, the construction of the circularized 3C library and the generation of the Illumina high throughput sequencing library. This protocol is particularly suited for the use of mammalian tissue samples, but can be used with minimal changes on circulating cells and cell lines from other sources as well. In the final section of this chapter, we provide a brief overview of data analysis approaches, accompanied by links to publicly available analysis tools.

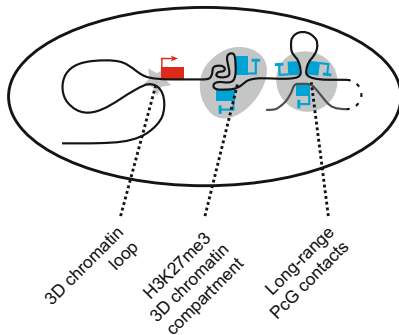
**Key words** Circular Chromosome Conformation Capture, 4C-seq, 3D chromatin organization, DNA interactions, Chromatin loops, Chromatin compartmentalization, Nuclear organization, High-throughput sequencing

---

## 1 Introduction

3D chromatin organization is an essential component of transcriptional regulation [1, 2]. The function of enhancers and insulators requires the formation of 3D chromatin loops ([3, 4] and Fig. 1, left), whereas the recently discovered Topological Associated Domains (TADs) appear to spatially structure and separate gene regulatory domains [5–7]. Genomic regions that bind the repressive Polycomb group proteins (PcG-proteins) and that carry the associated H3K27me3 histone mark form specialized 3D chromatin structures as well. The first 3C studies (Chromosome Conformation Capture) in human and *Drosophila* cells identified complex loop structures at the GATA-4 gene (in cultured human Tera-2

## In-vivo 3D chromatin organization:



## Proximity ligation:

### Formaldehyde cross-linking of chromatin:



### Enzymatic digestion of cross-linked chromatin:



### Proximity ligation:



**Fig. 1** The principle of proximity ligation for the detection of 3D chromatin interactions. Chromatin fragments that are in spatial proximity are cross-linked in vivo using formaldehyde (red bars). After enzymatic digestion, only cross-linked fragments are kept together. Proximity ligation circularizes those fragments that are together due to their shared cross-links

embryonic carcinoma cells, [8]) and the bithorax complex (BX-C in *Drosophila* embryonic cells, [9]). More recent 4C-seq studies in mouse embryos reported that the repressed *Hox* gene clusters form local 3D chromatin compartments. In these dynamic 3D compartments the H3K27me3 marked chromatin clusters together, following the temporal and spatial repressed state of the *Hox* genes ([10, 11] and Fig. 1, left). Moreover, in mouse and *Drosophila* cells, PcG targets form long-range contacts among each other despite being separated by many megabases on the same chromosome or being located on different chromosomes ([10, 12–14] and Fig. 1, left). Multiple specialized and 3D chromatin structures are therefore dynamically associated with PcG-mediated repression in mammalian and insect cell systems.

Among the most used techniques to study 3D genome organization at high resolution is the 4C-seq technique (Circular Chromosome Conformation Capture followed by Illumina sequencing). 4C-seq is a genome-wide adaptation to the original 3C technique [15] that was originally developed for readout with microarrays (4C, [16]). In a 4C-seq experiment, the genome-wide 3D interactions of preselected genomic sites (the so-called “viewpoints”) are identified and quantified. Due to this focus on individual viewpoints, typically several tens of thousands of interacting sites can be identified in a single experiment, which makes 4C the most comprehensive approach for individual genomic sites available (as compared to other genome-wide adaptations to the 3C approach like 5C, HiC, and ChIA-PET). The combination of 4C with Illumina sequencing allows sequencing of up to 20 viewpoints at a time, thereby considerably improving the throughput of the experiments.

4C-seq, like all 3C-based assays, relies on “proximity ligation” to detect contacts between DNA fragments (Fig. 1, right). First, 3D chromatin organization is fixed *in vivo* over short distance using formaldehyde cross-linking. Next, chromatin is fragmented using a restriction enzyme, only keeping fragments together that were cross-linked due to their initial spatial proximity. In a final step, the DNA is ligated under diluted conditions, thereby promoting ligation between fragments that are present in the same cross-linked complexes. Using the frequency of ligation between pairs of restriction fragments as readout, 3D chromatin interactions can next be determined. By interrogating the ligation events of a viewpoint, typically in around 100,000 cells, an average snapshot of 3D chromatin organization within the cell population can be obtained. Depending on the downstream bioinformatics analysis, local 3D organization, long-range interactions or differences in 3D organization between cell types or experimental conditions can be determined.

In this chapter, we provide a detailed description of the 4C-seq approach, with particular emphasis on the experimental procedures (*see* Fig. 2). This protocol generates a very high-resolution description of 3D chromatin organization for individual viewpoints, by using two sequential rounds of frequent cutting restriction enzymes (4 bp recognition sites, resulting in an average distance between informative restriction fragments of 1–1.5 kb). The first section of this protocol describes the preparation of cross-linked nuclei from tissue samples, which can be used with minor modifications for cultured or circulating cells as well (*see* **Notes 1** and **6**). In the second section, a circularized 3C library is generated using a first round of digestion and (proximity) ligation on cross-linked chromatin followed by a second round of digestion and ligation on naked DNA. In the third section, a 4C-seq library for Illumina sequencing is generated by PCR amplification of the circularized 3C library with viewpoint-specific inverse primers. In the final section, an overview of the data analysis strategy is provided, accompanied by links to recently published and publicly available resources that can be used for the bioinformatics analysis of 4C-seq data.

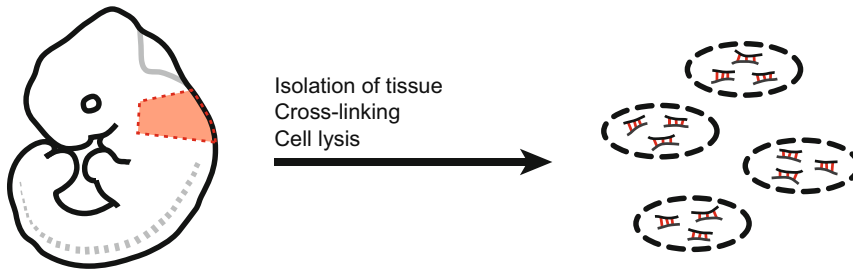
---

## 2 Materials

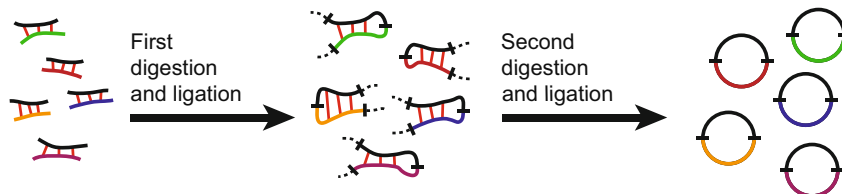
### 2.1 Isolation and Preparation of Tissue Samples

1. Isolation buffer: Phosphate buffered saline (PBS) (pH 7.2–7.6) supplemented with 10% filter-sterilized fetal bovine serum (FBS). Can be stored at 4 °C for 1 week.
2. 12.5% w/v collagenase solution: collagenase powder from *Clostridium histolyticum* dissolved in PBS.
3. Cell strainer: 35 µm cell strainer for round bottom tube or 40 µm cell strainer for 50 ml tube.

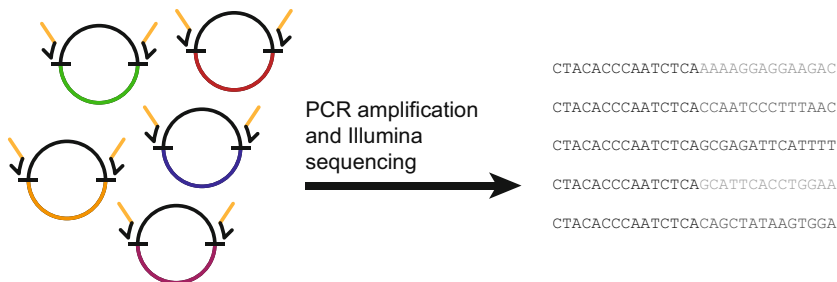
## Preparation of tissue samples:



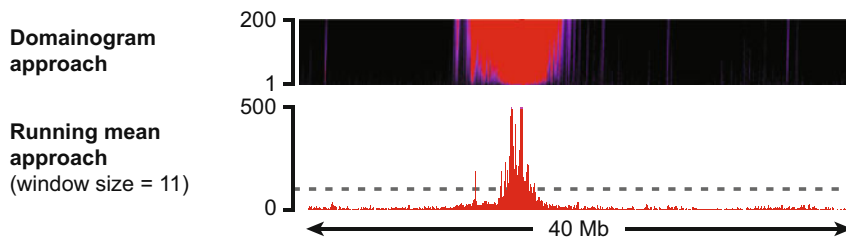
## Preparation of circularized 3C library:



## Preparation of 4C-seq library:



## Data analysis:



**Fig. 2** Schematic outline of the steps in the 4C-seq assay

- 37% w/v formaldehyde solution.
- 2% cross-linking solution: 650  $\mu$ l 37% formaldehyde solution, 11.35 ml isolation buffer. Prepare the solution fresh on the day of use. Keep at 20–25  $^{\circ}$ C (room temperature).
- 1 M glycine solution. Keep at 4  $^{\circ}$ C.
- Cell lysis buffer: 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5% v/v NP-40, 1% v/v Triton X-100, 1x

complete protease inhibitor cocktail, EDTA free. Prepare the solution on the day of use and keep at 4 °C.

Depending on the cell type, the cell lysis buffer may need to be optimized. *See Note 6* for other published lysis buffers.

## **2.2 Preparation of Circularized 3C Library**

1. 1.2× restriction buffer: 60 µl 10× restriction buffer (as supplied with restriction enzyme), 12 µl bovine serum albumin (BSA stock) (10 mg/ml stock), 428 µl ultrapure water.
2. 20 % sodium lauryl sulfate (SDS) solution.
3. 20 % Triton X-100 solution.
4. Selected restriction enzymes, preferably in high concentrated form.
5. 5 M NaCl solution.
6. 1 M Tris-HCl, pH 7.5 solution.
7. 0.5 M EDTA solution.
8. 20 mg/ml Proteinase K solution.
9. 10 mg/ml RNaseA solution.
10. Phenol-chloroform-IAA solution (25:24:1): 50 % v/v phenol, saturated with 10 mM Tris, pH 8.0 and 1 mM EDTA, 48 % v/v chloroform, 2 % v/v isoamyl alcohol (commercially available).
11. 2 M sodium acetate (NaAc), pH 5.6 solution.
12. 70 % and 100 % ethanol solutions.
13. 10× ligation buffer: 660 mM Tris-HCl, pH 7.5, 50 mM MgCl<sub>2</sub>, 50 mM dithiothreitol (DTT), and 10 mM adenosine triphosphate (ATP). 1 ml aliquots can be stored at -20 °C for 6 months.
14. T4 DNA ligase, preferably in high concentrated form.
15. PCR Purification Kit (e.g., Qiagen, prior to the first use, add four volumes of 100 % ethanol to wash buffer PE).

## **2.3 Optimization of PCR Conditions and Preparation of 4C Sequencing Libraries**

1. Expand Long Template PCR System (e.g., Roche life sciences).
2. 10 mM dNTP solution. Can be stored for 1 month at -20 °C. Repeated freeze-thaw cycles and prolonged storage at -20 °C can severely decrease PCR efficiency.
3. Locus specific inverse Forward and inverse Reverse primers, diluted at 100 µM (*see Note 4*).
4. PCR Purification Kit (e.g., Qiagen, additional PB binding buffer can be ordered separately).

## **2.4 Available Pipelines for 4C Data Analysis**

1. HTSstation: online data analysis and Python scripts; [17] and <http://htsstation.epfl.ch/>
2. 4Cseqpipe: R-package; [18] and [http://compgenomics.weizmann.ac.il/tanay/?page\\_id=367/](http://compgenomics.weizmann.ac.il/tanay/?page_id=367/)

3. FourCSeq: R/Bioconductor-package; [19] and <http://www.bioconductor.org/packages/release/bioc/html/FourCSeq.html>
4. Basic4Cseq: R/Bioconductor-package; <http://www.bioconductor.org/packages/release/bioc/html/Basic4Cseq.html>
5. r3Cseq: R/Bioconductor-package; [20] and <http://www.bioconductor.org/packages/release/bioc/html/r3Cseq.html>

---

## 3 Methods

See **Notes 1–4** and Fig. 3 for important guidelines on tissue quantities, cell types, choice of restriction enzymes and primer design that should be considered prior to starting the experiment.

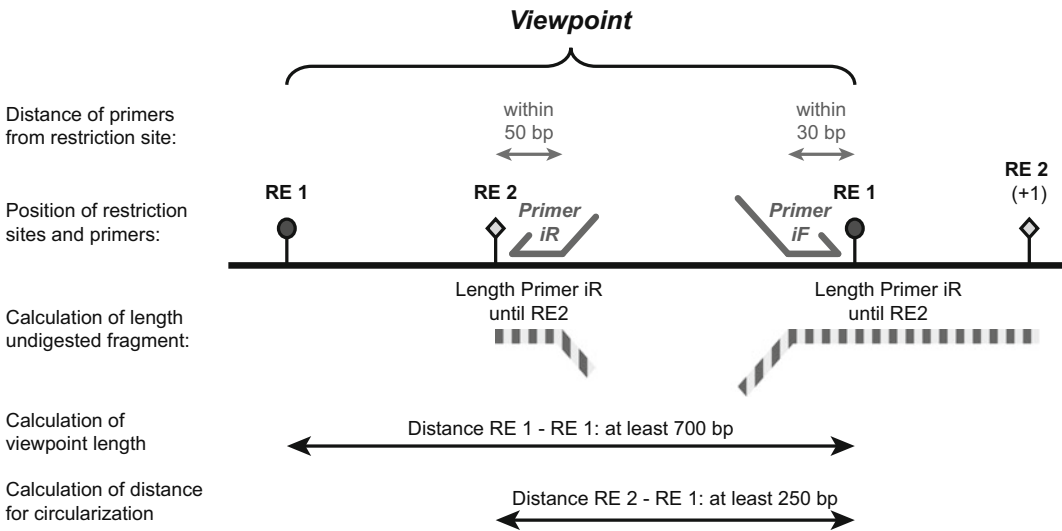
### 3.1 Isolation and Preparation of Tissue Samples

1. Pool dissected tissue fragments in 495  $\mu$ l of isolation buffer. Keep at 4 °C during dissection. See **Note 5** if samples of less than  $1 \times 10^7$  cells are used.
2. Add 5  $\mu$ l of 12.5% w/v collagenase solution (final concentration 0.125%) and incubate tissue fragments for 45 min at 37 °C in a shaker at 750 rpm. Dissociate cells by pipetting up and down several times with a blue tip.
3. Make single cell by forcing the solution through a cell strainer and transfer cells to a 15 ml conical tube. Add 9.5 ml of 2% cross-linking solution and incubate the cells for 10 min on a rotating wheel or rocking platform at 20–25 °C (room temperature).
4. Immediately transfer the tube with cells to ice. Quench the cross-link reaction by adding 1.43 ml of a cold 1 M Glycine solution. Centrifuge the cells for 8 min at  $225 \times g$ , 4 °C.
5. Remove the cross-linking solution. Resuspend the cells in 5 ml cold cell lysis buffer and incubate the cells for 10 min on ice followed by pipetting up and down several times with a blue tip. Centrifuge the nuclei for 5 min at  $400 \times g$ , 4 °C. See **Note 6**.
6. Remove 4.5 ml of the cell lysis buffer and resuspend the nuclei in the remaining 500  $\mu$ l of volume. Transfer the remaining volume to a 1.5 ml plastic micro tube and centrifuge the nuclei for 1 min at  $230 \times g$ , 4 °C.
7. Remove the remaining 500  $\mu$ l of the cell lysis buffer. At this stage, cells can either be frozen in liquid nitrogen and stored at –80 °C until further use, or the protocol can be immediately continued at Subheading 3.2.

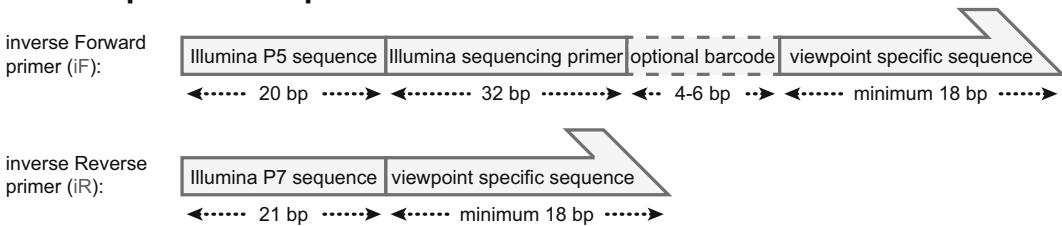
### 3.2 Preparation of Circularized 3C Library

Multiple preparations of cross-linked nuclei, as prepared in Subheading 3.1, can be pooled at this point. In this case, we advise to pool and wash the samples: add a total volume of 500  $\mu$ l 1.2 $\times$

## A. Design considerations for 4C-seq experiments:



## B. PCR primer components:



**Fig. 3** Design considerations for 4C-seq experiments and PCR primers. **(a)** Scheme with design considerations for 4C-seq experiments. A useful viewpoint should be of sufficient length (distance between the cut sites of the primary restriction enzyme RE 1) and should have sufficient distance between the cut sites of the primary (RE 1) and secondary (RE 2) restriction enzyme to allow circularization of the 4C-library. The inverse primers (iF and iR) should be located within the indicated distance next to the cut sites of the restriction enzymes. The length of the undigested fragment (see **Note 4**) can be calculated by adding up the length of both the iF and iR primers, including adapter sequences, up to their respective cut sites + the distance between RE 1 and the first downstream cut site of the secondary restriction enzyme RE 2 (+1) (total length of the striped bars). **(b)** Components of the inverse Forward (iF) and inverse Reverse (iR) primers. We standardly design the iF primer next to RE 1 and the iR primer next to RE 2

restriction buffer to the combined samples, pool the samples together in a single 1.5 ml plastic micro tube, centrifuge the nuclei for 1 min at  $230 \times g$  at 4 °C and remove the supernatant. Next, continue with **step 1** of this subheading.

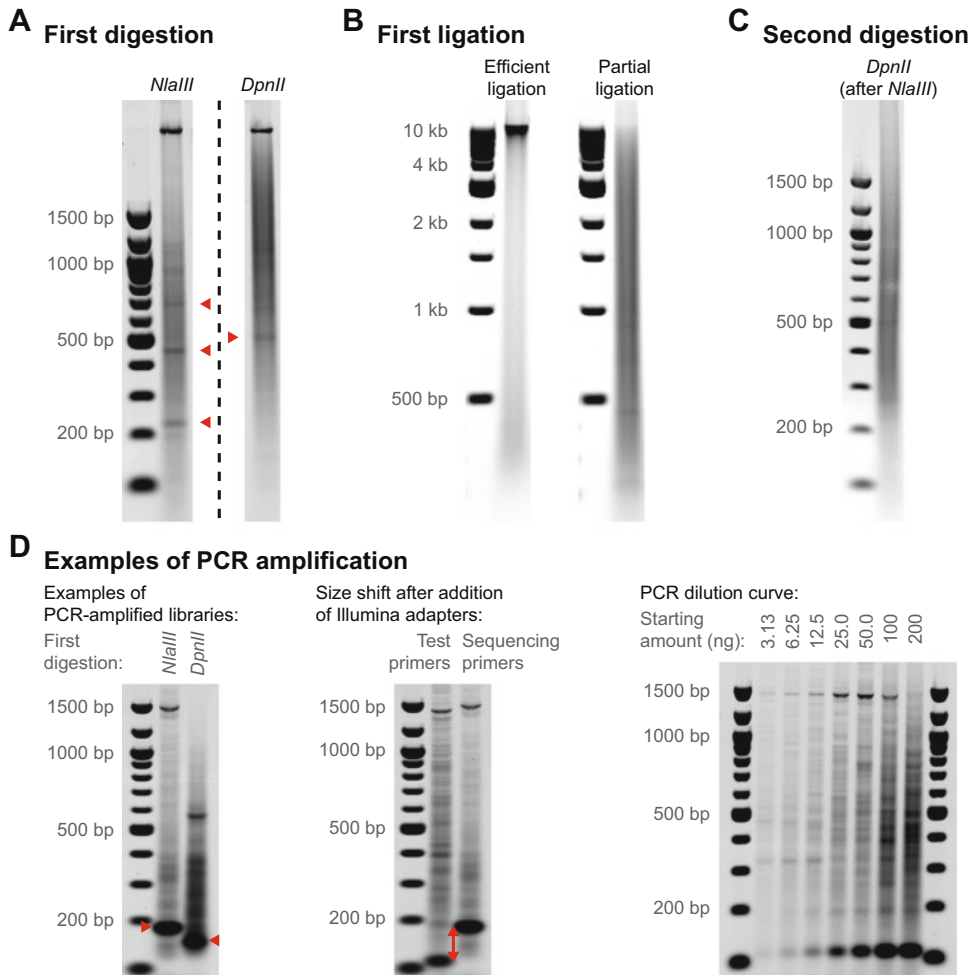
1. Take up nuclei in 500  $\mu$ l 1.2 $\times$  restriction buffer, place at 55 °C in a shaker at 750 rpm and immediately add 7.5  $\mu$ l of a 20% SDS solution (final concentration 0.3%). Incubate for no more than 10 min at 60 °C, followed by another 50 min at 37 °C in a shaker at 750 rpm. See **Note 7** for the appearance of the solution.



2. Sequester the SDS in the solution by adding 50  $\mu\text{l}$  of a 20% Triton X-100 solution (final concentration 2%) and incubate for 1 h at 37  $^{\circ}\text{C}$  in a shaker at 750 rpm. Optionally, a 5  $\mu\text{l}$  aliquot can be taken at this point as “undigested control”. See in **step 4** of this subheading how to revert cross-links and visualize this control.
3. Digest the cross-linked DNA by adding 400 Units of the selected restriction enzyme and incubate for 4–6 h at 37  $^{\circ}\text{C}$  in a shaker at 750 rpm. Add another 400 Units of the restriction enzyme and incubate overnight at 37  $^{\circ}\text{C}$  in a shaker at 750 rpm.
4. Prior to continuation, verify the efficiency of the DNA digestion. Take a 5  $\mu\text{l}$  aliquot as “digested control”. The optional “undigested control” from **step 2** should be added here as well. The remainder of the digested samples can be stored at 4  $^{\circ}\text{C}$  before continuing at **step 5**. Add 90  $\mu\text{l}$  ultrapure water and 5  $\mu\text{l}$  of a 5 M NaCl solution to the control(s) and incubate for 2 h at 65  $^{\circ}\text{C}$  in a shaker at 750 rpm. Lower the temperature to 45  $^{\circ}\text{C}$ , add 2  $\mu\text{l}$  of a 1 M Tris-HCl, pH 7.5 solution, 2  $\mu\text{l}$  of a 0.5 M EDTA solution and 2  $\mu\text{l}$  of a Proteinase K solution and incubate for 2 h at 45  $^{\circ}\text{C}$  in a shaker at 750 rpm. Lower the temperature to 37  $^{\circ}\text{C}$ , add 2  $\mu\text{l}$  of an RNaseA solution and incubate for 30 min at 37  $^{\circ}\text{C}$  in a shaker at 750 rpm. Transfer sample to a safety cabinet, add 120  $\mu\text{l}$  of a phenol-chloroform-IAA solution and shake vigorously. Centrifuge the sample for 15 min at maximum speed (typically around 20,000 $\times g$ ) at 20–25  $^{\circ}\text{C}$ . Transfer aqueous phase to a new plastic micro tube, add 12  $\mu\text{l}$  of a 2 M NaAc solution, add 300  $\mu\text{l}$  of 100% ethanol and centrifuge for 30 min at maximum speed at 4  $^{\circ}\text{C}$ . Remove the supernatant, add 200  $\mu\text{l}$  of a 70% ethanol solution and centrifuge for 10 min at maximum speed at 4  $^{\circ}\text{C}$ . Briefly air-dry sample and dissolve the sample in 20  $\mu\text{l}$  ultrapure water. Visualize the sample on a 1.5% agarose gel. The sample should run as a large smear with highest intensity between 500 and 1000 bp (*see* Fig. 4a). If the chromatin shows good digestion, the remaining sample can be further processed (continue to **step 5**). If chromatin is partially digested, showing a smear with highest intensity in the range of 1–5 kb, **step 3** of this subheading should be repeated with the remaining sample. If no or minimal digestion has occurred, tissue preparation, cell lysis or the choice of restriction enzyme should be optimized (*see* **Notes 1, 3 and 6**).
5. Take the remainder of the digested nuclei (between 500 and 600  $\mu\text{l}$ ) and inactivate the remaining restriction enzyme by adding 40  $\mu\text{l}$  of a 20% SDS solution (final concentration around 1.5%) and an incubation of no more than 20 min at 65  $^{\circ}\text{C}$ .
6. Transfer the sample to a 50 ml conical tube and transfer to a 37  $^{\circ}\text{C}$  water bath. Add 6.13 ml 1.15 $\times$  ligation buffer (pre-mixed: 710  $\mu\text{l}$  10 $\times$  ligation buffer + 5.42 ml ultrapure water),

add 375  $\mu\text{l}$  of a 20% Triton X-100 solution to sequester the SDS and incubate for 1 h at 37 °C, with occasional shaking.

7. Transfer the sample to a 16 °C water bath and incubate for 15 min. Ligate the diluted cross-linked DNA fragments by adding 100 Units of T4 DNA ligase and incubation at 16 °C for 4 h, followed by 30 min at 20–25 °C.
8. De-cross-link the sample by adding 15  $\mu\text{l}$  of a Proteinase K solution and overnight incubation in a water bath at 65 °C.
9. Transfer the sample to a 37 °C water bath, add 30  $\mu\text{l}$  of RNaseA solution and incubate for 45 min at 37 °C. Transfer the sample to a safety cabinet, add 7 ml of a phenol–chloroform–IAA solution, shake vigorously and centrifuge for 15 min at 3200  $\times g$  at 20–25 °C. Transfer aqueous phase to a new 50 ml conical tube and add 7 ml of ultrapure water, 1.5 ml of a 2 M NaAc solution, and 35 ml of 100% ethanol. Mix gently and place at –80 °C for at least 2 h. Centrifuge for 45 min at 3200  $\times g$  at 4 °C. A relatively large pellet will be visible that consists mainly of salts from the ligation buffer (*see Note 8*). Remove the supernatant, add 25 ml of a 70% ethanol solution and centrifuge for 15 min at 3200  $\times g$  at 4 °C. Remove the supernatant, dry for 1 h at 55 °C, transfer the sample to a 37 °C water bath and dissolve the pellet by adding 150  $\mu\text{l}$  of a 10 mM Tris–HCl, pH 7.5 solution and 1 h incubation combined with occasional tapping of the tube. After a 1-h incubation, the pellet should have fully dissolved.
10. Transfer the solution to a 1.5 ml plastic micro tube. Measure 2  $\mu\text{l}$  of the solution using a dye-incorporation based method (e.g., Qubit, *see Note 8*). When starting with  $1 \times 10^7$  cells, the remaining amount of DNA should typically be around 40  $\mu\text{g}$ . Run 500 ng of the sample on a 1.5% gel to confirm the efficiency of ligations (*see Fig. 4b*). The sample may be stored at –20 °C for several weeks at this stage.
11. Digest the sample with the second restriction enzyme (*see Note 3*) by diluting the DNA to a concentration of 100 ng/ $\mu\text{l}$  in the appropriate 1 $\times$  restriction buffer (supplemented with BSA, if required) and add 1 Unit of restriction enzyme/ $\mu\text{g}$  of DNA. Digest the DNA overnight at 37 °C in a shaker at 750 rpm.
12. Inactivate the restriction enzyme by 20 min incubation at 65 °C. Add an equal volume of a phenol–chloroform–IAA solution, shake vigorously and centrifuge the sample for 15 min at maximum speed (typically around 20,000  $\times g$ ) at 20–25 °C. Transfer aqueous phase to a new plastic micro tube, add 1/10 volume of a 2 M NaAc solution, add 2.5 volumes of 100% ethanol and place tube for 30 min at –80 °C. Centrifuge for 30 min at maximum speed at 4 °C, remove the supernatant,



**Fig. 4** Visualization of cross-linked chromatin and DNA at different stages of the 4C-seq procedure. (a) Digestion of cross-linked chromatin (2% formaldehyde, 10 min) with *NlaIII* (left) and *DpnII* (right). Digestion of cross-linked material is not 100% efficient, which results in a smear with an average length that is larger than randomly predicted. In sufficiently digested samples, within the large smear several specific products that stem from repeated sequences can be observed (red arrow heads for mouse genomic DNA). (b) Ligation of cross-linked chromatin (initially digested with *NlaIII*). Left: highly efficient ligation; right: partial ligation. Notice that specific bands that were visible after digestion have disappeared. In our hands, both efficiently and partially ligated samples can give good downstream results. (c) Second digestion of DNA, after de-crosslinking. A smear should be visible with a large distance distribution and no specific products. (d) PCR amplification of 4C-seq library with locus specific primers. After PCR amplification, a smear should be visible with fragment lengths that extend beyond 1 kb. Left: examples of PCR-amplified 4C-seq libraries first digested by *NlaIII* and *DpnII* (iF and iR sequencing primers). Undigested bands, which are specific to each primer set, are indicated by arrowheads. Middle: shift in the size of the undigested fragment upon addition of Illumina adapters (iF and iR test primers versus iF and iR sequencing primers, see Note 4). Right: titration to find the optimal concentrations for PCR amplification (iF and iR test primers). Too low concentrations result in poor yield, whereas overloading of the PCR results in an unwanted shift towards smaller fragments. Linearity here is observed in lanes 1–5 (3.13–50 ng per PCR reaction), after which the reaction becomes overloaded. The optimal concentration in this example is therefore 50 ng per PCR reaction. All samples were visualized on 1.5% agarose gels with a 100 bp or 1 kb ladder as reference (sizes of bands indicated in red). Primer sequences in Note 4

add 1 ml of a 70% ethanol solution and centrifuge for 10 min at maximum speed at 4 °C. Briefly air-dry sample and dissolve the sample in 100 µl ultrapure water. Verify the efficiency of the digestion by visualizing 2.5 µl of the sample on a 1.5% agarose gel (*see* Fig. 4c).

13. Subsequently, the final circularized 3C-library is generated by re-ligation under diluted conditions. Transfer the sample to a 50 ml conical tube, add 12.5 ml ultrapure water, add 1.4 ml 10× ligation buffer and transfer the tube to a 16 °C water bath. Add 200 Units of T4 DNA ligase and incubate for 4 h at 16 °C followed by 30 min at 20–25 °C.
14. Transfer the sample to a safety cabinet, add 14 ml of a phenol–chloroform–IAA solution, shake vigorously and centrifuge for 15 min at 3200×*g* at 20–25 °C. Transfer the aqueous phase to a new 50 ml conical tube, add 14 ml of ultrapure water, 2.8 ml of a 2 M NaAc solution, mix gently, and divide the solution equally over two 50 ml conical tubes. Add 35 ml of 100% ethanol to each tube, mix gently and place at –80 °C for at least 2 h (or overnight). Centrifuge for 45 min at 3200×*g* at 4 °C. In both tubes, a relatively large pellet will be visible (*see* **Note 8**). Remove the supernatant, add 25 ml of a 70% ethanol solution to both tubes and centrifuge for 15 min at 3200×*g* at 4 °C. Remove as much as possible of the supernatant, dry for 1 h at 55 °C, transfer the tubes to a 37 °C water bath and dissolve the pellets by adding 200 µl of a 10 mM Tris–HCl, pH 7.5 solution to each tube and 1 h incubation combined with occasional tapping of the tube. After a 1-h incubation, the pellet should have fully dissolved.
15. Clean the library by addition of 1 ml of Qiagen PB loading buffer to each tube and mix by pipetting up and down a few times. Load 4 QIAquick Spin Columns with 600 µl each of the solution and centrifuge for 1 min at maximum speed (typically around 20,000×*g*) at 20–25 °C. Remove the flow through, add 600 µl wash buffer PE and centrifuge for 1 min at maximum speed at 20–25 °C. Remove the last traces of wash buffer by transferring the spin columns to new plastic 1.5 ml micro tubes and 1 min centrifugation at maximum speed at 20–25 °C. Elute the samples by transferring the spin columns to new plastic 1.5 ml micro tubes, addition of 40 µl of a 10 mM Tris–HCl solution to each Spin column, 1 min incubation at 20–25 °C and 1 min centrifugation at maximum speed at 20–25 °C. Pool all samples in a single 1.5 ml plastic micro tube and measure 2 µl of the solution using a dye-incorporation based method (e.g., Qubit, *see* **Note 8**). When starting with  $1 \times 10^7$  cells, the DNA concentration should typically be around 100 ng/µl. At this stage, the sample can be stored long-term at –20 °C.

### 3.3 Optimization of PCR Conditions and Preparation of 4C-seq Libraries

1. First, the optimal amount of the circularized 3C library for PCR amplification needs to be determined. This amount depends on the concentration of the circular DNA and the amount of salt contamination in the library. To determine the optimal concentration, a dilution curve with the following amounts of DNA/50  $\mu$ l PCR reaction should be run: 12.5, 25, 50 and 100 ng. The total volume of the circularized 3C library that is added to each PCR reaction should not exceed 2  $\mu$ l. If possible, use a confirmed primer set (*see Note 4*). Primers with added Illumina adapter sequences are not required at this step.

Components of PCR reaction:

- 5  $\mu$ l 10 $\times$  Buffer 1 (Roche Expand Long Template PCR System)
- 1  $\mu$ l 10 mM dNTP solution
- 0.5  $\mu$ l 100  $\mu$ M iF primer (50 pmol final concentration)
- 0.5  $\mu$ l 100  $\mu$ M iR primer (50 pmol final concentration)
- Variable amount of circularized 3C library
- Ultrapure water up to 49.25  $\mu$ l
- 0.75  $\mu$ l DNA Polymerase mix (e.g., 5 U/ $\mu$ l, Roche Expand Long Template PCR System)

PCR program:

- 1 $\times$  (2 min at 94  $^{\circ}$ C)
- 30 $\times$  (15 s at 94  $^{\circ}$ C, 1 min at 55  $^{\circ}$ C; 3 min at 68  $^{\circ}$ C)
- 1 $\times$  (7 min at 68  $^{\circ}$ C)

After PCR, visualize 20  $\mu$ l of the material on a 1.5% agarose gel. The optimal concentration is the sample where the intensity of the smear is still linearly increased, where the size range in PCR products is maximal and where minimal primer dimers are detected (*see Fig. 4d, right*).

2. Prepare as many PCR reactions as necessary to amplify a total amount of 1  $\mu$ g circularized 3C library, containing iF and iR primers with Illumina sequences (*see Note 4* for design strategy). Use the same PCR program as mentioned in **step 1**.
3. After PCR, pool products from the same viewpoint, add five volumes of Qiagen PB loading buffer and load two QIAquick Spin Columns with the PCR product (columns may need to be loaded multiple times). After each loading step, centrifuge for 1 min at maximum speed (typically around 20,000 $\times g$ ) at 20–25  $^{\circ}$ C and remove the flow through. After loading, add 600  $\mu$ l wash buffer PE and centrifuge for 1 min at maximum speed at 20–25  $^{\circ}$ C. Remove the last traces of wash buffer by transferring the spin columns to new plastic 1.5 ml micro tubes and 1 min centrifugation at maximum speed at 20–25  $^{\circ}$ C. Elute

the samples by transferring the spin columns to new plastic 1.5 ml micro tubes, add 50  $\mu$ l of ultrapure water to each Spin column, incubate for 1 min at 20–25 °C and 1 min centrifugation at maximum speed at 20–25 °C. Load the Spin columns a second time with 50  $\mu$ l ultrapure water and centrifuge for 1 min at maximum speed at 20–25 °C.

4. Pool the eluate of both columns and measure 2  $\mu$ l of the solution using a dye-incorporation based method (e.g., Qubit).
5. Verify that the material has amplified well and that most unincorporated primers have been removed by visualizing the material on a 1.5% agarose gel (*see* Fig. 4d).
6. Pool PCR products of up to 20 viewpoints together (or bar-coded PCR products, if the same viewpoint from different samples should be pooled, *see* **Note 4**) at a final concentration of 1.625  $\mu$ g/ $\mu$ l. To avoid imbalances within the first six bases that are used for base calibration, particularly when few different viewpoints or barcodes are used, we advise mixing in 25% PhiX balancer DNA. Samples can be sequenced on the Illumina HiSeq system using at least 75 bp read length without further processing. Due to the wide length distribution of the PCR amplified material, we load the flow cell at moderately reduced cluster density (around 50–75%).

### 3.4 Data Analysis

After Illumina sequencing, bioinformatics tools should be used to generate a density profile of (averaged) DNA contacts for each viewpoint. In recent years, several 4C-seq data analysis pipelines have been developed. Publicly available examples are indicated in Subheading 2.4.

All abovementioned pipelines essentially follow the same data flow, as briefly outlined below (discussed in more detail in ref. 21). Most pipelines were developed with a specific biological question in mind, therefore using somewhat different strategies to identify significant interactions and/or changes in 3D genome organization. The choice of pipeline should be made with the biological question and availability of bioinformatics infrastructure in mind.

Summary of common data analysis steps:

1. Demultiplexing of data using the sequence specific iF sequence (and optional barcode) to identify the viewpoint-origin. Followed by removal of iF primer and barcode sequences.
2. Mapping of the remainder of reads to the reference genome followed by exclusion of reads that do not map next to restriction sites. This mapping serves as the quantitation of proximity ligation events between the viewpoint and other restriction fragments, thereby providing the snapshot of average genome-wide 3D chromatin interactions. Depending on the pipeline, the reads can further be translated into restriction fragments

and can be fitted to polymer-based models that compensate for reduced contact frequencies at increasing distance from the viewpoint.

3. In the final step, significant interactions within data sets or significant changes between data sets are determined. Depending on the pipeline, different approaches may be taken:
  - Running mean approaches with fixed window-size and thresholding to identify distant interacting regions.
  - Domainogram approaches [22], using multi-scale clustering in windows of variable size to determine local interaction trends or to identify significant distant interacting regions and determine their approximate size.
  - Statistical approaches, using variance-stabilizing transformation, to reliably determine differences in 3D organization between different samples at different length scales.

---

## 4 Notes

1. This protocol is optimized for mammalian tissue samples. With minor modifications it may be used for circulating cells, cell lines and tissues samples from other organisms as well. In these cases, the cell lysis may need to be optimized (*see Note 6*). The collagenase treatment can be omitted for circulating cells and for cell lines that do not produce collagen. If required for cell lines, we advise to do trypsin treatment prior to the start of the described protocol.
2. This protocol is optimized for 10 million mammalian cells ( $1 \times 10^7$  cells), which corresponds to about 60  $\mu\text{g}$  of chromatin. If different starting amounts are used, all volumes should be scaled accordingly. Doing the experiment on fewer than 1 million cells ( $1 \times 10^6$  cells) is not advised as loss of material during the procedure will generally result in insufficient material for the final PCR amplification.
3. High-resolution 4C-seq uses two sequential digestions with restriction enzymes that have a 4 bp recognition site. Depending on the species and the choice of restriction enzymes, the distance between informative restriction fragments is in the order of 1–1.5 kb. The choice of restriction enzymes should depend on the following parameters:
  - The primary and secondary restriction enzyme should be insensitive to the methylation state of the DNA (e.g., CpG methylation in mammalian cells).
  - The primary restriction enzyme should be capable of digesting cross-linked DNA with high efficiency at 37 °C. For mammalian cells, results with only four restriction enzymes

have been published so far: *DpnII* and its isoschizomer *MboI* (recognizing GATC), *NlaIII* (recognizing CATG), and *Csp6I* (recognizing GTAC) [10, 18, 23].

- At the intended viewpoint, digestion with the primary restriction enzyme should generate a restriction fragment that is at least 700 bp long. Though we and other groups have used smaller fragments (up to 500 bp, [23]), the success rate of smaller fragments in our hands is less reliable. *See also* Fig. 3a.
  - Several secondary restriction enzymes have been successfully used: *DpnII* (GATC), *NlaIII* (CATG), *Csp6I* (GTAC), *Taq $\alpha$ I* (TCGA, 65 °C), and *MseI* (TTAA) [10, 18, 24–26]. At the intended viewpoint, the secondary restriction enzyme should cut at least 250 bp away from the primary restriction enzyme to allow for efficient circularization [27]. *See also* Fig. 3a.
4. Good quality primers at the intended viewpoint are essential for obtaining high-quality 4C-seq data. The design strategy we present here is optimized for the Illumina HiSeq system with single-end (SE) sequencing and 100 bp read length. The use of paired-end (PE) reads or other Illumina sequencing platforms requires a modified Illumina P7 sequence. Guidelines for primer design (*see also* Fig. 3b):

- Inverse Forward (iF) and inverse Reverse (iR) primers are designed with Primer3 (version 4.0.0; <http://bioinfo.ut.ee/primer3/>) using standard settings, except for Primer size (min 18, opt 20, max 27) and Primer Tm (min 54.0, opt 55.0, max 57.0). iF and iR primers are designed independently. To avoid nonspecific amplification, primers should not map to repeated sequences.
- Illumina sequencing will be directed from the iF primer, whose sequence up to the restriction site will be included into the Illumina read. We design the iF primer within the first 30 bp from the primary restriction site to keep sufficient read length for the identification of interacting fragments (RE 1 in Fig. 3a).
- The optimal length of PCR products for Illumina sequencing, without added Illumina adapters is around 150 bp (supplemental Fig. 5 in [10]). To promote the efficiency of sequencing, we keep the average fragment length as short as possible by designing the iR primer within the first 50 bp from the secondary restriction site (RE 2 in Fig. 3a).
- After primer design, we test the quality of the iF and iR primers, without Illumina adapters, by PCR amplification on a circularized 3C template followed by gel electrophoresis (*see* Fig. 4d, left). The resulting PCR products should



run as a large smear with maximum size between 1 and 1.5 kb. In many cases, an undigested fragment can be observed with a predictable length (*see also* Fig. 3a). Few other discrete bands may be visible as well. Primers are rejected if (1) a very weak smear is detected, (2) a smear with a maximum below 1 kb is detected, (3) a very faint smear with a very strong undigested band is detected, (4) a faint smear with multiple strong (over 5) discrete bands is detected or (5) mainly primer-dimers or other fragments below 100 bp are detected.

- After verification of the primers, Illumina adapter sequences are added (*see* Fig. 3b):

iF: AATGATACGGCGACCACCGA (Illumina P5 sequence)—  
ACACTCTTTCCTACACGACGCTCTTCCGATCT  
(Illumina sequencing primer)—optional barcode—viewpoint  
specific primer sequence.

iR: CAAGCAGAAGACGGCATAACGA (short Illumina P7  
sequence)—viewpoint specific primer sequence.

Primers should be PAGE purified. Similar to the short primers, the quality should first be tested by PCR amplification on a circularized 3C template followed by gel electrophoresis (*see* Fig. 4d, middle). An optional 4–6 bp barcode can be added to the iF primer if the same viewpoint from different samples is sequenced in the same Illumina lane. We advise to use base-balanced and varied barcodes that can tolerate one or two sequencing errors (e.g., similar to the Illumina TruSeq LT Kit barcodes, as provided in the TruSeq Sample Preparation Pooling Guide).

- Primers used for Fig. 4d are as follows:

Mouse, primary enzyme: *NlaIII*, secondary enzyme: *DpnII*, viewpoint: *Hoxd13* gene [10], iF test primer: AAAATCCTAGACCTGGTCATG, iF sequencing primer: AATGATACGGCGACCACCGAACACTCTTTCCTACACGACGCTCTTCCGATCTAAAATCCTAGACCTGGTCATG, iR test primer: GGCCGATGGTGCTGTATAGG, iR sequencing primer: CAAGCAGAAGACGGCATAACGAGGCCGATGGTGCTGTATAGG

Mouse, primary enzyme: *DpnII*, secondary enzyme: *NlaIII*, viewpoint: *Amn* gene, iF sequencing primer: ATGATACGGCGACCACCGAACACTCTTTCCTACACGACGCTCTTCCGATCTCTCAGGCGCACTCTTAGCTG, iR sequencing primer: CAAGCAGAAGACGGCATAACGATATGGTAAGGCTCGGGCTG

5. If less than  $1 \times 10^7$  cells can be isolated at once, the procedure should be scaled down. Multiple smaller samples of cross-linked nuclei can be pooled prior to continuing **step 3.2**.
6. Depending on the type of cells, the cell lysis step may need to be optimized. The following cell lysis buffers have been successfully used as well:  
Primary mammalian cells and cell lines, insect cell lines: 10 mM Tris-HCl, pH 8.0, 10 mM NaCl, 0.2% NP-40, 1× complete protease inhibitor cocktail, EDTA free [12, 21] or 10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 5 mM MgCl<sub>2</sub>, 0.1 mM EGTA, 1× complete protease inhibitor cocktail, EDTA free [27].  
Primary insect cells: 15 mM Tris-HCl, pH 7.4, 0.34 M sucrose, 15 mM NaCl, 60 mM KCl, 0.2 mM EDTA, 0.2 mM EGTA, 1× complete protease inhibitor cocktail, EDTA free [28, 29].
7. Upon addition of the restriction buffer, the nuclei may visibly aggregate, which negatively influences the efficiency of the restriction enzyme reaction. A short incubation of the nuclei at higher temperature (60 °C) in the presence of 0.3% SDS will dissociate the aggregates, resulting in a somewhat milky solution without visible aggregates. The duration of the exposure to high temperature should be minimized though, as it negatively influences the integrity of formaldehyde cross-links.
8. The large volumes of ligation buffer result in large quantities of salt in the DNA solution, which are incompletely removed in the following cleanup steps (phenol-chloroform-IAA, ethanol, and QIAquick Spin Columns). Particularly DTT contamination results in a strong absorption at 260 nm, thereby interfering with the use of standard spectrophotometric determination of DNA concentration (e.g., using a NanoDrop). Dye incorporation based assays (e.g., Qubit or PicoGreen) are insensitive to the presence of salt contamination, and therefore provide a more reliable output. After PCR amplification, dye incorporation can discriminate between double stranded and single stranded DNA as well, thereby accurately quantifying the amounts of PCR-amplified DNA without measuring unincorporated primers.

---

## Acknowledgements

We thank Céline Hernandez for comments on the manuscript. Work in the D.N. laboratory is supported by funds from the Fondation pour la Recherche Médicale (FRM—Amorçage de jeunes équipes 2014, grant AJE20140630069), the Biologie Intégrative des Génomes project funded by the Initiative d'Excellence Paris-Saclay (ANR-11-IDEX-0003-02) and the Centre National de la Recherche Scientifique (CNRS).

## References

1. de Laat W, Duboule D (2013) Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* 502(7472):499–506. doi:10.1038/nature12753
2. Gorkin DU, Leung D, Ren B (2014) The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell* 14(6):762–775. doi:10.1016/j.stem.2014.05.017
3. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* 10(6):1453–1465
4. Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, Zhao Z, Lobanekov V, Reik W, Ohlsson R (2006) CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc Natl Acad Sci U S A* 103(28):10684–10689. doi:10.1073/pnas.0600326103
5. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485(7398):376–380. doi:10.1038/nature11082
6. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Bluthgen N, Dekker J, Heard E (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485(7398):381–385. doi:10.1038/nature11049
7. Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, Ong CT, Hookway TA, Guo C, Sun Y, Bland MJ, Wagstaff W, Dalton S, McDevitt TC, Sen R, Dekker J, Taylor J, Corces VG (2013) Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* 153(6):1281–1295. doi:10.1016/j.cell.2013.04.053
8. Tiwari VK, McGarvey KM, Licchesi JD, Ohm JE, Herman JG, Schubeler D, Baylin SB (2008) PcG proteins, DNA methylation, and gene repression by chromatin looping. *PLoS Biol* 6(12):2911–2927. doi:10.1371/journal.pbio.0060306
9. Lanzuolo C, Roure V, Dekker J, Bantignies F, Orlando V (2007) Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat Cell Biol* 9(10):1167–1174. doi:10.1038/ncb1637
10. Noordermeer D, Leleu M, Splinter E, Rougemont J, De Laat W, Duboule D (2011) The dynamic architecture of Hox gene clusters. *Science* 334(6053):222–225. doi:10.1126/science.1207194
11. Noordermeer D, Leleu M, Schorderet P, Joye E, Chabaud F, Duboule D (2014) Temporal dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. *Elife* 3:e02557. doi:10.7554/eLife.02557
12. Bantignies F, Roure V, Comet I, Leblanc B, Schuettengruber B, Bonnet J, Tixier V, Mas A, Cavalli G (2011) Polycomb-dependent regulatory contacts between distant Hox loci in *Drosophila*. *Cell* 144(2):214–226. doi:10.1016/j.cell.2010.12.026
13. Tolhuis B, Blom M, Kerkhoven RM, Pagie L, Teunissen H, Nieuwland M, Simonis M, de Laat W, van Lohuizen M, van Steensel B (2011) Interactions among Polycomb domains are guided by chromosome architecture. *PLoS Genet* 7(3):e1001343. doi:10.1371/journal.pgen.1001343
14. Denholtz M, Bonora G, Chronis C, Splinter E, de Laat W, Ernst J, Pellegrini M, Plath K (2013) Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell* 13(5):602–616. doi:10.1016/j.stem.2013.08.013
15. Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. *Science* 295(5558):1306–1311. doi:10.1126/science.1067799
16. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* 38(11):1348–1354. doi:10.1038/ng1896
17. David FP, Delafontaine J, Carat S, Ross FJ, Lefebvre G, Jarosz Y, Sinclair L, Noordermeer D, Rougemont J, Leleu M (2014) HTSstation: a web application and open-access libraries for high-throughput sequencing data analysis. *PLoS One* 9(1):e85879. doi:10.1371/journal.pone.0085879
18. van de Werken HJ, Landan G, Holwerda SJ, Hoichman M, Klous P, Chachik R, Splinter E, Valdes-Quezada C, Oz Y, Bouwman BA, Verstegen MJ, de Wit E, Tanay A, de Laat W (2012) Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat Methods* 9(10):969–972. doi:10.1038/nmeth.2173
19. Klein FA, Anders S, Pakozdi T, Ghavi-Helm Y, Furlong EEM, Huber W (2014) FourCSeq:

- analysis of 4C sequencing data. *bioRxiv*. doi:10.1101/009548
20. Thongjuea S, Stadhouders R, Grosveld FG, Soler E, Lenhard B (2013) r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. *Nucleic Acids Res* 41(13):e132. doi:10.1093/nar/gkt373
  21. Gheldof N, Leleu M, Noordermeer D, Rougemont J, Reymond A (2012) Detecting long-range chromatin interactions using the chromosome conformation capture sequencing (4C-seq) method. *Methods Mol Biol* 786:211–225. doi:10.1007/978-1-61779-292-2\_13
  22. de Wit E, Braunschweig U, Greil F, Bussemaker HJ, van Steensel B (2008) Global chromatin domain organization of the *Drosophila* genome. *PLoS Genet* 4(3):e1000045. doi:10.1371/journal.pgen.1000045
  23. van de Werken HJ, de Vree PJ, Splinter E, Holwerda SJ, Klous P, de Wit E, de Laat W (2012) 4C technology: protocols and data analysis. *Methods Enzymol* 513:89–112. doi:10.1016/B978-0-12-391938-0.00004-5
  24. Kernohan KD, Vernimmen D, Gloor GB, Berube NG (2014) Analysis of neonatal brain lacking ATRX or MeCP2 reveals changes in nucleosome density, CTCF binding and chromatin looping. *Nucleic Acids Res* 42(13):8356–8368. doi:10.1093/nar/gku564
  25. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C, Aneas I, Credidio FL, Sobreira DR, Wasserman NF, Lee JH, Puvion-Vandier V, Tam D, Shen M, Son JE, Vakili NA, Sung HK, Naranjo S, Acemel RD, Manzanares M, Nagy A, Cox NJ, Hui CC, Gomez-Skarmeta JL, Nobrega MA (2014) Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* 507(7492):371–375. doi:10.1038/nature13138
  26. Woltering JM, Noordermeer D, Leleu M, Duboule D (2014) Conservation and divergence of regulatory strategies at Hox Loci and the origin of tetrapod digits. *PLoS Biol* 12(1):e1001773. doi:10.1371/journal.pbio.1001773
  27. Simonis M, Kooren J, de Laat W (2007) An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods* 4(11):895–901. doi:10.1038/nmeth1114
  28. Bonn S, Zinzen RP, Perez-Gonzalez A, Riddell A, Gavin AC, Furlong EE (2012) Cell type-specific chromatin immunoprecipitation from multicellular complex samples using BiTS-ChIP. *Nat Protoc* 7(5):978–994. doi:10.1038/nprot.2012.049
  29. Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W, Furlong EE (2014) Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* 512(7512):96–100. doi:10.1038/nature13417