

## Methods

journal homepage: [www.elsevier.com/locate/ymeth](http://www.elsevier.com/locate/ymeth)



# C-TALE, a new cost-effective method for targeted enrichment of Hi-C/3C-seq libraries

Arkadiy K. Golov<sup>a,f</sup>, Sergey V. Ulianov<sup>a,b</sup>, Artem V. Luzhin<sup>a</sup>, Ekaterina P. Kalabusheva<sup>c,d</sup>, Omar L. Kantidze<sup>a</sup>, Ilya M. Flyamer<sup>e</sup>, Sergey V. Razin<sup>a,b</sup>, Alexey A. Gavrilov<sup>a,\*</sup>

<sup>a</sup> Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia

<sup>b</sup> Faculty of Biology, M.V. Lomonosov Moscow State University, Moscow, Russia

<sup>c</sup> Koltsov Institute of Developmental Biology, Russian Academy of Sciences, Moscow, Russia

<sup>d</sup> Pirogov Russian National Research Medical University, Research Institute of Translational Medicine, Department of Regenerative Medicine, Moscow, Russia

<sup>e</sup> MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK

<sup>f</sup> Mental Health Research Center, Moscow, Russia

## ARTICLE INFO

### Keywords:

Genome folding

C-methods

Hi-C

In-solution hybridization

## ABSTRACT

Studies performed using Hi-C and other high-throughput whole-genome C-methods have demonstrated that 3D organization of eukaryotic genomes is functionally relevant. Unfortunately, ultra-deep sequencing of Hi-C libraries necessary to detect loop structures in large vertebrate genomes remains rather expensive. However, many studies are in fact aimed at determining the fine-scale 3D structure of comparatively small genomic regions up to several Mb in length. Such studies typically focus on the spatial structure of domains of coregulated genes, molecular mechanisms of loop formation, and interrogation of functional significance of GWAS-revealed polymorphisms. Therefore, a handful of molecular techniques based on Hi-C have been developed to address such issues. These techniques commonly rely on in-solution hybridization of Hi-C/3C-seq libraries with pools of biotinylated baits covering the region of interest, followed by deep sequencing of the enriched library. Here, we describe a new protocol of this kind, C-TALE (Chromatin TArget Ligation Enrichment). Preparation of hybridization probes from bacterial artificial chromosomes and an additional round of enrichment make C-TALE a cost-effective alternative to existing many-versus-all C-methods.

## 1. Introduction

Spatial organization of chromatin inside the eukaryotic nucleus is interconnected with its functional activities such as transcription, replication, and repair [1,2]. Recent advances in our understanding of genome architecture are largely based on the usage of high-throughput biochemical protocols collectively referred to as C-methods or 3C technologies [3]. All C-methods including the basic 3C (Chromosome Conformation Capture) [4], utilize the proximity ligation procedure. The idea of proximity ligation is that spatially proximal fragments of endonuclease-digested genomic DNA preferentially ligate to each other independently of whether or not they are adjacent in a linear DNA. Thereby, the frequency of cross-ligation products between two DNA

fragments in proximity ligation output (which is commonly termed a 3C template) approximately reflects the proportion of cells from a population in which these two fragments are in close vicinity. The first approach aimed to study spatial DNA folding by analyzing proximity ligation products was developed even before the 3C protocol [5], but addition of the cross-linking step in 3C significantly improved the sensitivity and reproducibility of the method [4].

Originally, 3C was a rather low-throughput hypothesis-driven method which allowed evaluation of the frequency of spatial interactions of one or several genomic sites of interest (baits) with a few pre-selected partners by PCR. However, interrogation of a 3C template opened wide perspectives and had clear advantages compared to microscopic methods such as FISH, which played a key role in 3D genome

**Abbreviations:** 3C, chromosome conformation capture; 4C, circular chromosome conformation capture; 5C, chromosome conformation capture carbon copy; BAC, bacterial artificial chromosome; C-TALE, Chromatin TArget Ligation Enrichment; ChIP-seq, chromatin immunoprecipitation sequencing; FISH, fluorescence in situ hybridization; GWAS, genome-wide association study; Hi-C, high-throughput chromosome conformation capture; NGS, next generation sequencing; RT, room temperature; SNP, single nucleotide polymorphisms; SPRI, solid phase reversible immobilization; T2C, targeted chromatin capture; TAD, topologically associated domain

\* Corresponding author.

E-mail address: aleksey.a.gavrilov@gmail.com (A.A. Gavrilov).

<https://doi.org/10.1016/j.ymeth.2019.06.022>

Received 25 March 2019; Accepted 22 June 2019

Available online 26 June 2019

1046-2023/ © 2019 Elsevier Inc. All rights reserved.

organization studies prior to the advent of C-methods. First of all, potential resolution of C-methods is equivalent to the genomic distance between two recognition sites of a utilized endonuclease (approximately 5 kb for “6-cutters”, 500 bp for “4-cutters”, and down to 1 bp in the case of DNase) which is several orders of magnitude higher than the resolution achieved in most refined microscopic studies. In addition, the throughput of C-methods is highly scalable: the 3C template contains information regarding spatial distribution of all genomic regions relative to each other. This “hidden” potential of the 3C method was embodied by joining proximity ligation with next-generation sequencing (NGS) in high-throughput derivatives of 3C. The most powerful of the 3C technologies, Hi-C, provides for a comprehensive quantitative description of DNA folding on a whole-genome scale [6]. In this data-driven approach, it became possible to evaluate the spatial proximity of each genomic region to every other (so called all-versus-all) with up to 1 kb resolution [7]. Hi-C libraries have immense combinatorial complexity due to the huge variety of theoretically possible ligation products (hundreds and thousands of billions), a substantial part of which are actually generated in experiments. Therefore, ultradeep sequencing of billions of read pairs is necessary to build whole-genome maps with a resolution sufficiently high for detection of the finest structural features of vertebrate genome folding, such as looping interactions between pairs of remote loci. However, many studies aim at detailed description of spatial DNA organization of comparatively short genomic regions. Specifically, for achieving such goals, the 5C variant of high-throughput C-methods was developed [8] and then gradually improved [9]. In 5C, the 3C template is enriched with ligation products from selected genomic regions of up to several megabases using a technique referred to as ligation-mediated amplification. Using 5C, one can analyze frequency of spatial interactions of many DNA fragments from one genomic region between each other but not with the rest of the genome (many-versus-many approach). Unfortunately, 5C suffers from a number of crucial flaws, such as the inability to acquire maps with high resolution readily achievable by Hi-C, the impossibility of analyzing interactions of a studied region with other parts of the genome, complex primer design and relatively labor-intensive procedures. Hence, the last five years have seen almost complete displacement of 5C with techniques exploiting an in-solution hybridization strategy for detailed analysis of DNA folding in extended genomic loci of interest [10].

Initially, in-solution hybridization was used for enrichment of genomic DNA with 1–2% of coding sequences followed by NGS [11]. The method referred to as exome sequencing is now widely applied for mapping of rare genetic variants causing Mendelian diseases or variants associated with common diseases. In liquid hybridization-based C-methods discussed here, pools of biotinylated DNA or RNA probes, covering a locus of interest, are hybridized with melted DNA of Hi-C or 3C-seq libraries to enrich them with junctions from the targeted regions (hereinafter by 3C-seq we will mean a method similar to Hi-C, in which stages of filling restriction ends with biotinylated nucleotides and subsequent biotin pull-down of chimeric ligation products are absent; originally, such a method was termed GCC — genome conformation capture [12]). Remarkably, not only are ligation events inside targeted regions captured in such approaches, but also interactions between the region and the rest of genome (many-versus-all approaches). Probes for targeted enrichment of Hi-C/3C-seq libraries are most frequently purchased as a part of commercial custom kits. Several different variants of targeted enrichment of Hi-C libraries have been published in recent years, differing with respect to the details of probe manufacture and hybridization protocol [13–15].

Here, we describe a new flexible and cost-effective method for the analysis of 3D organization of extended genomic loci, which is based on in-solution hybridization-mediated enrichment. The key feature of the approach, which we term C-TALE (Chromatin TArget Ligation Enrichment) [16], is the manufacturing hybridization probes from DNA of bacterial artificial chromosome (BAC) clones harboring loci of interest. C-TALE allows one to obtain DNA folding maps of vertebrate

megabase-sized genomic regions with up to 1 kb resolution at a sequencing depth of several tens of millions of reads. We successfully applied C-TALE to map fine-scale DNA spatial organization of extended genomic loci in cells of diverse vertebrate species such as zebrafish, chickens, and humans. Its relative simplicity and low cost make C-TALE an ideal tool for studying the influence of various experimental factors (physicochemical conditions, drug treatment, genetic or epigenetic perturbations) on the spatial organization of chromatin at the scale of enhancer-promoter or CTCF-mediated loops. Below, we discuss key points significant for successful conduction of C-TALE and provide a step-by-step protocol of the method.

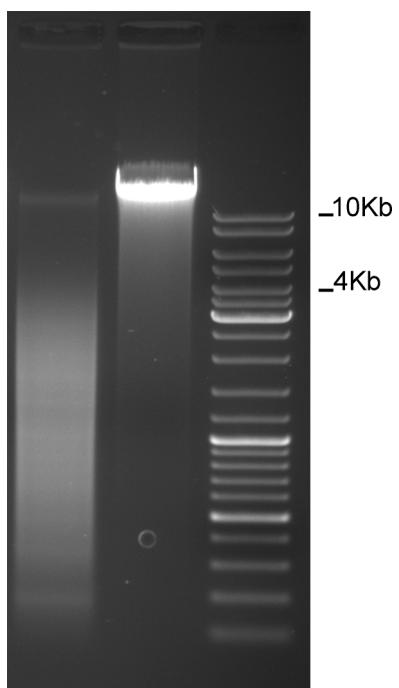
## 2. Method overview

### 2.1. Proximity ligation and 3C-seq library preparation

The C-TALE experiment begins with preparation of a 3C template with subsequent production of an NGS library from this template. We use the conventional 3C-protocol without Hi-C-specific steps of biotin fill-in of cohesive ends and streptavidin pull-down of chimeric ligation products. In our experience, these steps substantially diminish the complexity of the final libraries [17]. This latter fact makes acquiring high-quality spatial maps problematic. Indeed, at a certain point, further increase of the sequencing depth of C-TALE libraries made on the basis of Hi-C libraries does not lead to the increase in detected unique ligation products but only multiplies PCR duplicates (particularly when using less than five million cells at the start). Thus, we prefer to sacrifice enrichment with ligation junctions (only 10–20% of raw reads in our 3C-seq libraries contain information about ligation events, compared to 30–70% of such reads in Hi-C libraries) for library complexity and ultimately for map resolution. We also recommend using 4-cutter restriction enzymes (MboI, DpnII or NlaIII) and a protocol of in-situ ligation to improve data quality [18,19]. It is preferable to control the quality of the 3C template by assessing restriction endonuclease digestion and proximity ligation efficiency using electrophoresis of DNA from corresponding stages. DNA smear distribution in successful experiments should be almost entirely under 4 kb after digestion with 4-cutters and above 10 kb after proximity ligation (Fig. 1). After digestion and proximity ligation, a high molecular weight 3C template is sonicated to 150–700 bp length distribution, followed by a standard protocol for NGS library preparation including DNA end repair, A-tailing, and Illumina Y-adapter ligation. SPRI (Solid Phase Reversible Immobilization) magnetic bead clean-up after each enzymatic stage is applied. Usage of different sequencing barcodes in Illumina adapters for libraries from diverse experimental contexts (biological replicates, different cell types, differently treated cells, genetically or epigenetically perturbed cells) makes it possible to mix the libraries and perform enrichment in a single hybridization reaction. Demultiplexing of reads with different barcodes after sequencing allows the building of separate heatmaps of spatial folding for each specific experimental context. After final SPRI clean-up, analytical PCR with Illumina primers is done to choose the optimal amplification cycle. The remaining part of a pre-PCR library is then amplified using an appropriate number of cycles.

### 2.2. Generation of biotinylated baits from BAC DNA

In C-TALE, biotinylated probes for targeted enrichment of 3C-seq libraries are generated from BACs bearing the locus of interest (Fig. 2A). In other techniques such as Capture Hi-C, T2C (Targeted Chromatin Capture), or Hi-C<sup>2</sup> (HYbrid Capture Hi-C), pools of long (60–120 bases) synthetic single-stranded oligonucleotides (RNA or DNA) are used. These pools can be pre-labeled with biotin, in which case they are ready for annealing immediately [13,14]. Otherwise, biotinylated nucleotides can be inserted for example by in-vitro transcription [15]. C-TALE involves manufacturing of double-stranded DNA probes from cheap and easily available material — BAC DNA.

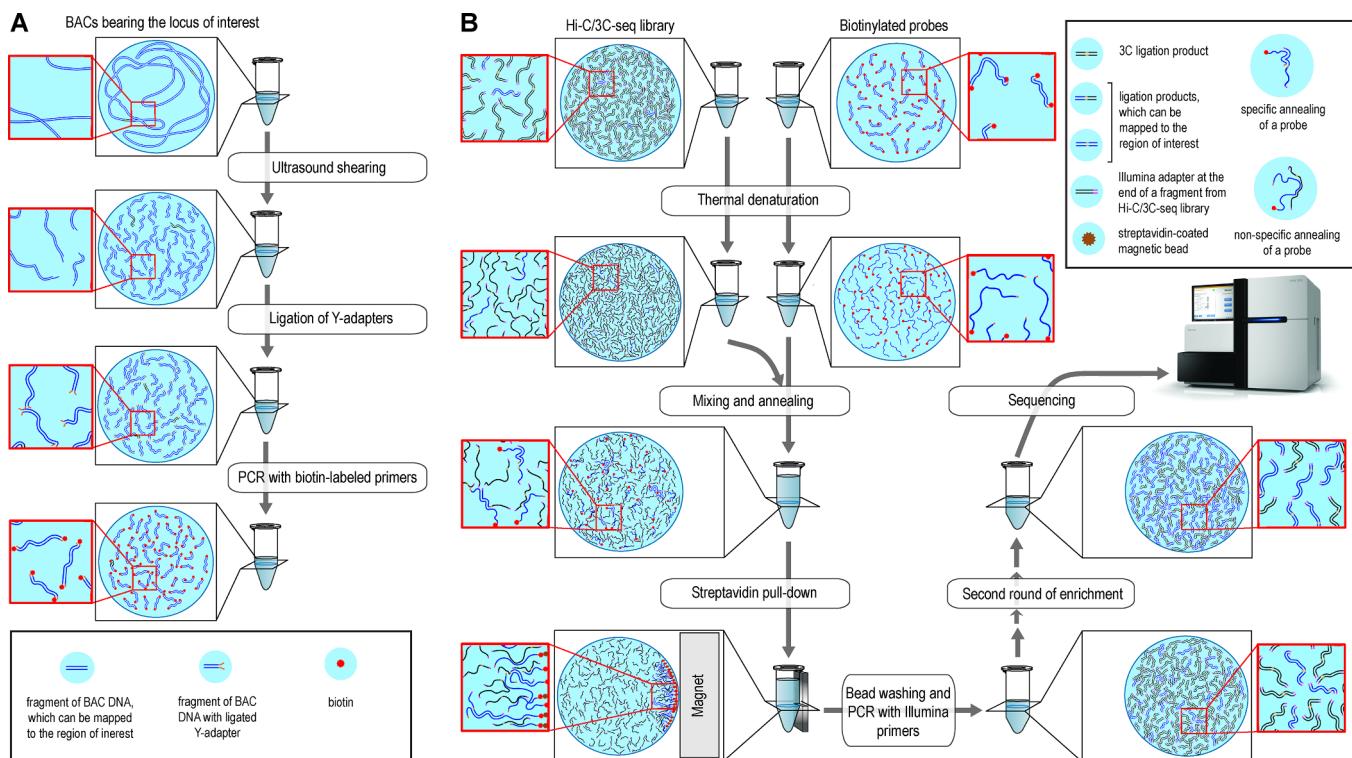


**Fig. 1.** Typical example of chromatin digestion control and proximity-ligated 3C template length distribution on 0.8% agarose electrophoresis. Left to right: chromatin digested with DpnII, in-situ religated 3C template, DNA ladder mix (ThermoScientific, cat.# SM0331). Digested DNA should almost completely run below 4-kb marker band, whereas 3C template should appear as a tight band above 10 kb.

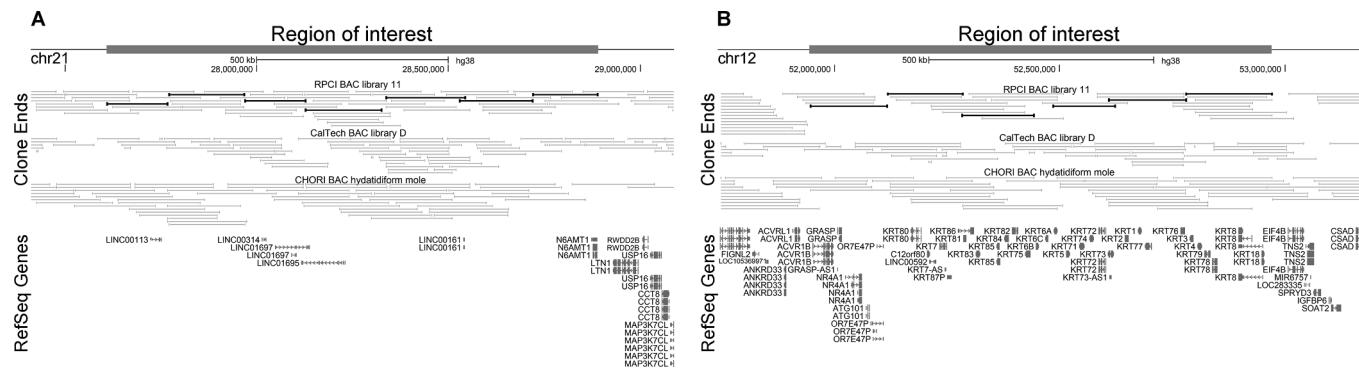
Appropriate BAC clones (carrying genomic fragments covering a region of interest) for human and many model organisms can be selected via the UCSC genome browser (“Clone Ends” track) and purchased through

BACPAC Resources (<https://bacpacresources.org>) or Invitrogen (Fig. 3). If some genomic region is not covered with BAC clones or suitable clones cannot be purchased, the gap can be filled with self-made PCR-product clones. One should not be concerned about partial overlapping of BAC clones: differences in hybridization efficiency caused by such overlaps can be corrected afterwards bioinformatically (see 2.4). BAC clones covering non-adjacent genomic regions can also be used for preparation of a joint set of probes for C-TALE hybridization. We prepared probes for a single C-TALE experiment from up to seven BAC clones, inserts from which covered approximately 1.5 Mb, and we suggest that even larger regions can be analyzed. We also observed that the larger the region covered with the probes is, the more significant is the overall enrichment efficiency. It should be mentioned that we are not the first to produce hybridization probes for targeted enrichment of NGS libraries from BAC DNA [20,21]. However, in these protocols, probes were generated in a different manner (using a nick translation technique) and to our knowledge, there are no publications where such probes would be exploited to enrich Hi-C or 3C-seq libraries.

An overall process of probe preparation in C-TALE is analogous to preparation of libraries for high-throughput sequencing. BAC DNA from chosen clones is isolated either with isolation kits or with the standard alkaline lysis/alcohol precipitation technique. Then, an equimolar mix of all clones is made and sonicated to 150–500 bp size distribution. Enzymatic reactions alternate with SPRI clean-ups as in the 3C-seq library preparation procedure, including end repair, A-tailing, and Y-adapter ligation. The only distinction is that the fork adapter’s non-complementary parts should substantially differ from the Illumina PE primers used for sequencing library amplification. Finally, a library of probes is amplified using PCR with 5'-biotinylated primers, giving a pool of dsDNA probes. Only a small part of a pre-PCR probe library is typically enough to obtain an amount of baits sufficient for one C-TALE experiment. The residual part of the library can be stored frozen until necessary. Once prepared, these libraries can be used as a source of probes for tens of C-TALE hybridizations during many months.



**Fig. 2.** A schematic of C-TALE experimental procedure (see text for details). (A) Generation of biotinylated probes from BAC DNA. (B) In-solution hybridization of Hi-C/3C-seq libraries with biotinylated probes, subsequent biotin pull-down, PCR-amplification, and sequencing.



**Fig. 3.** Choice of BAC clones for C-TALE probe preparation. Regions on human chromosome 21 (left) and 12 (right) are shown with available BAC clones underneath. Clones purchased for C-TALE experiments are highlighted in black. Note that it is convenient to select minimally overlapping BAC clones covering the region of interest.

### 2.3. Targeted enrichment

The stage of hybridization and enrichment in C-TALE is almost indistinguishable from similar stages in many published protocols (e.g. [15]) (Fig. 2B). Two mixes are prepared: a “3C-seq” mix and a “probe” mix, the first containing a 3C-seq library(ies) and the second containing biotinylated baits. Both mixes are supplemented with blocking oligos, complementary to corresponding universal DNA ends, an approach which prevents adapter-mediated cross-hybridization. To decrease non-specific annealing of GC-rich and repetitive sequences to probes, salmon DNA and Cot-1 DNA are added to the mixes. If Cot-1 DNA of species used for 3C-seq library generation is absent, sonicated genomic DNA can be used instead.

As already mentioned, several differentially barcoded libraries can be combined into one “3C-seq” mix to be demultiplexed afterwards bioinformatically. It should be taken into account that the total amount of DNA in one annealing remains constant, and thus an increase in library multiplexing leads to the decrease of complexity of each library. Hence, we recommend increasing the number of hybridization reactions in case one is performing C-TALE on more than five 3C-seq libraries.

The two prepared mixes are melted in a thermal cycler, then hybridization buffer is added to both mixes, repetitive and GC-rich sequences are blocked, and finally mixes are combined in one hybridization reaction and annealed for at least overnight. Afterwards, fragments of 3C-seq libraries hybridized to biotinylated probes are pulled-down on streptavidin beads and extensively washed. These beads are then used to PCR-amplify 3C-seq libraries enriched with ligation junctions from the target region. Unfortunately, the procedure of hybridization in C-TALE is rather inefficient, so that less than five percent of valid pairs from a singly enriched 3C-seq library map to the targeted region. Preferential annealing of individual strands of dsDNA probes with each other and imperfect annealing of chimeric 3C-seq libraries on probes originating from un-reshuffled genomic sequences are among the most significant reasons for this issue. To make the enrichment efficiency approximately ten times higher, we introduced a second round of hybridization equivalent to the first one. In our experience, this additional hybridization, which has been extensively employed previously [17,20,22], does not lead to any considerable bias. The second round of enrichment allowed us to map 20–50% of valid read pairs to the targeted region of the genome.

## 3. Step-by-step protocol

### 3.1. Proximity ligation and 3C-seq library preparation

#### 3.1.1. Cross-linking of cells

Grow the cells in an appropriate culture medium. Carefully estimate

the concentration of cells in suspension or their confluence in case of adherent culture.

#### 3.1.1.1. Adherent cells:

1. Aspirate culturing medium from sub-confluent 10 cm plate, containing 3–5\*10<sup>6</sup> cells.
2. Wash the cells once with 10 ml of 3C buffer.
3. Cross-link the cells by adding 9 ml of 1.5% formaldehyde solution in 3C buffer and incubating the plate at room temperature (RT) for exactly 10 min.
4. Add 1 ml of 2 M glycine to a final concentration of 0.2 M, mix gently and incubate for 10 min at RT to quench the crosslinking reaction.
5. Discard the supernatant, add 9 ml of 3C buffer and 1 ml of 2 M glycine, mix gently and incubate for 10 min at RT.
6. Discard the supernatant, add 8 ml of PBS, scrape the cells from the plate with a cell scraper, transfer them to a 15 ml tube.
7. Spin down the cells at 600 g for 5 min, remove the supernatant completely.

#### 3.1.1.2. Suspension cells

1. Centrifuge a volume of growth medium, containing 3–5\*10<sup>6</sup> of suspension cells, at 400 g for 5 min.
2. Discard the supernatant, resuspend the pellet in 1 ml of 3C buffer and transfer it to a microcentrifuge tube.
3. Pellet the cells at 400 g for 3 min, discard the supernatant.
4. Gently resuspend the cells in 900 µl of 1.5% formaldehyde solution in 3C buffer. Cell clumps should be broken by pipetting. Cross-link the cells, agitating them gently on a rotator for exactly 10 min at room temperature (RT).
5. Add 100 µl of 2 M glycine to a final concentration of 0.2 M. Agitate the cells on a rotator for 10 min at RT to quench the crosslinking reaction.
6. Centrifuge the cells at 600 g for 3 min, discard the supernatant.
7. Resuspend the pellet in 900 µl of 3C buffer, add 100 µl 2 M glycine, agitate on a rotator for 10 min at RT.
8. Spin down the cells at 600 g for 3 min, discard the supernatant.
9. Wash the cells by resuspending them in 1 ml of PBS and subsequent centrifugation at 600 g for 3 min. Remove the supernatant completely.

#### 3.1.2. Cell lysis, restriction endonuclease digestion

10. Place the pellet on ice for 2 min.
11. Resuspend the pellet in 1 ml of ice-cold lysis buffer, incubate for 15 min on ice.
12. Vortex lysed cells briefly and pellet them in pre-cooled centrifuge at

600 g for 3 min at 4 °C, discard supernatant.

Note: at this stage, the pellet of lysed cells can be snap-frozen and stored at –70 °C for at least one year. When convenient, one can thaw frozen material and continue with the subsequent procedure.

13. Resuspend the pellet in 270 µl of 1x NEBuffer2, transfer the material to a new microtube, add 3.75 µl of 20% SDS solution, mix gently by pipetting.
14. Incubate in a ThermoMixer with shaking (1,000 rpm) at 37 °C for 1 h.
15. Add 25 µl of 20% Triton X-100 solution, mix gently by pipetting.
16. Incubate in a ThermoMixer with shaking (1,000 rpm) at 37 °C for 1 h.
17. Centrifuge the nuclei at 900 g for 2 min at RT.
18. Resuspend the pellet in 300 µl of appropriate 1x digestion buffer, transfer the suspension to a new tube, add 200–500 units of chosen “4-cutter” enzyme (highly concentrated MboI (25 U/µl), DpnII (50 U/µl) or NlaIII (10 U/µl)), mix by vortexing.

Note: buffer in which the chosen endonuclease shows maximal activity should be utilized. We use NEBuffer 2.1 for MboI, NEBuffer DpnII — for DpnII and CutSmart buffer — for NlaIII.

19. Incubate with vigorous shaking (1,000 rpm) in a ThermoMixer at 37 °C overnight or for at least 4 h.
20. Heat inactivate the enzyme by incubation the tube in a ThermoMixer with shaking at 65 °C for 20 min.

### 3.1.3. In-situ proximity ligation and DNA purification

21. Spin down the nuclei with digested chromatin at 900 g for 2 min.
22. Resuspend the pellet in 300 µl of 1x T4 DNA ligase buffer, take 15 µl of the suspension in a new tube as a control of digestion efficiency. Transfer remaining part of suspension to another clean tube, centrifuge at 900g for 2 min.
23. Resuspend the pellet in 300 µl of 1x T4 DNA ligase buffer, add 5 µl of T4 DNA ligase (5 U/µl), mix by vortexing.
24. Incubate with shaking (1,000 rpm) in a ThermoMixer with cooling at 16 °C overnight or at least 4 h to perform proximity ligation.
25. Add 100 µl of PBS, 20 µl of 20% SDS solution, and 5 µl of proteinase K (10 mg/ml).
26. Add to 15 µl of digestion control from step 22 375 µl of PBS, 20 µl of 20% SDS solution, and 3 µl of proteinase K (10 mg/ml).
27. Incubate both samples (control and ligated material) at 65 °C in a ThermoMixer without shaking overnight or at least for four hours to reverse cross-linking.
28. Cool down both samples after de-cross-linking to RT, add 420 µl of saturated phenol (pH 8.0): chloroform (1:1) mixture, vortex thoroughly, centrifuge at maximum speed for 3 min.
29. Transfer 400 µl of top aqueous phase from each sample to a new microtube, add 400 µl of chloroform, vortex thoroughly, centrifuge at maximum speed for 3 min.
30. Transfer 370 µl of top aqueous phase from each sample to a new microtube, add 30 µl of 3 M sodium acetate solution (pH 5.2), 1 µl of glycogen (20 mg/ml), mix thoroughly.
31. Add 1,300 µl of 96% ice-cold ethanol to each sample, mix thoroughly, pellet DNA in a pre-cooled centrifuge at maximum speed at 4 °C for 10 min.
32. Discard the supernatants, wash the pellets with 1 ml of cold 70% ethanol, centrifuge at maximum speed at 4 °C for 5 min.
33. Discard the supernatants, remove ethanol drops with water jet pump, and air-dry the pellets.
34. Dissolve digestion control DNA pellet in 25 µl of 10 mM Tris-HCl (pH 8.0) and 3C template in 100 µl of the same solvent.

Note: One may not treat these samples with RNase A because the overwhelming majority of RNA molecules are eluted from nuclei prior to the proximity ligation step. The remaining traces of RNA do not interfere with further steps.

35. Measure DNA concentration of each sample fluorometrically with a Qubit dsDNA Broad Range kit according to the manufacturer's instructions.

Note: if starting with 3–5\*10<sup>6</sup> cells, the total amount of DNA is normally 0.5–2 µg in the digestion control and 10–40 µg in the 3C template.

36. Run 200–300 ng of digestion control, the same amount of 3C template, and 1 µl of DNA size marker (e.g. GeneRuler DNA ladder mix) on a 0.7% agarose gel to check the efficiency of chromatin digestion and proximity ligation.

Note: The smear of digested DNA should almost completely reside below the 4-kb marker band, while the 3C template should run as more or less a tight band above the 10-kb marker band. The portion of high-quality 3C template can be used for immediate 3C-seq library preparation, and the remaining part can be stored at –20 °C for months. Otherwise, the 3C-template can be frozen entirely until the need for a 3C-seq library will arise. If DNA size distribution does not correspond to the mentioned pattern, we recommend to repeat 3C template preparation from the beginning or starting with a frozen pellet from step 12.

### 3.1.4. 3C-seq library preparation

#### 3.1.4.1. 3C template sonication

37. Shear 5 µg of the 3C template to a size of 150–700 bp using a sonicator. We use the VirSonic 100 (VirTis) sonicator with the following parameters: DNA is diluted in 500 µl of 1x sonication buffer, shearing is performed on ice with 4 30-sec pulses on “15” power setting, separated with 3 min periods of recovery.

Note: parameters of shearing should be chosen depending on the utilized instrument model.

38. Centrifuge the solution of sheared DNA for five minutes at maximum speed.
39. Transfer the supernatant to a 30-kDa Amicon Column, centrifuge for five minutes at maximum speed.
40. Discard the flow-through, add 450 µl of 10 mM Tris-HCl (pH 8.0) to the column, centrifuge for five minutes at maximum speed.
41. Transfer concentrated DNA solution from the column to a new microtube, add 1.5 volumes of AmpureXP beads, mix thoroughly, incubate for 10 min.
42. Place the tube on the magnet for 2 min, discard the supernatant.
43. Wash beads three times with 200 µl of 80% ethanol, carefully remove drops of ethanol after last wash, air-dry the beads.

Note: here and elsewhere, do not displace the tube from the magnet during the washing steps so that beads remain attached to the tube wall.

44. Remove the tube from the magnet, resuspend the beads in 50 µl of 10 mM Tris-HCl (pH 8.0), incubate for 5 min at RT.
45. Place the tube on the magnet for 2 min, transfer the supernatant to a fresh tube.

#### 3.1.4.2. End repair

46. Prepare 2x end repair mix as follows:

Ingredient	volume per 1 reaction, $\mu$ l
10x T4 DNA ligase buffer	10
dNTP mix (10 mM)	5
T4 Polynucleotide Kinase (10 U/ $\mu$ l)	5
T4 DNA polymerase (3 U/ $\mu$ l)	4
Klenow fragment of DNA polymerase I (5 U/ $\mu$ l)	1
milliQ	25

47. Add 50  $\mu$ l of 2x end repair mix to sheared 3C template DNA from step 45, mix thoroughly.
48. Incubate the tube at RT for 30 min to blunt sonication-damaged DNA ends.
49. Add 150  $\mu$ l of AmpureXP beads to the reaction tube, mix thoroughly, incubate for 10 min.
50. Place the tube on the magnet for 2 min, discard the supernatant.
51. Wash beads three times with 300  $\mu$ l of 80% ethanol, carefully remove drops of ethanol after last wash, air-dry the beads.
52. Remove the tube from the magnet, resuspend beads in 50  $\mu$ l of 10 mM Tris-HCl (pH 8.0), incubate for 5 min at RT.
53. Place the tube on the magnet for 2 min, transfer the supernatant to a fresh tube.

#### 3.1.4.3. A-tailing

54. Prepare 2x A-tailing mix as follows:

Ingredient	volume per 1 reaction, $\mu$ l
10x NEBuffer 2	10
Klenow Fragment 3'→5' exo-(5 U/ $\mu$ l)	5
dATP (10 mM)	5
milliQ	30

55. Add 50  $\mu$ l of 2x A-tailing mix to DNA solution from step 53, mix thoroughly.
56. Incubate tube at 37 °C for 30 min to add dA to 5' DNA ends.
57. Add 150  $\mu$ l of AmpureXP beads to reaction tube, mix thoroughly, incubate for 10 min.
58. Place the tube on the magnet for 2 min, discard the supernatant.
59. Wash beads three times with 300  $\mu$ l of 80% ethanol, carefully remove drops of ethanol after last wash, air-dry the beads.
60. Remove the tube from the magnet, resuspend beads in 20  $\mu$ l of 10 mM Tris-HCl (pH 8.0), incubate for 5 min at RT.
61. Place the tube on the magnet for 2 min, transfer the supernatant to a fresh tube.

#### 3.1.4.4. Adapter ligation

62. Prepare 2x adapter ligation mix as follows:

Ingredient	volume per 1 reaction, $\mu$ l
10x T4 DNA ligase buffer	5
T4 DNA ligase (5 U/ $\mu$ l)	2
milliQ	18

63. Add 5  $\mu$ l of desired Illumina PE single-indexed TruSeq forked adapter to A-tailed DNA solution from step 61.
64. Add 25  $\mu$ l of 2x adapter ligation mix to DNA solution from previous step, mix thoroughly.
65. Incubate tube at RT overnight

Note: in our experience, longer incubation in this step improves the overall efficiency of library preparation.

66. On the next day, add 75  $\mu$ l of AmpureXP beads to the reaction tube, mix thoroughly, incubate for 10 min.
67. Place the tube on the magnet for 2 min, discard the supernatant.
68. Wash beads three times with 200  $\mu$ l of 80% ethanol, carefully remove drops of ethanol after last wash, air-dry the beads.
69. Remove the tube from the magnet, resuspend beads in 27  $\mu$ l of 10 mM Tris-HCl (pH 8.0), incubate for 5 min at RT.
70. Place the tube on the magnet for 2 min, transfer the supernatant to a fresh tube.

Note: the pre-PCR 3C-seq library can be stored at 4 °C for several days.

#### 3.1.4.5. PCR amplification of 3C-seq library.

71. To determine the number of PCR cycles adequate for generating an amount of 3C-seq library sufficient for C-TALE hybridization, set up a trial PCR reaction as follows:

Ingredient	volume per 1 reaction, $\mu$ l
KAPA HiFi 5x buffer	5
dNTP mix (10 mM)	0.75
Illumina-for primer (10 $\mu$ M)	1.25
Illumina-rev primer (10 $\mu$ M)	1.25
DNA solution from step 70	5
KAPA HiFi HotStart polymerase (1 U/ $\mu$ l)	1
milliQ	10.75

Use the following PCR program:

step	temperature, °C	time, s
1	95	300
repeat steps 2–4 7 times		
2	98	20
3	65	15
4	72	20
5	72	180

72. Take 3  $\mu$ l aliquots of PCR reaction at the end of the 3rd, 5th, and 7th cycle.
73. Run taken samples on 1% agarose gel with 1  $\mu$ l of 50-bp DNA ladder to choose an optimal PCR cycle.

Note: we typically use cycles in which the smear of amplified 3C-seq library is readily seen on the gel, but before PCR overamplification takes place. The latter manifests on electrophoresis as a smear lifting caused by formation of so-called “PCR-bubbles”. The DNA concentration in the reaction tube in a chosen cycle should be in the range of 10–30 ng/ $\mu$ l. If the library smear is not seen on the gel even after seven cycles of amplification, the library preparation procedure should be repeated.

74. Once a PCR cycle has been chosen, mix up four preparative PCR reactions equal to the one mentioned in step 71. Use the program described in step 71 with a selected number of cycles.

Note: because C-TALE includes extensive PCR amplification with the same pair of Illumina primers, we strongly recommend taking precautions (e.g. use isolated space and time to work with post-PCR sequencing libraries) to avoid contamination of pre-PCR samples with previously made libraries. This note particularly relates to contamination of the pre-PCR 3C-seq libraries with enriched ones. Thus, it is better to carry out analytical and preparative PCR on different days.

75. Transfer amplified libraries from four PCR tubes to one clean

- microtube, add 150 µl of AmpureXP beads, mix thoroughly, incubate for 10 min.
76. Place the tube on the magnet for 2 min, discard the supernatant.
  77. Wash beads three times with 300 µl of 80% ethanol, carefully remove drops of ethanol after last wash, air-dry the beads.
  78. Remove the tube from the magnet, resuspend beads in 30 µl of 10 mM Tris-HCl (pH 8.0), incubate for 5 min at RT.
  79. Place the tube on the magnet for 2 min, transfer the supernatant to a fresh tube.
  80. Measure DNA concentration fluorometrically with a Qubit dsDNA High Sensitivity kit according to the manufacturer's instructions.

Note: successful library preparation gives a total of 1–3 µg of amplified 3C-seq library. The material can be utilized in hybridization immediately, otherwise it can be stored at 4 °C for up to several weeks. If hybridization is not performed for a longer period of time, the prepared library can be frozen and thawed when necessary.

### 3.2. Generation of biotinylated baits from BAC DNA

#### 3.2.1. Preparation of forked adapters for probe amplification

81. Prepare BAC-adapter annealing reaction in a PCR tube as follows:

Ingredient	Volume, µl
10 × NEBuffer 2.1	10
BAC oligo 1 (100 µM)	10
BAC oligo 2 (100 µM)	10
milliQ	70

Note: for probe preparation, we use adapters and corresponding primers utilized in 454-sequencing [23]. One can use adapters and primers with any custom sequences suitable for ligation to A-tailed DNA and substantially differing from Illumina TruSeq adapters and primers.

82. Place the tube in a thermal cycler, heat the mix to 98 °C, then gradually cool to 4 °C at a rate of 1 °C per 15 s. Do not warm the tube to RT.
83. Dilute annealed Y-adapters 10 times with cold 10 mM Tris-HCl (pH 8.0), split in 10 aliquots, and store at –20 °C.

#### 3.2.2. BAC DNA sonication

84. Isolate BAC DNA from selected BAC clones, spanning the genomic region of interest.
85. Make an equimolar mixture of isolated BAC DNA to obtain a total amount of 5 µg.
86. Shear the mixture of BAC DNA to a size of 150–500 bp using a sonicator. We exploit the VirSonic 100 (VirTis) sonicator using the following parameters: DNA is dissolved in 500 µl of 1x sonication buffer, shearing is performed on ice with five 30-sec pulses on the “15” power setting, separated with 3-min periods of recovery.

Note: shearing parameters should be chosen depending on utilized instrument model.

87. Centrifuge the solution of sheared DNA for five minutes at maximum speed.
88. Transfer the supernatant to a 30-kDa Amicon column, centrifuge for five minutes at maximum speed.
89. Discard flow-through, add 450 µl of 10 mM Tris-HCl (pH 8.0) to the column, centrifuge for five minutes at maximum speed.
90. Transfer concentrated DNA solution to a new microtube, add 1.5 volumes of AmpureXP beads, mix thoroughly, incubate for 10 min.
91. Place the tube on the magnet for 2 min, discard the supernatant.

92. Wash beads three times with 200 µl of 80% ethanol, carefully remove drops of ethanol after the last wash, air-dry the beads.
93. Remove the tube from the magnet, resuspend beads in 50 µl of 10 mM Tris-HCl (pH 8.0), incubate for 5 min at RT.
94. Place the tube on the magnet for 2 min, transfer the supernatant to a fresh tube.

#### 3.2.3. End repair

95. Prepare a 2x end repair mix as in step 46.
96. Add 50 µl of 2x end repair mix to the sheared BAC DNA from step 94, mix thoroughly.
97. Incubate tube at RT for 30 min to blunt DNA ends
98. Add 150 µl of AmpureXP beads to the reaction tube, mix thoroughly, incubate for 10 min
99. Place the tube on the magnet for 2 min, discard the supernatant.
100. Wash beads three times with 300 µl of 80% ethanol, carefully remove drops of ethanol after the last wash, air-dry the beads.
101. Remove the tube from the magnet, resuspend beads in 50 µl of 10 mM Tris-HCl (pH 8.0), incubate for 5 min at RT.
102. Place the tube on the magnet for 2 min, transfer the supernatant to a fresh tube.

#### 3.2.4. A-tailing

103. Prepare 2x A-tailing mix as in step 54.
104. Add 50 µl of 2x A-tailing mix to the DNA solution from step 102, mix thoroughly.
105. Incubate tube at 37 °C for 30 min to add dA to 5' DNA ends.
106. Add 150 µl of AmpureXP beads to the reaction tube, mix thoroughly, incubate for 10 min.
107. Place the tube on the magnet for 2 min, discard the supernatant.
108. Wash beads three times with 300 µl of 80% ethanol, carefully remove drops of ethanol after the last wash, air-dry the beads.
109. Remove the tube from the magnet, resuspend beads in 20 µl of 10 mM Tris-HCl (pH 8.0), incubate for 5 min at RT.
110. Place the tube on the magnet for 2 min, transfer the supernatant to a fresh tube.

#### 3.2.5. Adapter ligation

111. Prepare a 2x adapter ligation mix as in step 62.
112. Add 5 µl of the diluted BAC adapters from step 83 to the dA-tailed DNA solution from step 110.

Note: avoid warming of adapters to RT.

113. Add 25 µl of 2x adapter ligation mix to the DNA solution from the previous step, mix thoroughly.
114. Incubate tube at RT overnight.
115. On the next day, add 75 µl of AmpureXP beads to the reaction tube, mix thoroughly, incubate for 10 min.
116. Place the tube on the magnet for 2 min, discard the supernatant.
117. Wash beads three times with 200 µl of 80% ethanol, carefully remove drops of ethanol after the last wash, air-dry the beads.
118. Remove the tube from the magnet, resuspend beads in 100 µl of 10 mM Tris-HCl (pH 8.0), incubate for 5 min at RT.
119. Place the tube on the magnet for 2 min, transfer the supernatant to a fresh tube.

#### 3.2.6. Generation of baits by PCR

120. To determine the optimal number of PCR cycles generating the maximum amount of baits without overamplification, set up an analytical PCR reaction with a pair of 5' biotinylated primers as follows:

Ingredient	volume per 1 reaction, $\mu$ l
KAPA HiFi 5x buffer	5
dNTP mix (10 mM)	0.75
BAC-for-bio primer (10 $\mu$ M)	1.25
BAC-rev-bio primer (10 $\mu$ M)	1.25
DNA solution from step 119	2
KAPA HiFi HotStart polymerase (1 U/ $\mu$ l)	1
milliQ	13.75

Use the following PCR program:

step	temperature, °C	time, s
1	95	300
repeat steps 2–4 12 times		
2	98	20
3	60	15
4	72	20
5	72	180

121. Take 3  $\mu$ l aliquots of PCR reaction at the end of the 8th, 10th, and 12th cycle.
122. Run the taken samples on 1% agarose gel with 1  $\mu$ l of 50-bp DNA ladder to choose the optimal PCR cycle.

Note: we typically use cycles in which the smear of amplified probes is readily seen in the gel, but before PCR overamplification takes place. DNA concentration in the reaction tube in a chosen cycle should be in the range of 10–30 ng/ $\mu$ l. If the probe smear is not observed in the gel even after 12 cycles of amplification, the bait preparation procedure should be repeated.

123. Once a PCR cycle has been chosen, mix up several preparative PCR reactions equal to the one mentioned in step 120. Use the program described in step 120 with optimized number of cycles.

Note: A quantity of probes sufficient for two subsequent C-TALE hybridizations (greater than 2  $\mu$ g) is usually generated in 5–8 such PCR reactions. If one wants to carry out more hybridizations with the same probe-set at once, one can scale up the number of reactions.

124. Transfer amplified libraries from PCR tubes to a clean microtube, add 1.5 volumes of AmpureXP beads, mix thoroughly, incubate for 10 min.
125. Place the tube on the magnet for 2 min, discard the supernatant.
126. Wash beads three times with 500  $\mu$ l of 80% ethanol, carefully remove drops of ethanol after the last wash, air-dry the beads.
127. Remove the tube from the magnet, resuspend beads in 60  $\mu$ l of 10 mM Tris-HCl (pH 8.0), incubate for 5 min at RT.
128. Place the tube on the magnet for 2 min, transfer the supernatant to a fresh tube.
129. Measure DNA concentration fluorometrically with a Qubit dsDNA High Sensitivity kit according to the manufacturer's instructions.

Note: Amplified probes can be stored at 4 °C for up to one week. The remaining pre-amplified probe library can be stored at –70 °C for an unlimited period of time, thawed, and amplified as required.

### 3.3. Target enrichment

#### 3.3.1. Preparation of hybridization buffer, “3C-seq” and “probe” mixes

130. Prepare 2x hybridization buffer as follows:

Ingredient	Volume, $\mu$ l
20 × SSPE buffer	250
50 × Denhardt's solution	100
EDTA (0.5 M)	10
SDS 20%	5
milliQ	135

Note: warm mixture to 65 °C in the ThermoMixer before addition of 20 × SSPE to avoid precipitate formation.

131. Place microtube with 2x hybridization buffer in the ThermoMixer at 65 °C in close proximity to the thermal cycler in which annealing will be carried out.

132. Prepare two mixes in PCR tubes as follows:

“3C-seq” mix		“Probe” mix	
Ingredient	Amount	Ingredient	Amount
3C-seq library or libraries (from step 80)	1 $\mu$ g totally	Probes (from step 129)	1 $\mu$ g totally
Sheared salmon DNA(1 $\mu$ g/ $\mu$ l solution in milliQ)	3 $\mu$ l	Cot-1 DNA of appropriate species (1 $\mu$ g/ $\mu$ l)	3 $\mu$ l
Illumina Uni oligo (10 $\mu$ M)	2 $\mu$ l	BAC oligo 1 (10 $\mu$ M)	2 $\mu$ l
Illumina PE 2.0 primer (10 $\mu$ M)	2 $\mu$ l	BAC-rev primer (10 $\mu$ M)	2 $\mu$ l
milliQ	to 30 $\mu$ l	milliQ	to 30 $\mu$ l

Note: up to five 3C-seq libraries intended to be enriched with ligation products of the same region can be mixed at this step in one “3C-seq” mix (given that they are differentially indexed at step 63). If more than five libraries are to be enriched, carry out several hybridization reactions to avoid complexity issues.

Note: if Cot-1 DNA is not available for studied species, one can use sheared genomic DNA instead.

#### 3.3.2. Melting of mixes and annealing of probes for 3C-seq libraries

133. Place the PCR tubes with “sequencing” and “probe” mixes in the thermal cycler, start the following program:

step	temperature, °C	time, s
1	95	300
2	65	$\infty$

134. After melting at 95 °C and ten minutes at 65 °C, add to each mix 30  $\mu$ l of hot (65 °C) 2x hybridization buffer (see step 131). Incubate mixes for one hour at 65 °C to block sticky sequences.
135. Transfer carefully 60  $\mu$ l of “probe” mix to the tube with the “3C-seq” mix, mix thoroughly by pipetting and incubate at 65 °C for another 20–50 h to allow annealing of probes to complementary sequences in 3C-seq libraries.

#### 3.3.3. Biotin pull-down and washing out non-hybridized part of 3C-seq library

136. Place 1 ml of HS buffer in ThermoMixer to warm it to 65 °C.
137. Wash 15  $\mu$ l of streptavidin-coated magnetic beads (Dynabeads MyOne C1) in 200  $\mu$ l of WB buffer twice on the magnet, resuspend them finally in 300  $\mu$ l of WB buffer.
138. Transfer 120  $\mu$ l of hybridization reaction (from step 135) to the tube with magnetic beads in WB buffer, mix thoroughly, incubate for 30 min with occasional vortexing.
139. Place the tube on the magnet for 2 min, discard the supernatant.
140. Resuspend beads in 500  $\mu$ l of LS buffer, incubate for 15 min with occasional vortexing.
141. Place the tube on the magnet for 2 min, discard the supernatant

Note: because the targeted region of the genome in C-TALE is comparatively small (less than 0.1% of the genome) and the hybridization reaction is not particularly efficient, a very small proportion of the initial 3C-seq library is bound to streptavidin beads. Therefore, to avoid contamination with previously made Illumina libraries, we recommend moving to a library-free space (see note for step 74) to perform subsequent washing steps.

142. Resuspend beads in 300 µl of prewarmed HS buffer from step 138, place the tube in the ThermoMixer at 65 °C with shaking (1,000 rpm) for ten minutes.
143. Place the tube on the magnet for 2 min, discard the supernatant.
144. Repeat steps 142–143 two more times.
145. Resuspend beads in 200 µl of S buffer, transfer them to a clean, low-binding tube.

Note: hereafter, use low-binding tips to manipulate beads.

146. Place the tube on the magnet for 2 min, discard the supernatant.
147. Wash the beads with 200 µl of S buffer two more times, not displacing the tube from the magnet.
148. Finally, resuspend beads in 27 µl of 10 mM Tris-HCl (pH 8.0).

### 3.3.4. PCR amplification of singly enriched 3C-seq libraries

149. To determine the required number of PCR cycles to generate an amount of singly enriched 3C-seq library sufficient for second hybridization, set up an analytical PCR reaction as in step 71. Use the suspension of beads from step 148 as a template and 19 cycles of amplification.

Note: resuspend beads by pipetting before adding them to the PCR reaction.

150. Take a 3-µl aliquot of the PCR reaction at the end of elongation of the 15th, 17th, and 19th cycles.
151. Run taken samples on 1% agarose gel with 50-bp DNA ladder to choose the optimal PCR cycle.

Note: If the enriched library smear is not observed in the gel even after 19 cycles of amplification, the hybridization procedure should be repeated.

152. Once a PCR cycle has been chosen, mix up four preparative PCR reactions as in step 149. Use the PCR program with an optimized number of cycles.
153. Transfer amplified libraries from four PCR tubes to one clean microtube, add 150 µl of AmpureXP beads, mix thoroughly, incubate for 10 min.
154. Place the tube on the magnet for 2 min, discard the supernatant.
155. Wash beads three times with 300 µl of 80% ethanol, carefully remove drops of ethanol after the last wash, air-dry the beads.
156. Remove the tube from the magnet, resuspend beads in 30 µl of 10 mM Tris-HCl (pH 8.0), incubate for 5 min at RT.
157. Place the tube on the magnet for 2 min, transfer the supernatant to a fresh tube.
158. Measure DNA concentration fluorometrically with a Qubit High sensitivity kit according to the manufacturer's instructions.
159. Note: a successful library enrichment gives a total of 1–3 µg of singly enriched 3C-seq library. The material can be utilized in a second hybridization immediately, otherwise it can be stored at + 4 °C for up to several weeks.

### 3.3.5. Second round of enrichment

159. Repeat steps 130–158 with the following modifications:

- a) use the singly enriched 3C-seq library instead of the 3C-seq library during "3C-seq" mix preparation (step 132)
- b) check the 7th, 9th, and 11th cycles in the analytical PCR (step 150)
- c) use two PCR reactions to finally amplify the enriched library (step 152), proportionally scale down all volumes in subsequent steps (steps 153–156)

DNA purified after final preparative PCR is ready for high-throughput sequencing.

## 3.4. Equipment and reagents required

### 3.4.1. Specific equipment

- Thermomixer with cooling (ThermoMixer C, Eppendorf, cat.# 5382000252)
- Qubit fluorometer (Invitrogen, cat.# Q32857)
- Sonicator (VirSonic 100, VirTis)
- Amicon Ultra – 0.5 ml 30 K centrifugal filters (Millipore, UFC503096)
- Magnetic bead separator (DynaMag-2, Invitrogen, cat.# 12321D)
- Low retention microtubes (Eppendorf, cat.# 022431048)
- Low retention tips (Eppendorf, cat.# 022493022)

### 3.4.2. Reagents and enzymes

- Formaldehyde, 37% (Sigma-Aldrich, cat.# 252549)
- 100x Protease Inhibitor Cocktail (Bimake, cat.# B14001)
- 10x NEBuffer 2 (NEB, cat.# B7002S)
- 10x NEBuffer 2.1 (NEB, cat.# B7202S)
- 10x CutSmart buffer (NEB, cat.# B7204S)
- 10x NEBuffer DpnII (NEB, cat.# B0543S)
- 10x T4 DNA ligase buffer (ThermoScientific, cat.# B69)
- Phenol saturated, pH 8.0 (Sigma-Aldrich, cat.# P4557)
- Glycogen, 20 mg/ml (ThermoScientific, cat.# R0561)
- Qubit dsDNA BR Assay Kit (Invitrogen, cat.# Q32853)
- Qubit dsDNA HS Assay Kit (Invitrogen, cat.# Q32854)
- GeneRuler DNA Ladder Mix (ThermoScientific, cat.# SM0331)
- GeneRuler 50 bp DNA Ladder Mix (ThermoScientific, cat.# SM0371)
- AMPure XP beads (Beckman Coulter, cat.# A63881)
- dATP, 100 mM (Sileks, cat.# N0110)
- dCTP, 100 mM (Sileks, cat.# N0210)
- dGTP, 100 mM (Sileks, cat.# N0310)
- dTTP, 100 mM (Sileks, cat.# N0410)
- KAPA HiFi HotStart PCR Kit (Roche, cat.# 07958897001)
- Illumina TruSeq single-indexed adapters (Illumina, cat.# 20,020,492 and 20020493)
- DNA from salmon testes (Sigma-Aldrich, cat.# D1626)
- Human Cot-1 DNA, 1 mg/ml (Invitrogen, cat.# 15279011)
- MboI, 25U/ul (NEB, cat.# R0147M)
- DpnII, 50 U/ul (NEB, cat.# R0543M)
- NlaIII, 10 U/ul (NEB, cat.# R0125L)
- T4 DNA ligase, 5 U/ul (ThermoScientific, cat.# EL0012)
- Proteinase K, 20 mg/ml (ThermoScientific, cat.# EO0491)
- T4 Polynucleotide Kinase, 10 U/ul (NEB, cat.# M0201L)
- T4 DNA polymerase, 3 U/ul (NEB, cat.# M0203L)
- Klenow Fragment of DNA polymerase I, 5 U/ul (NEB, cat.# M0210L)
- Klenow Fragment 3'→5' exo-, 5 U/ul (NEB, cat.# M0212L)
- Streptavidin-coated magnetic beads Dynabeads MyOne C1 (Invitrogen, 65001)

### 3.4.3. Buffers and mixes

- 3C buffer (store at + 4 °C)

**Table 1**

Comparison of different published protocols of in-solution hybridization of 3C-seq/Hi-C libraries with biotinylated probes in terms of enrichment efficiency. All valid read pairs in which at least one end maps to region of interest are considered as “on-target” pairs. \*Only valid read pairs with both ends mapping to regions of interest are taken into account; \*\*only numbers for subsampled data sets of 30,009,111 read pairs are shown because efficiency of enrichment is substantially masked by PCR duplicates in entire libraries.

Reference	Method	Region size, Mb	% unique valid pairs of total reads	% on-target pairs of unique valid pairs	% on-target pairs of total reads	Bait library cost
Current publication	C-TALE	1–1,3	10–20%	30–50%	3–10%	\$
[15]	Hi-C <sup>2</sup>	1,9–2	50–70%	5–20%*	2,5–15%*	\$\$
[26]	Capture Hi-C	7,5	40–60%	45–90%	25–35%	\$\$\$
[27]	Capture Hi-C	4,7	25–30%**	35–70%**	13–20%**	\$\$\$
[14]	T2C	2,1	5–10%	45–85%	2–10%	\$\$\$
[13]	Capture Hi-C	3,3	45–65%	7,5–15%	5–7%	\$\$\$

- o 10 mM Hepes, pH 8.0
- o 145 mM NaCl
- o 0.3 mM MgCl<sub>2</sub>
- Lysis buffer (store at + 4 °C)
- o 10 mM Tris-HCl, pH 8.0
- o 10 mM NaCl
- o 0.5% NP-40
- o 1 × PICs (add directly before usage)
- 2 × sonication buffer (store at RT)
- o 50 mM Tris-HCl, pH 8.0
- o 20 mM EDTA
- o 0.2% SDS
- 20 × SSPE buffer (store at RT)
- o 0.2 M NaH<sub>2</sub>PO<sub>4</sub>, pH 7.5
- o 3 M NaCl
- o 20 mM EDTA
- 50 × Denhardt's solution (store at + 4 °C)
- o 1% Ficoll (type 400)
- o 1% polyvinylpyrrolidone
- o 1% BSA
- 20 × SSC buffer (store at RT)
- o 0.3 M sodium citrate
- o 3 M NaCl
- WB buffer (store at + 4 °C)
- o 10 mM Tris-HCl, pH 7.5
- o 1 M NaCl
- o 1 mM EDTA
- LS buffer (store at + 4 °C)
- o 1 × SSC buffer
- o 0.1% SDS
- HS buffer (store at + 4 °C)
- o 0.1 × SSC buffer
- o 0.1% SDS
- S buffer (store at + 4 °C)
- o 10 mM Tris-HCl, pH 8.0
- o 10 mM NaCl

#### 3.4.4. Sequences of custom oligos

Illumina-for primer	5'AATGATAACGGCGACCACCGAGAT 3'
Illumina-rev primer	5'CAAGCAGAACAGCCGCATACGA 3'
BAC oligo 1	5'CCATCTCATCCCTCGCTGTCGACTACACTACTCGT 3'
BAC oligo 2	5'PO <sub>4</sub> -CGAGTAGTGTTTCAGCAAGGCACACAGGGGATAGG 3'
BAC-for-bio primer	5'bio-CCATCTCATCCCTCGCTGTC 3'
BAC-rev-bio primer	5'bio-CCTATCCCCGTGTCGCTTG 3'
BAC-rev primer	5'CCTATCCCCGTGTCGCTTG 3'
Illumina Uni oligo	5'AATGATAACGGCGACCACCGAGATCTACACTCTTCCCTAC ACGACGCTCTCCGATC 3'

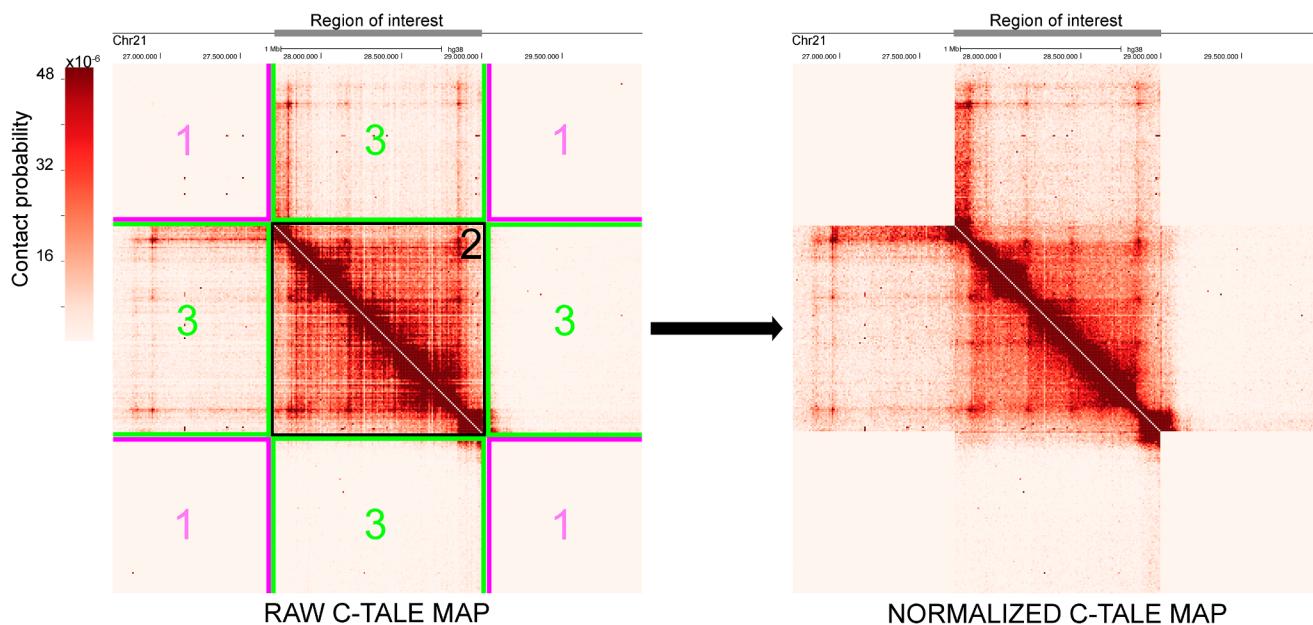
#### 4. Sequencing and data analysis

After the final PCR and clean-up, enriched C-TALE libraries are

subjected to Illumina paired-end sequencing with 50–150 bp read length. At this step, reads from different libraries enriched together in one hybridization reaction are de-multiplied according to ligated sequencing barcodes. In our experience, 20 million pairs of raw reads per megabase of enriched region is more than enough to achieve loop-scale resolution of Hi-C maps. Billions of reads in conventional Hi-C experiment are required to obtain a map with comparable resolution inside the region of interest. If several biological/technical replicates of an experiment have been performed, reads could be distributed among these replicates. Raw reads are then mapped to a reference genome, filtered, and binned by means of one of the available Hi-C data analysis pipelines (reviewed in [24]). We commonly use the hiclib package for these purposes (<https://bitbucket.org/mirnylab/hiclib>). In some cases, it is convenient to map C-TALE reads not to the whole genome, but to the chromosome or even a specific locus carrying the targeted region, an approach which can substantially speed up data analysis. Several replicates for each experiment can be merged at this step. Pearson's correlation coefficient for biological replicates is most frequently estimated prior to merging to ensure a high level of similarity.

Due to the fact that C-TALE libraries are not enriched with ligation junctions, a substantial part of read pairs (more than 75%) map to the same restriction fragment, do not carry any information about spatial proximity, and are therefore discarded (see Table 1). Of the remaining unique valid pairs (which comprise approximately 10–20% of raw reads), 25–50% typically map to the targeted region of the genome with at least one end. Less than 10% of unique valid pairs may indicate low efficiency of the restriction enzyme digestion (high proportion of the same fragment reads) or low library complexity (high proportion of PCR duplicates), which can arise from inefficient 3C-seq library preparation or inefficient capture in the enrichment step. A low proportion (less than 25%) of on-target reads points to issues with the enrichment step of the protocol. We have observed that the percentage of on-target reads depends on the total length of the enriched genomic region(s): the larger a targeted region is, the higher the fraction of reads in the final C-TALE library is localized to this region.

Raw Hi-C matrices are distorted by many intrinsic and technical biases of the C-TALE procedure. Uneven distribution of restriction sites inside genomic bins, biases of PCR amplification and hybridization efficiency, and low mappability of some genomic regions are among them. We devised a rather simple algorithm to correct for these biases applicable specifically for C-TALE Hi-C maps. This algorithm generally stems from the idea that in an unbiased Hi-C matrix, each genomic bin should be equally covered with ligation products. For normalization of whole-genome Hi-C data, this idea has been embodied as the ICE procedure [25]. Hi-C matrices in C-TALE can be subdivided into three distinct sections (Fig. 4): 1) entries of the matrix corresponding to interactions between genomic loci both of which are outside of the targeted region; 2) entries of the matrix corresponding to interactions between genomic loci, both of which are inside the targeted region (these entries are localized in the square on the diagonal axis of the matrix); and 3) entries of the matrix corresponding to interactions



**Fig. 4.** Hi-C heatmap normalization in C-TALE. A raw C-TALE heatmap of 1.3 Mb region on chromosome 21 in HeLa cells binned at 10-kb matrix resolution is shown on the left. Section 1 is outlined in purple, Section 2 in black, and Section 3 in green. The same heatmap after Section 3 multiplication by 1.54 and iterative coverage normalization (see supplementary data for details) is depicted on the right side of the figure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

between genomic loci, only one of each is inside the targeted region (these entries are localized on the Hi-C matrix in the cross, emanating from the square in which the entries of the previous part reside). First of all, entries from Section 1 are discarded from further analysis. Then, each entry from Section 3 is multiplied by a normalization coefficient to correct for less efficient enrichment of these parts of the map. This normalization step is designed to compensate for less effective hybridization of ligation products from Section 3 which are only half covered with C-TALE probes. By comparing C-TALE matrices with conventional Hi-C matrices from corresponding cell lines, we estimated that in general, Section 3 is approximately 1.5 times less efficiently enriched than Section 2, and we therefore typically use 1.5 as the normalization coefficient (see [supplementary data](#)). Finally, we iteratively normalize the coverage of each bin inside the targeted region (Fig. 4; for the detailed algorithm, see [supplementary data](#)).

Once the C-TALE data have been normalized, specific features of the spatial folding of the targeted genomic region can be analyzed in a variety of ways, according to the goals of each particular study. One may wish to annotate self-interacting domains or topological domains inside the targeted region with the use of different tools (Fig. 5). Loops with at least one anchor inside the targeted region can be annotated as well. The position of domain borders and loop anchors can be compared with genome-wide ChIP-seq profiles for different histone modifications and DNA-interacting proteins including known architectural proteins, such as CTCF, cohesin/condensin subunits, and YY1 (Fig. 5). A genome-wide spatial proximity profile for each bin inside the targeted region can also be retrieved from C-TALE data as virtual 4C curves. A subsequent search of significant interactions of the chosen specific bait bins (e.g. bins containing promoters) can be performed (Fig. 5).

## 5. Comparison of C-TALE enrichment efficiency with similar techniques

Protocols for the targeted enrichment of Hi-C/3C-seq libraries with ligation products from specific genomic regions aim to reduce sequencing depth for obtaining high-quality data pertaining to the spatial folding of these regions. A key indicator of the efficiency of such protocols is the proportion of raw sequencing reads contributing to the

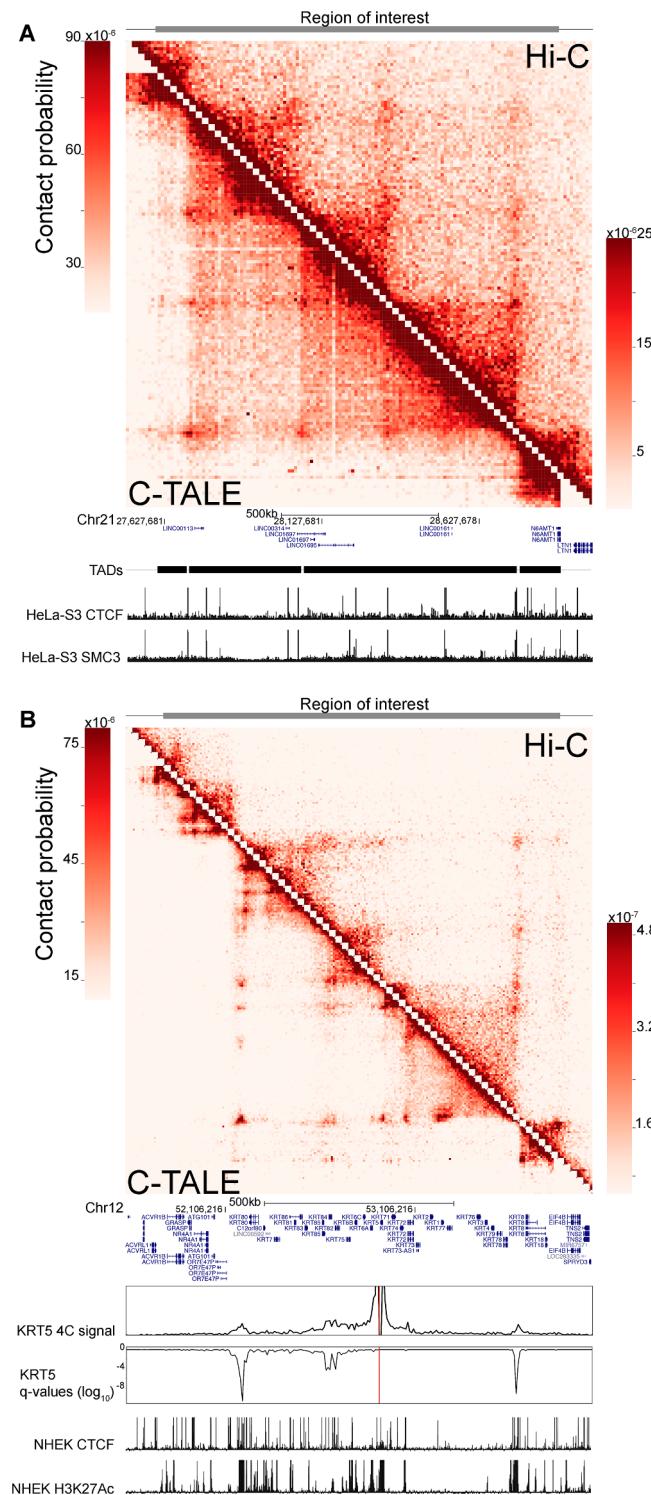
building of Hi-C maps inside the regions of interest (sections 2 and 3 of the Hi-C map from the previous paragraph). In Table 1, we attempted to summarize the available data on the efficiency of C-TALE and other published enrichment protocols. As one can see, the ultimate efficiency of C-TALE (3–10% of reads) is close to the efficiency of methods based on commercially purchased probe pools. The main disadvantage of C-TALE in comparison to other techniques (except T2C) is a low percentage of valid pairs (10–20%) due to the absence of ligation junction capture steps or owing to a short length of the targeted region. Indeed, an almost identical Capture Hi-C protocol applied for larger regions shows much higher overall efficiency. Hence, we anticipate that increasing the targeted region length in C-TALE will also lead to a growth of enrichment efficiency.

## 6. Conclusions

Integration of C-methods with next generation sequencing in Hi-C allowed data-driven interrogation of genomic DNA folding with unprecedented resolution and throughput [7]. However, highly deep sequencing is required to study the most functionally relevant features of DNA folding in vertebrates — chromatin loops. The vast majority of looping interactions occur between pairs of DNA sites divided by no more than several hundred thousands of base pairs inside so-called topologically associated domains (TADs) [7,28]. To address a variety of issues concerning interactions inside restricted genomic regions, it is reasonable to choose many-versus-many (5C) and many-versus-all (Capture Hi-C, T2C, HiC<sup>2</sup>) versions of 3C technologies. One example of an experimental design in which such approaches were exceptionally useful is the study of DNA folding and its link to transcription regulation inside domains of coregulated paralogous genes such as domains of globin, keratin, and Hox genes [8,29–32]. Another issue efficiently addressed with the use of many-versus-all C-methods is the elucidation of universal mechanisms underlying the formation of specific chromatin loop types (e.g. CTCF-mediated loops) in genomes of vertebrates. The relative low cost of these methods has enabled their usage for evaluation of the influence of various experimental impacts on the stability of looping structures: physico-chemical conditions, drug treatments, genetic and epigenetic perturbations [15]. Finally, in the

last few years, the practice of employing many-versus-all C-methods for prioritization of SNP in GWAS regions has gained popularity [33]. The capacity of many-versus-all 3C technologies to detect spatial interactions with loci located outside the targeted regions is of particular importance for such GWAS-deciphering research [13,26,27,34].

C-TALE, a new variant of the many-versus-all C-methods that we have developed, enables analysis of DNA folding at the loop scale at a sequencing depth of approximately 10 million read pairs per 500 kb of a targeted genomic region. Despite the fact that C-TALE has somewhat



**Fig. 5.** Comparison of C-TALE and Hi-C heatmaps and primary bioinformatic analysis of C-TALE data. (A) Normalized C-TALE heatmap of studied region on chromosome 21 in HeLa cells (lower left) and corresponding fragment of conventional Hi-C heatmap in the same cells [7] (upper right) along with annotation of topologically associated domains from C-TALE matrix\*. Heatmaps are binned at 10-kb resolution. ChIP-seq profiles for CTCF and cohesin subunit SMC3 in HeLa S3 are presented below. (B) Normalized C-TALE heatmap of keratin gene domain in normal human keratinocytes (lower left) and corresponding fragment of conventional Hi-C heatmap in NHEK cells [7] (upper right) along with virtual 4C and interaction significance profile for bin containing Ker5 gene (red vertical bar). Both profiles were generated from C-TALE data\*. Heatmaps are binned at 5-kb resolution. ChIP-seq profiles for CTCF and H3K27ac of corresponding region in NHEK cells are presented below. All represented ChIP-seq profiles were extracted from ENCODE. \*See supplementary data for detailed algorithm. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

lower enrichment efficiency than other many-versus-all C-methods exploiting pools of synthetic baits, C-TALE is highly competitive by means of its affordability and design flexibility. These advantages of C-TALE can make it a convenient tool for obtaining high-resolution C-data for genomic regions of interest, data which is becoming increasingly useful in both fundamental and translational research.

## Acknowledgements

This work was supported by the Russian Science Foundation (grant 18-14-00011). S.V.U. and A.V.L. were supported by the Russian Foundation for Basic Research (project 18-34-20104).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ymeth.2019.06.022>.

## References

- [1] W.A. Bickmore, Annual review of genomics and human genetics, 14, 2013, 67–84.
- [2] B. Bonev, G. Cavalli, *Nat Rev Genet* 17 (2016) 772.
- [3] A. Denker, W. de Laat, *Genes Dev* 30 (2016) 1357–1382.
- [4] J. Dekker, K. Rippe, M. Dekker, N. Kleckner, *Science* 295 (2002) 1306–1311.
- [5] K.E. Cullen, M.P. Kladde, M.A. Seyfred, *Science* 261 (1993) 203–206.
- [6] E. Lieberman-Aiden, N.L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, et al., *Science* 326 (2009) 289–293.
- [7] S.S. Rao, M.H. Huntley, N.C. Durand, E.K. Stamenova, I.D. Bochkov, J.T. Robinson, et al., *Cell* 159 (2014) 1665–1680.
- [8] J. Dostie, T.A. Richmond, R.A. Arnaout, R.R. Selzer, W.L. Lee, T.A. Honan, et al., *Genome Res* 16 (2006) 1299–1309.
- [9] J.H. Kim, K.R. Titus, W. Gong, J.A. Beagan, Z. Cao, J.E. Phillips-Cremins, *Methods* 142 (2018) 39–46.
- [10] A.D. Schmitt, M. Hu, B. Ren, *Nat Rev Mol Cell Biol* 17 (2016) 743–755.
- [11] M.J. Clark, R. Chen, H.Y. Lam, K.J. Karczewski, R. Chen, G. Euskirchen, A.J. Butte, M. Snyder, *Nat Biotechnol* 29 (2011) 908–914.
- [12] C.D. Rodley, F. Bertels, B. Jones, J.M. O’Sullivan, *Fung Genet Biol: FG & B* 46 (2009) 879–886.
- [13] N.H. Dryden, L.R. Broome, F. Dudbridge, N. Johnson, N. Orr, S. Schoenfelder, et al., *Genome Res* 24 (2014) 1854–1868.
- [14] P. Kolovos, H.J. van der Werken, N. Kepper, J. Zuin, R.W. Brouwer, C.E. Kockx, et al., *Epigenet Chromat* 7 (2014) 10.
- [15] A.L. Sanborn, S.S. Rao, S.C. Huang, N.C. Durand, M.H. Huntley, A.I. Jewett, et al., *Proc Natl Acad Sci USA* 112 (2015) E6456–6465.
- [16] S.V. Ulianov, A.A. Galitsyna, I.M. Flyamer, A.K. Golov, E.E. Khrameeva, M.V. Imakaev, et al., *Epigenet Chromat* 10 (2017) 35.
- [17] J.O. Davies, J.M. Telenius, S.J. McGowan, N.A. Roberts, S. Taylor, D.R. Higgs, J.R. Hughes, *Nat Methods* 13 (2016) 74–80.
- [18] T. Nagano, C. Varma, S. Schoenfelder, B.M. Javierre, S.W. Wingett, P. Fraser, *Genome Biol* 16 (2015) 175.
- [19] H. Belaghzal, J. Dekker, J.H. Gibcus, *Methods* 123 (2017) 56–65.
- [20] S. Bashirades, R. Veile, C. Helms, E.R. Mardis, A.M. Bowcock, M. Lovett, *Nat Methods* 2 (2005) 63–69.
- [21] E. Yigit, Q. Zhang, L. Xi, D. Grilley, J. Widom, J.P. Wang, A. Rao, M.E. Pipkin, *Nucleic Acids Res* 41 (2013) e87.
- [22] S. Selvaraj, A.D. Schmitt, J.R. Dixon, B. Ren, *BMC Genom* 16 (2015) 900.
- [23] Z. Zheng, A. Advani, O. Melefors, S. Glavas, H. Nordstrom, W. Ye, L. Engstrand, A.F. Andersson, *Nat Protoc* 6 (2011) 1367–1376.

- [24] F. Ay, W.S. Noble, *Genome Biol* 16 (2015) 183.
- [25] M. Imakaev, G. Fudenberg, R.P. McCord, N. Naumova, A. Goloborodko, B.R. Lajoie, J. Dekker, L.A. Mirny, *Nat Methods* 9 (2012) 999–1003.
- [26] P. Martin, A. McGovern, G. Orozco, K. Duffus, A. Yarwood, S. Schoenfelder, et al., *Nat Communica* 6 (2015) 10069.
- [27] R.G. Jager, M. Migliorini, R. Henrion, H.E. Kandaswamy, A. Speedy, N. Heindl, et al., *Nat commun* 6 (2015) 6178.
- [28] J. Cairns, P. Freire-Pritchett, S.W. Wingett, C. Varnai, A. Dimond, V. Plagnol, et al., *Genome Biol* 17 (2016) 127.
- [29] D. Bau, A. Sanyal, B.R. Lajoie, E. Capriotti, M. Byron, J.B. Lawrence, J. Dekker, M.A. Marti-Renom, *Nat Struct Mol Biol* 18 (2011) 107–114.
- [30] M.A. Ferraiuolo, M. Rousseau, C. Miyamoto, S. Shenker, X.Q. Wang, M. Nadler, M. Blanchette, J. Dostie, *Nucleic Acids Res* 38 (2010) 7472–7484.
- [31] I. Williamson, S. Berlivet, R. Eskeland, S. Boyle, R.S. Illingworth, D. Paquette, J. Dostie, W.A. Bickmore, *Genes Dev* 28 (2014) 2778–2791.
- [32] A.P. Kovina, N.V. Petrova, E.S. Gushchanskaya, K.V. Dolgushin, E.S. Gerasimov, A.A. Galitsyna, et al., *Mol Biol Evol* 34 (2017) 1492–1504.
- [33] Y.G. Tak, P.J. Farnham, *Epigenet Chromat* 8 (2015) 57.
- [34] J.S. Baxter, O.C. Leavy, N.H. Dryden, S. Maguire, N. Johnson, V. Fedele, et al., *Nat Commun* 9 (2018) 1028.