Chapter 15

Unbiased Interrogation of 3D Genome Topology Using Chromosome Conformation Capture Coupled to High-Throughput Sequencing (3C-Seq)

AU1 Rutger W.W. Brouwer, Mirjam C.G.N. van den Hout, Wilfred F.J. van IJcken, Eric Soler, and Ralph Stadhouders

Abstract

The development and widespread implementation of chromosome conformation capture (3C) technology has allowed unprecedented new insight into how chromosomes are folded in three-dimensional (3D) space. 3C and its derivatives have contributed tremendously to the now widely accepted view that genome topology plays an important role in many major cellular processes, at a chromosome-wide scale, but certainly also at the level of individual genetic loci. A particularly popular application of 3C technology is to study transcriptional regulation, allowing researchers to draw maps of gene regulatory connections beyond the linear genome through addition of the third dimension. In this chapter, we provide a highly detailed protocol describing 3C coupled to high-throughput sequencing (referred to as 3C-Seq, analogous to 4C-Seq reported by others), allowing the unbiased interrogation of genome-wide chromatin interactions with specific genomic regions of interest. Interactions between spatially clustered DNA fragments are revealed by crosslinking the cells with formaldehyde, digesting the genome with a restriction endonuclease and performing a proximity ligation step to link interacting genomic fragments. Next, interactions with a selected DNA fragment are extracted from the 3C library through a second round of digestion and ligation followed by an inverse PCR. The generated products are immediately compatible with high-throughput sequencing, and amplicons from different PCR reactions can easily be multiplexed to dramatically increase throughput. Finally, we provide suggestions for data analysis and visualization.

Key words Chromosome conformation capture (3C), 3C-Seq, 4C, Genome-wide, Long-range gene regulation, Chromatin looping, DNA, Bioinformatics

1 Introduction

Chromosome conformation capture (3C) and high-throughput derivatives (4C, 3C-/4C-Seq, Capture-C, T2C, Hi-C, Capture-Hi-C, ChIA-PET) enable the reconstruction of average topological genome conformations from populations of cells [1, 2]. This knowledge is essential to fully understand gene regulatory principles, as it has been clearly demonstrated that gene transcription is

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcription and Post-Transcription Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_15,

intimately linked to three-dimensional (3D) genome organization. Indeed, gene regulatory elements may interact with their target genes over large genomic distances (hundreds to thousands of kilobases (kb)) via chromatin looping mechanisms (reviewed in [3]) to modulate the transcriptional activity of target promoters. Therefore, defining long-range chromatin interactions and their dynamics across different experimental settings has the potential to yield critical insights into gene regulatory processes. 3C, for example, allows researchers to detect gene distal cis-regulatory elements through their physical co-association with target promoters in large and complex genomes. 3C also represents an essential step toward understanding genotype–phenotype relationships in mammals and the functional impact of common noncoding genomic variants in humans (i.e., [4–9]).

The principle underlying 3C is one of the elegant simplicity. 3C (and all 3C-related technologies) relies on proximity ligation-mediated capture of chromatin interactions (Fig. 1). Typically, formaldehyde crosslinking is used to fix the native 3D genome organization, followed by restriction enzyme digestion of the genome. This way, distal chromatin fragments that were in close proximity in the nuclear space at the time of fixation remain physically linked and can be ligated to each other. Subsequent PCR strategies and deep sequencing approaches are used to detect these ligation products and to reconstruct the 3D chromatin conformation of the cells under investigation.

We have used 3C-sequencing (3C-Seq, analogous to 4C-Seq [10]), a high-throughput 3C derivative to analyze topological conformations of genomic regions of interest with genome-wide coverage. 3C-Seq uses a single chosen genomic region of interest (e.g., enhancer, promoter, domain boundary; referred to as the "viewpoint") and interrogates the whole genome for chromatin co-associations with this region. As such, 3C-Seq by itself is not truly a genome-wide assay as it is focused on chosen genomic regions. To overcome this limitation, we apply 3C-Seq in a multiplexed fashion to enable analysis of dozens of viewpoints in parallel with genome-wide coverage [11].

2 Materials

All materials, reagents, and software listed are divided using the same subheadings as used in Subheading 3 describing the methods. All buffers and solutions described below are prepared using nuclease-free deionized water. Reagents can be stored at room temperature unless stated otherwise.

2.1 Design of a 3C-Seq Experiment

1. Computer with standard restriction analysis/PCR primer design software.

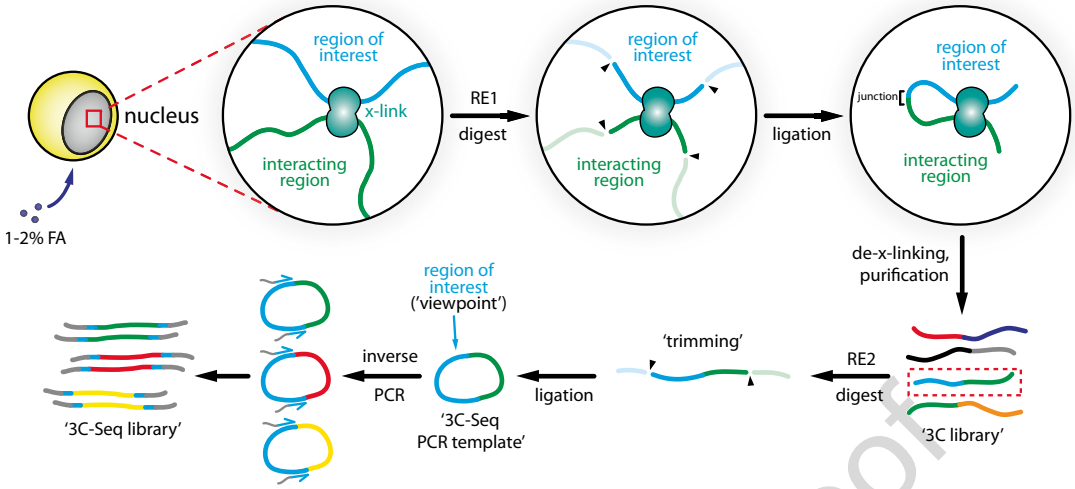

Fig. 1 Principles underlying the 3C-Seq protocol. Depicted is a schematic of the different experimental steps required to obtain a 3C-Seq library. Cells are treated with formaldehyde ("FA"; 1–2 %), causing nuclear architecture to be stably fixed through the formation of protein–protein and protein–DNA crosslinks (shown as an "x-link"). As a result, interacting genomic regions normally separated on the linear chromosome (exemplified by the *blue* and *green* strands) are maintained in close proximity during chromatin fragmentation (digestion with the primary restriction enzyme, "RE1") and proximity ligation. The latter step creates chimeric molecules (the "3C library") with a ligation junction that allows their identification. In 3C-/4C-Seq protocols, 3C libraries are digested a second time (with RE2) to trim the average size of the individual molecules, allowing for efficient PCR amplification after self-circularization through a second ligation. PCR primers are designed on a region of interest (the "viewpoint," depicted in *blue* here) and face outward to amplify all unknown fragments that have been ligated to the viewpoint (represented by the *green*, *red*, and *yellow* fragments). The addition of sequencing adapters to the PCR primers (shown as *gray* overhangs) allows for direct high-throughput sequencing of 3C-Seq libraries

2.2 Crosslinking of Cells and Isolation of Nuclei

1. Mammalian cells of interest (i.e., cultured cell lines or primary cells, FACS purified cells, or freshly isolated tissue samples) (*see* **Note 1**).
2. Cell strainer (40 μM or 70 μM depending on the cell size).
3. Appropriate cell culture medium or 10 % heat-inactivated FCS in PBS (FCS/PBS).
4. 37 % formaldehyde (*Toxic:* handle in chemical fume hood) (*see* **Note 2**).
5. 1-M glycine in water (0.22 μM filtered).
6. Protease inhibitors (e.g., complete protease inhibitor, Roche).
7. Freshly prepared cell lysis buffer (10-mM Tris–HCl, pH 8.0, 10-mM NaCl, 0.2 % NP-40 containing protease inhibitors), store on ice until needed (*see* **Note 3**).

2.3 Generation of the 3C Library: Digestion and Ligation

1. Primary restriction enzyme and appropriate buffer (*see* **Note 4**).
2. 10 % (wt/vol) SDS.
3. 20 % (wt/vol) Triton X-100.

4. Proteinase K (10 mg/ml).
5. Basic agarose gel electrophoresis equipment and reagents.
6. 10× T4 DNA ligation buffer (Roche) (*see* **Note 5**).
7. T4 DNA ligase (highly concentrated, 5 U/μl, Roche).

2.4 De-crosslinking and DNA Purification

1. Phenol/chloroform/isoamyl alcohol (25:24:1), saturated with 100-mM Tris–HCl, pH 8.0 (*Toxic:* handle in chemical fume hood).
2. DNase-free RNase A (10 mg/ml).
3. 2-M sodium acetate, pH 5.6 or 3-M sodium acetate, pH 5.2.
4. 100% and 70% ethanol.
5. 10-mM Tris–HCl, pH 7.5.

2.5 Generation of the 3C-Seq PCR Template: Digestion and Ligation

1. Secondary restriction enzyme and appropriate buffer (*see* **Note 6**).
2. Glycogen (20 mg/μl), molecular biology grade.

2.6 DNA Purification, PCR Amplification, and 3C-Seq Library Preparation

1. Spin column-based DNA purification kit (*see* **Note 7**).
2. DNA quantification system (e.g., spectrophotometric or fluorometric).
3. Expand Long Template PCR System.
4. dNTPs (10 mM).
5. Viewpoint-specific inverse PCR oligonucleotides.
6. High Pure PCR Product Purification kit or AMPure XP beads (Beckman Coulter) (*see* **Note 8**).

2.7 High-Throughput Sequencing and Data Analysis

1. Bioanalyzer and DNA 12000 chip cartridges (Agilent) (*see* **Note 9**).
2. Accurate DNA quantification methodology.
3. Illumina high-throughput sequencing system and required reagents.

2.8 Primary Data Analysis and Downstream Data Analysis (*See* **Note 26**)

1. *SAMtools* available at http://www.htslib.org/ [12].
2. *BEDTools* available at https://github.com/arq5x/bedtools2 [13].
3. *Bowtie* available at http://bowtie-bio.sourceforge.net/ [14].
4. Illumina base-calling software (*bcl2fastq*) available from http://www.illumina.com/.
5. *IGV genome browser* available at http://www.broadinstitute.org/igv/ [15].
6. *Python* available at http://www.python.org/. All simple 3C-Seq analysis scripts have been developed using Python release 2.x.

7. *Pysam* available at https://github.com/pysam-developers/pysam.
8. 3C analysis scripts available at https://github.com/RWWB/simple3Cseq.
9. FastA file with the reference genome sequence.
10. Chromosome sizes for the reference genome (generated from FastA).
11. Bowtie index for the reference genome (generated from FastA).
12. Software packages for in-depth data analysis (e.g., *r3Cseq*, *Basic4CSeq*, *fourSig*, *4Cseqpipe*, *FourCSeq*).

3 Methods

3.1 Design of a 3C-Seq Experiment

1. Determine "bait(s)" or "viewpoint(s)" of interest to be analyzed. Typically, these represent discrete genomic regions defined by a known or suspected regulatory function (e.g., enhancer or promoter)—however, any genomic region is in principle amenable to 3C-Seq analysis.
2. Select a (set of) suitable primary restriction enzyme(s) based on (1) general performance in 3C-type conditions, (2) insensitivity to mammalian CpG methylation, (3) no/very low star activity, (4) excellent (>95%) ligation efficiencies after digestion, and (5) recognition site distribution surrounding the desired viewpoint(s) and the desired resolution (*see* **Note 4** and Fig. 2).
3. Select suitable secondary restriction enzymes based on (1) insensitivity to mammalian CpG methylation, (2) excellent (>95%) ligation efficiencies after digestion, and (3) compatibility with the primary restriction enzyme in generating a suitable fragment for inverse PCR primer design (*see* Fig. 2) (*see* **Note 6**).
4. Design inverse PCR primer pairs on restriction fragment ends of the viewpoint(s) of interest (*see* Fig. 2) (*see* **Note 10**).

3.2 Crosslinking of Cells and Isolation of Nuclei

1. Collect cells or tissues; resuspend in fresh culture medium or FCS/PBS. If required (e.g., when using tissue samples), prepare a single-cell suspension by filtering through a cell strainer (*see* **Note 1**).
2. Determine cell concentration and further dilute the cell suspension as necessary using fresh culture medium or FCS/PBS (*see* **Note 11**). We prefer to crosslink cells at a 1×10^6 cells per ml density to ensure standardized conditions (e.g., 10×10^6 cells in 10 ml of the medium). We have conducted successful 3C-Seq experiments using this protocol starting from $10–20 \times 10^6$ to as little as $1–0.5 \times 10^6$ cells [16–18].

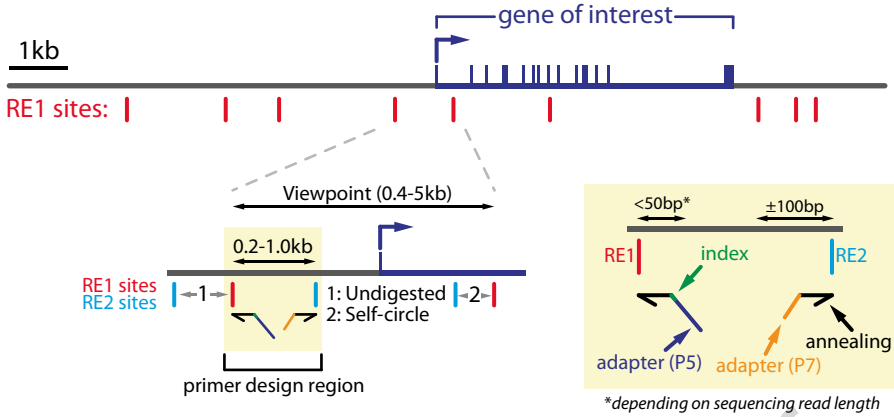

Fig. 2 Primer design considerations. A hypothetical locus of interest is shown to illustrate several aspects of 3C-Seq primer design. Primary restriction enzyme (RE1) sites are scattered around the locus, determining local resolution of the 3C-Seq experiment. Viewpoint fragment length is considered optimal between 0.4 and 5.0 kb and is generally chosen to be located as close as possible to (or preferably overlapping with) the specific genomic element of interest—in this case the promoter region of a chosen gene (the transcription start site is depicted as an *arrow*, exons as *vertical blue lines*). Selected viewpoint fragments are then examined for secondary restriction enzyme (RE2) sites. Primers can be designed at both the 5′- and 3′-ends of the viewpoint fragment (as delineated by the RE1 sites), as long as the final RE1–RE2 fragment is >200 bp (thus ensuring efficient self-circularization) and allows primer design near the RE1 and RE2. In the example, the 5′ RE1–RE2 combination was chosen for primer design (*see* magnification marked in *yellow*). Sequencing is initiated from the RE1 side of the viewpoint; hence, the corresponding primer should contain an optional index and the P5 Illumina sequencing adapter as overhang. This "reading primer" should be positioned as close to the RE1 site as possible (preferably encompassing it) to ensure enough sequence read length is left for aligning the unknown interacting sequence to the genome. The design of the nonreading primer with P7 adapter overhang near the RE2 site allows for more flexibility, as sequencing is not initiated from this side. Inherent to the experimental strategy of 3C-/4C-Seq, two fragments will be amplified with a much higher frequency (also *see* Fig. 3): (1) the "undigested" fragment due to a failure to cut the RE1 site next to the reading primer (the RE1–RE2 fragment directly adjacent to the viewpoint RE1–RE2 fragment) and (2) the "self-circle" fragment due to the self-circularization of the viewpoint fragment in the first round of ligation (the RE1–RE2 fragment at the opposite end of the viewpoint fragment not used for primer design)

3. Add 37% formaldehyde directly to the cell suspensions to a final concentration of 1–2% (*see* **Note 12**). Immediately mix by inverting and incubate at room temperature for 10 min under rotation.
4. Transfer the cells to ice and add glycine to a final concentration of 0.125 M.
5. Immediately centrifuge the crosslinked cells for 8 min at 340 × *g* (4 °C) and remove all supernatant.
6. Gently resuspend the cell pellet in 5 ml of prechilled cell lysis buffer. Incubate for 10–15 min on ice.
7. Centrifuge the mixture for 5 min at 650 × *g* (4 °C) to pellet the nuclei.

8. Pelleted nuclei can be washed with PBS and transferred to 1.5-ml microcentrifuge tubes (e.g., as aliquots), snap-frozen in liquid N_2, and stored at −80 °C for at least a year.

3.3 Generation of the 3C Library: Digestion and Ligation

1. Gently resuspend the nuclei in 0.5 ml of 1.2× restriction buffer. Do not use more than 10×10^6 nuclei per individual reaction (*see* **Note 13**).
2. Place the tubes in a thermomixer at 37 °C and add 15 µl of 10% SDS (final, 0.3%).
3. Incubate at 37 °C for 1 h while shaking at 900 × *g*.
4. Add 50 µl of 20% Triton X-100 (final, 2%) (*see* **Note 14**).
5. Incubate at 37 °C for 1 h while shaking at 900 × *g*.
6. Remove a small (5–10 µl) aliquot from each sample and store it overnight at −20 °C. The DNA extracted from these nuclei will serve as an undigested control sample (*see* **steps 9** and **10**).
7. Add 400 U of the preferred restriction enzyme to the remaining sample and incubate overnight at 37 °C while shaking at 900 × *g* (*see* **Note 15**).
8. Remove a small (5–10 µl) aliquot from each sample. The DNA extracted from these nuclei will determine the digestion efficiency (*see* **steps 9** and **10**).
9. Add 10 µl of proteinase K to the two control aliquots and add 10-mM Tris–HCl, pH 7.5 to a final volume of 100 µl. Incubate at 65 °C for 1 h to reverse formaldehyde crosslinks.
10. Directly run a 20 µl aliquot of the control samples alongside each other on a 0.6% (wt/vol) standard agarose gel. The DNA from undigested control samples should run as a discrete high molecular weight band (>12 kb), while a DNA smear should appear after a successful digestion (Fig. 3) (*see* **Note 16**).
11. If digestion was successful, heat-inactivate the restriction enzyme by incubating at 65 °C for 20 min. If the enzyme cannot be heat-inactivated, add 80 µl of 10% SDS (final, 1.6%) before incubating at 65 °C for 20 min.
12. Transfer the sample to a 50-ml centrifugation tube and add 6.125 ml of 1.15× ligation buffer (*see* **Note 17**).
13. Only if SDS was added in **step 11**, add 375 µl of 20% Triton X-100 (final, 1%) and incubate for 1 h at 37 °C in a water bath.
14. Add 100 U of T4 DNA ligase and incubate at 16 °C for 4 h or overnight.

3.4 De-crosslinking and DNA Purification

1. Add 30 µl of proteinase K and incubate for at least 4 h (overnight is also possible) at 65 °C to de-crosslink the samples.
2. Add 30 µl of RNase A and incubate for 30–45 min at 37 °C.

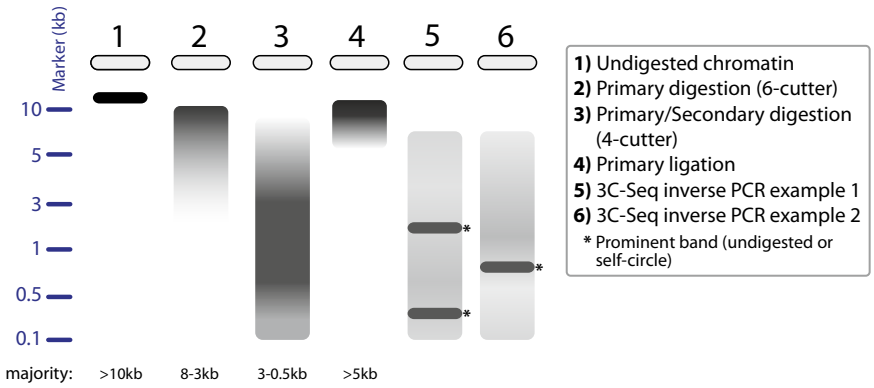

Fig. 3 3C-Seq wet-lab quality control. Artificial representation of typical DNA smears obtained after agarose gel electrophoresis during the various quality control steps of the 3C-Seq protocol (typical average sizes of the majority of DNA fragments are shown below the "gel"). Undigested chromatin (1) runs as a sharp high molecular weight band. After the first digestion (2 + 3), this sharp band should largely dissolve into a downward smear (with an average size depending on whether a 6-bp- or 4-bp-recognizing enzyme was selected). After the first ligation (4), the DNA fragment size returns upward as a compact high molecular weight smear/band. The second round of digest (3)—using a frequent 4-bp-recognizing enzyme—results again in a smear of low molecular weight fragments. The second ligation (the self-circularization step) does not result in visible DNA fragment size changes on the gel. Inverse PCR (5 + 6) typically produces a wide range of DNA fragments (representing detectable interacting fragments) that appear as a smear on the gel. Two prominent bands (the "undigested" and "self-circle" fragments, *see* Fig. 2 for an explanation of their origin) are often readily visible (5). Note that these fragments can be very small (<100 bp) and therefore they do not always appear on the gel (6)

3. Cool the samples to room temperature and add 7 ml of phenol/chloroform/isoamyl alcohol and shake the samples vigorously.
4. Centrifuge for 15 min at 3200 × *g*.
5. Transfer the upper aqueous phase into a new 50-ml tube. Add 7 ml of Milli-Q H_2O and 1.5 ml of sodium acetate and add 35 ml of 100 % ethanol.
6. Mix thoroughly and place the sample at −80 °C for 2–3 h until the liquid is frozen solid.
7. Directly centrifuge the frozen samples for 45 min at 3200 × *g* (4 °C).
8. Remove the supernatant and add 10 ml of 70 % ethanol.
9. Centrifuge the mixture for 15 min at 3200 × *g* (4 °C).
10. Remove the supernatant, air-dry, and dissolve the pellet in 150 µl of 10-mM Tris–HCl, pH 7.5 by incubating for 30 min at 37 °C.
11. Determine ligation efficiency by running a small aliquot (1–5 µl) of 3C material on a 0.6 % (wt/vol) standard agarose gel. A successful ligation should result in a significant shift of the digested DNA upward in the gel (Fig. 3) (*see* **Note 18**).

12. Store the 3C library at −20 °C or proceed with the second round of digestion. If desired, quantitative PCR can be performed on the 3C library to either ensure the presence of chimeric ligation products or validate library quality by probing for a known genomic interaction (*see* **Note 19**).

3.5 Generation of the 3C-Seq PCR Template: Digestion and Ligation

1. Digest the 3C library overnight with the selected four-base recognition restriction enzyme using 50 U of enzyme in a 500-µl total reaction volume. Use buffers and incubation temperatures as recommended in the manufacturer's instructions.
2. Remove a small aliquot (10–20 µl) from the reaction to assess digestion efficiency on a 1.5% (wt/vol) standard agarose gel. The high molecular weight DNA observed after the first ligation should have been digested into a smear of low molecular weight fragments (usually the majority is <1000 bp) (Fig. 3). If digestion is suboptimal, consider repurifying the samples using a phenol–chloroform extraction (as described below) followed by an additional round of overnight digestion.
3. If digestion was successful, heat-inactivate the restriction enzyme by incubating at 65 °C for 20 min. If the enzyme cannot be heat-inactivated, remove the enzyme by adding 500 µl of phenol/chloroform/isoamyl alcohol. Shake vigorously and centrifuge for 15 min at 16,100 × *g*. Transfer the aqueous phase to a new tube and add 2 µl of glycogen, 50 µl of sodium acetate, and 850 µl of 100% ethanol. Mix thoroughly, place the sample at −80 °C until frozen solid, and directly centrifuge the frozen samples for 20 min at 16,100 × *g* (4 °C). Wash the pellet once with 70% ethanol, air-dry, and dissolve in 500 µl of 10-mM Tris–HCl, pH 7.5.
4. Transfer the sample to a 50-ml centrifugation tube. Add 1.4 ml 10× ligation buffer, 200 U of T4 DNA ligase, and water up to 14 ml. Incubate at 16 °C for 4 h or overnight (*see* **Note 20**).

3.6 DNA Purification, PCR Amplification, and 3C-Seq Library Preparation

1. To the 14-ml ligation sample, directly add 14-µl glycogen and a 1/10 volume of sodium acetate, mix the contents, and add 35 ml of 100% ethanol (*see* **Note 21**).
2. Place the tubes at −80 °C for 2–3 h until the liquid is frozen solid.
3. Directly centrifuge the frozen samples for 45 min at 3200 × *g* (4 °C).
4. Remove the supernatant and add 10 ml of 70% ethanol.
5. Centrifuge the mixture for 15 min at 3200 × *g* (4 °C).
6. Remove the supernatant, air-dry, and dissolve the pellet in 150 µl of 10-mM Tris–HCl, pH 7.5 by incubating for 30 min at 37 °C.

7. Purify the DNA using a spin column-based DNA purification kit (QIAquick gel or PCR purification kits work well in our hands) according to the manufacturer's recommendations. Realize that most columns used in such kits have a maximum binding capacity of 10–20 μg of DNA. Use one column when working with $1–2 \times 10^6$ cells, increase the number of columns to two ($2–5 \times 10^6$ cells) or three ($5–10 \times 10^6$ cells) as more cells are used at the start of the protocol.
8. Estimate the DNA concentration of the resulting "3C-Seq PCR template" sample using spectrophotometric (e.g., Nanodrop) or fluorometric (e.g., Qubit) measurements.
9. Store the 3C-Seq PCR template at −20 °C or proceed with inverse PCR.
10. PCR primer sets designed in Subheading 3.1 can now be tested on dilution series of the 3C-Seq PCR template. Ensure that primer pairs yield reproducible fragment smears and amplify in a linear fashion (*see* **Note 22** and Fig. 3).
11. After assessing primer pair quality, perform several PCR reactions (we generally amplify the equivalent of 500–1000-ng input DNA per bait fragment, using the Roche Expand Long Template PCR System) using PCR primers containing Illumina P5/P7 adapters as overhang. The PCR reaction setup and program are indicated below. Ensure that the amount of input 3C-Seq PCR template used allows for a linear and reproducible PCR reaction. We generally do not exceed 200 ng of input material.

PCR reaction mix

5 μl of 1× buffer I

1 μl of 10 mM dNTPs

25-pmol forward primer

25-pmol reverse primer

0.75-μl polymerase mix (5 U/μl)

25–200 ng of 3C-Seq PCR template

H_2O up to 50 μl.

PCR program

94 °C—2 min

30 cycles: 94 °C, 15 s; [primer-specific Tm] °C, 1 min; 68 °C, 3 min

68 °C—7 min

12. Verify PCR success by analyzing a small aliquot (5–10 μl) of each reaction on a 1.5 % (wt/vol) standard agarose gel (Fig. 3).
13. Pool all successful reactions from the same bait fragment and purify the DNA using the High Pure PCR Product Purification kit or AMPure XP beads (*see* **Note 23**). Elute the columns/beads with 40 μl of elution buffer.

14. Verify the purification procedure success by running a 5-μl aliquot on a 1.5% (wt/vol) standard agarose gel and measuring sample purity using spectrophotometric measurements. We adhere to similar guidelines as those published by van de Werken et al. [19], with $A_{260/A280}$ values between 1.8 and 2.0 and $A_{260/A230}$ values >1.5 considered compatible with Illumina sequencing procedures. Also ensure that primer dimers (70–120 bp) have been successfully removed.
15. 3C-Seq libraries can be stored at −20 °C for at least 6 months and can be directly sequenced using Illumina high-throughput sequencing equipment (*see* **Note 24**).

3.7 High-Throughput Sequencing

1. Estimating the average fragment size of 3C-Seq libraries can be challenging. We quantify the average fragment size of the individual 3C-Seq libraries on an Agilent Bioanalyzer (ensure that the cartridge and kit used allow detection of DNA fragments up to 4 kb). If PCR smears show one or two dominant fragments, this size (or the average of the two) is generally a good approximation. Otherwise, a "smear analysis" quantification using the Bioanalyzer software provides a reliable size estimate.
2. Quantify 3C-Seq library DNA concentrations using a reliable library quantification strategy. We have successfully used Bioanalyzer, Qubit, and quantitative PCR (i.e., KAPA Library Quantification Kit) strategies.
3. Determine DNA molarities of the individual 3C-Seq libraries using the average size estimate and DNA concentration values. The following formula can be used: (library concentration [in ng/μl] × 10^6)/(650 D × library size [in bp]) = *x*nM 3C-Seq library.
4. Create different pools of multiple 3C-Seq libraries for Illumina sequencing by pooling equimolar amounts of individual libraries in a single tube (*see* **Note 25**).
5. Proceed with single-read sequencing as described by the manufacturer. Depending on the primer design, either a 50- or 100-bp read length is sufficient for accurate alignment of the reads (*see* **Note 10**). To enable custom de-multiplexing using the viewpoint primer at the beginning of the 3C-/4C-Seq sequence reads, do not specify the "I7_index_ID" and "index" columns in the SampleSheet.csv loaded upon starting the sequencer.

3.8 Primary Data Analysis

1. Proceed with standard sequencing data processing methods from Illumina (*bcl2fastq*) to retrieve the reads in FastQ format. After this procedure, all data will be located (by default) in the data run folder, in the Data/Intensities/BaseCalls/subfolder, organized as one gzipped FastQ file per lane.

```
> bcl2fastq --runfolder-dir /path/to/run-
folder/
```

```
> cd /path/to/runfolder/Data/Intensities/
BaseCalls; gunzip *.fastq.gz
```

2. Obtain the reads from the samples by matching the 5′-ends to the viewpoint sequence (*see* **Note 27**).

```
> demultiplex.py -i allreads.fastq -s view-
point_sequence -o sample.fastq
```

3. Next, remove the viewpoint and the restriction site from the 5′-end of the reads to keep only the interacting target sequence. The remaining sequence should be trimmed to a fixed length from the 3′-end of the read. Differences in read length between different viewpoints can introduce variation between experiments and should thus be avoided. In our experience 36 base-pair reads yield good mappability and high sensitivity to a wide size range of restriction fragments.

```
> subsection.py -i sample.fastq \
      -o infile.fastq \
      -s length_of_viewpoint_plus_restric-
tion_site \
      -e length_of_viewpoint_plus_restric-
tion_site+36
```

4. Check for the presence of primary restriction enzyme recognition sites in the reads. Restriction sites can sometimes be absent from a subset of the reads, which is likely due to unexpected endonuclease activity and/or PCR artifacts. To check for such issues, the bases comprising the restriction site are extracted from the reads, and the percentage of reads containing the expected restriction site is determined.

```
> subsection.py -i sample.fastq \
      -o sites.txt \
      -s length_of_viewpoint \
      -e length_of_viewpoint+6 \
      --table
> cut -f 2 sites.txt | sort | uniq -c | sort
-nr>sites.cnt
> grep -e recognition_sequence sites.cnt
```

   As a rule of thumb, over 90% of the reads should contain the primary restriction enzyme site.

5. Align the trimmed reads to the reference genome with *Bowtie* [14] (*see* **Note 28**).

```
> bowtie -qS -l 32 -n 2 -p 4 --best -m 1 /
path/to/genome/index \
    infile.fastq>infile.sam 2>bowtie.error.log
> samtools view -Sb infile.sam>infile.bam
> samtools sort infile.bam infile.srt
> samtools index infile.srt.bam
```

6. Convert the BAM files into BEDgraph files for visualization and quality control (*see* **Note 29**).

```
> echo 'track type=bedGraph name=$*
visibility=full'>infile.cvrg.bedgraph
> bedtools genomecov -ibam infile.srt.bam
>>infile.cvrg.bedgraph
```

3.9 Downstream Data Analysis

1. Generate a restriction map from the target genome.

```
> findSequence.py -f genome.fasta -s recog-
nition_sequence -b occurences.bed
> regionsBetween.py -I occurences.bed -s
chromsizes.txt -o regions.bed
> bedtools sort -I regions.bed>regions.srt.
bed
```

2. Count the number of reads per restriction fragment.

```
> alignCounter.py -b infile.srt.bam -r re-
gions.srt.bed -o infile.table
```

3. Convert the tables to BEDgraph files for visualization in genome browsers.

```
> echo 'track type=bedGraph name=$*
visibility=full'>infile.bedgraph
> gawk '/[^#]/{if($4>0) print $1 "\t" $2
"\t" $3 "\t" $4;}' \
       infile.table >>infile.bedgraph
```

4. (*Optional*) Convert the BEDgraph files to .tdf files for viewing in the IGV genome browser. IGV can also perform this indexing upon loading BEDgraph files.

```
> java -Xmx2g -Djava.awt.headless=true -jar
igvtools.jar \
      -f median,mean,max -z 10 infile.bedgraph
file.tdf genome-name
```

5. Verify 3C-Seq library quality by examining several important parameters. First and foremost, the vast majority (>50%) of reads should map back to the cis-chromosome, clearly clustering in close proximity (i.e., within 1 Mb) of the viewpoint. Read count percentages on the viewpoint fragment itself and the fragment directly adjacent to the location of the reading primer (the "self-circle" and "undigested" prominent signals, respectively; *see* Figs. 2 and 3) are informative for assessing crosslinking and digestion efficiencies. Van de Werken et al. [19] provide excellent information on 3C-/4C-Seq library quality control parameters.

6. Normalization, in-depth statistical analysis, and visualization of 3C-/4C-Seq data can be performed using several existing software packages. 3C/4C data analysis is usually performed using either a restriction fragment-based approach or a window-based approach. Using the restriction fragment-based

approaches, the reference genome is digested with the restriction enzyme in silico (*see* **step 1** above). These restriction fragments are then assigned a score based on the number of reads aligning to that fragment (*see* **step 2** above). In window-based approaches, sequencing read enrichments are determined per chosen genomic interval (i.e., a 20-kb window). Restriction fragment-based approaches normally yield bar and line graphs as a standard visual output (Fig. 4a, b). Window-based approaches are often represented using smoothened line graphs and "domainograms" (Fig. 4b, c). Below, we briefly describe the main 3C-/4C-Seq analysis software packages currently available.

(a) *r3Cseq* [20] is a Bioconductor package available for the R statistical software environment. The package performs data normalization, identifies statistically significant interactions (within and between experimental datasets), and provides several options for data visualization. The package uses BAM alignment files to count the number of reads per restriction fragment. These are then normalized using a reverse-cumulative fitting procedure in which the area around the viewpoint is not considered. Significant interactions are detected by comparing the observed read counts per fragment to those of a smoothened background signal. Biological replicates are combined using Fisher's combined probability test. r3Cseq creates restriction fragment bar graphs, log2 fold change comparison tracks, line plots, and domainograms.

(b) *Basic4Cseq* [21] is another Bioconductor package available in R. This software is focused on data visualization, but does provide data smoothening and a basic "reads-per-million" normalization function. Basic4Cseq visualizes the raw and normalized data per restriction fragment across the genome and creates running mean and median plots for varying window sizes. Data can be exported in either the .csv or .wig formats for downstream processing using other tools.

(c) *fourSig* [22] is a command-line tool written in Perl and R. It performs a per restriction fragment analysis, but also features a novel statistical approach to determine significant interactions with greater confidence ("interaction prioritization").

Fig. 4 (continued) frequent interactions between transcription factor-bound regions and the *MYB* promoter. Panel **c** displays 3C-Seq data as a so-called domainogram generated using r3CSeq software [20], in which an increasing window size (2–30 kb resolution) is used to identify broader regions of high interaction frequency. Two conditions (red blood cell progenitors and brain cells) were analyzed. Note how certain regions in the *Myb* locus display strong tissue-specific interactions with the *Myb* promoter. Many of the red blood cell-specific interactions contain transcription factor-bound enhancer regions, in agreement with the much higher *Myb* expression levels in these cells. Data were taken from Stadhouders et al. [5, 16]

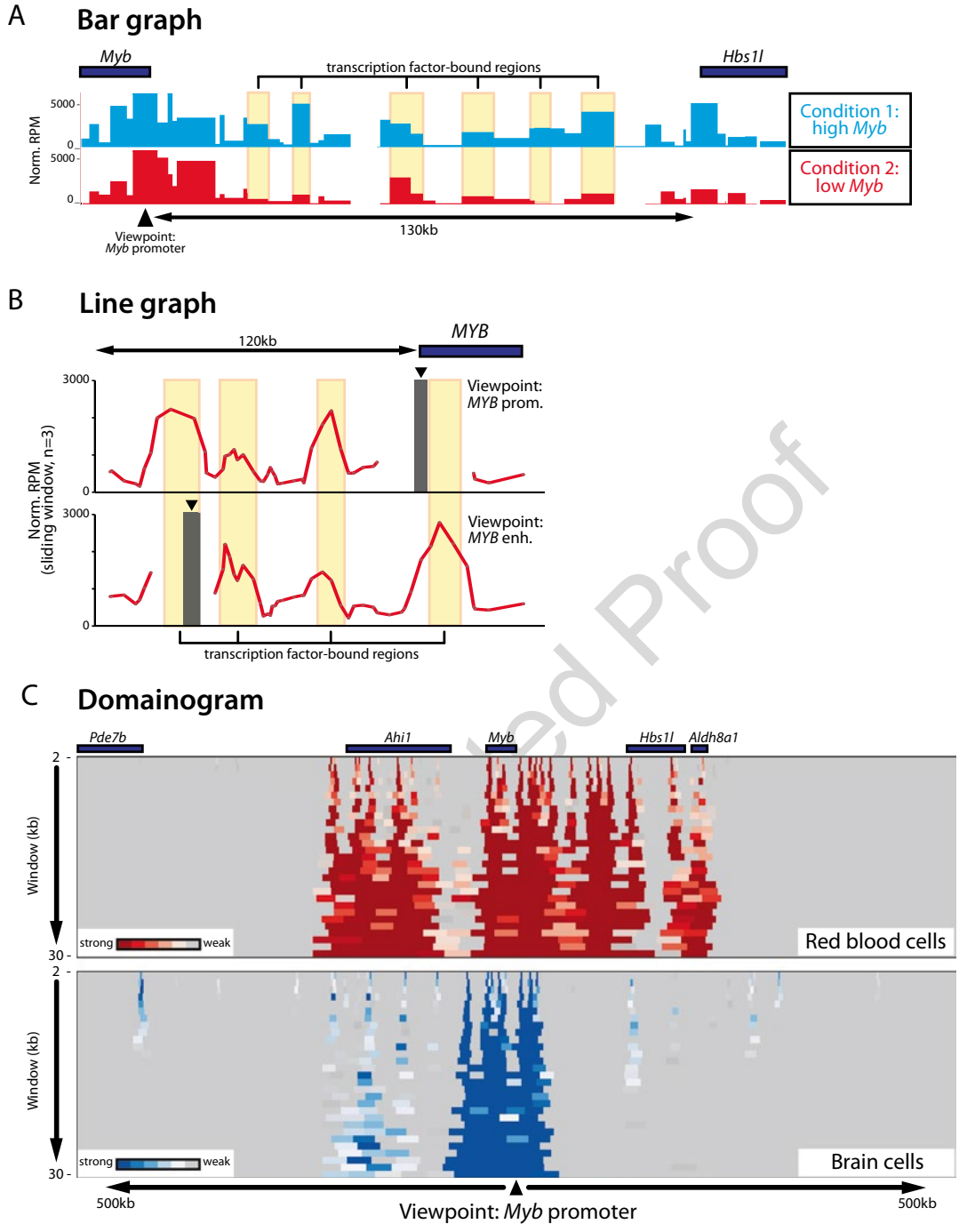

Fig. 4 Examples of 3C-Seq data visualization. 3C-Seq data can be displayed using different strategies. In the different examples shown, the mouse (panels **a** and **c**) or human (panel **b**) *Myb*/*MYB* promoter was used as a viewpoint in order to detect genomic regions in the near vicinity interacting with the gene. Panel **a** shows a standard bar graph representation, with each individual bar representing the (normalized) number of sequences retrieved from that particular restriction fragment—a measure for the interaction frequency between that fragment and the viewpoint. Two conditions (red blood cell progenitors and differentiating red blood cells) were analyzed. Note that interactions between the *Myb* promoter and transcription factor-bound intergenic regions (marked in *yellow*) are lost upon differentiation. Panel **b** depicts 3C-Seq data as a line graph, again highlighting

In terms of visualization, fourSig employs restriction fragment read count files for genome browser visualization.

(d) *4Cseqpipe* [23] software is a complete analysis package that not only performs data normalization and visualization but also includes a specific 4C-specific aligner that matches reads with the theoretical restriction fragments in the genome. The processed data is visualized in several ways, including high-resolution domainograms and genome-wide ideograms.

(e) *FourCSeq* [24] is an R package that uses a variance-stabilizing transformation and a trend-fitting approach for advanced data normalization. Statistically significant interactions above the expected background trend can be compared between experimental conditions, resulting in line graphs and log2 fold change comparison tracks that highlight significantly different interactions.

4 Notes

1. Any mammalian cell is in principle suitable for 3C-Seq analyses. To ensure equivalent formaldehyde exposure (and thus comparable crosslinking efficiencies) among individual cells and between different cell types, it is highly desirable to obtain single-cell suspensions before crosslinking. Especially when using tissue samples, the use of cell strainers and even collagenase treatment should be considered.
2. Typically, formaldehyde used in 3C (and many other biochemical assays, such as chromatin immunoprecipitation) comes from 37% formaldehyde stocks that contain 10–15% methanol as a stabilizing agent. Although suitable for 3C-based technologies, one should realize that over time these formaldehyde stocks are subjected to polymerization and oxidation. We consider the maximum shelf life of a 37% formaldehyde bottle to be 6 months after opening. Alternatively, one could use single-use ampules of (methanol-free) formaldehyde.
3. The recipe used here has not changed since the original publication of 3C by Dekker et al. [25], and it remains the most common lysis buffer for 3C-based approaches. It allows for (a partial) extraction of nuclei under mild conditions. As efficient lysis is considered important for allowing the restriction enzyme access to the chromatin, other lysis buffers as well as the use of a douncer have been employed (i.e., in Splinter et al. [10]) to increase lysis efficiency. Nevertheless, one should strive to be as gentle as possible considering that relevant chromatin co-associations originate from intact nuclei [26, 27]. In our hands, cell lysis is often poor using only the original 3C

lysis buffer. This is usually not problematic, as SDS addition (0.25–0.5 %) in the subsequent steps prior to digestion often efficiently lyses the cell membrane while keeping crosslinked nuclei intact. Nevertheless, some cell types are resistant to lysis (even in the presence of SDS), and it is therefore important to monitor cell lysis during the first phase of the procedure to obtain high digestion efficiencies.

4. As described in Subheading 3.1 (**step 2**), primary restriction enzyme choice is primarily determined by general performance under 3C conditions, sensitivity to CpG methylation, and ligation efficiencies of the resulting fragment ends. Other important parameters are recognition site distribution around the viewpoint of choice, as well compatibility with the second restriction enzyme (*see* Fig. 2 and its legend for a detailed example).
5. When diluting the nuclei as traditionally done in 3C-based protocols, large quantities of T4 DNA ligation buffer are used. We routinely produce our own ligation buffer using the manufacturer's standard recipe, which is stored at −20 °C as single-use aliquots.
6. The second restriction enzyme digestion is not performed on crosslinked chromatin under the somewhat harsh 3C conditions. Therefore, any 4-bp-recognizing restriction enzyme insensitive to CpG methylation leaving fragment ends with high ligation efficiencies can be used. Important is to ensure that the secondary restriction enzyme cuts the viewpoint fragment while creating a DNA fragment suitable for self-circularization in the second ligation (>200 bp) and inverse PCR primer design (*see* Fig. 2 and its legend for a detailed example).
7. After two rounds of digestion–ligation, 3C-Seq PCR template samples remain contaminated with impurities that could potentially affect the inverse PCR. Standard spin column-based DNA purification kits improve sample purity, although in our hands the persistence of suboptimal 260/280 absorbance ratios does not influence PCR efficiency and linearity under the conditions described in Subheading 3.6.
8. Final 3C-Seq library purification can be performed using spin column-based DNA purification kits or AMPure XP beads. Important at this step is to remove as much unused primer and primer dimer as possible, which resides in the <100–120-nt range. Substantial amounts of remaining primer dimers will negatively affect sequencing yield and quality. The Roche High Pure PCR Product Purification kit suggested by Splinter et al. [10] performs well in separating informative PCR product from primers and primer dimers.
9. 3C-Seq libraries have a broad size range (120 bp to >3 kb, *see* Fig. 3), which should be taken into account when choosing Agilent Bioanalyzer cartridges for quantification.

10. Standard primer design guidelines (for, i.e., length, %GC, annealing temperature, and secondary structures) can be followed. Critical is to position the reading primer—located near the primary restriction enzyme site—as close to the enzyme recognition site as possible, preferably on top (Fig. 2). As single-read sequencing commences from this side only, any extra sequence in between the primer and the primary restriction site will lower the number of nucleotides available for mapping sequences stemming from unknown interaction fragments to the genome (*see* Fig. 2 and Subheading 3.8). The design of the primer near the secondary restriction site allows for more flexibility (within ±100 bp of the enzyme recognition site, *see* Fig. 2), as sequencing is not initiated from this site. Ensure that PCR primers used to construct the final 3C-Seq libraries contain the appropriate P5 and P7 Illumina adapters as overhang: place the P5/P7 sequence directly upstream of the annealing part. Short 3–6-nt barcodes can be placed in between the P5 adapter, and annealing part of the reading primer if multiplexing of libraries generated with the same viewpoint PCR primers is desired. Adapter sequences: P5: 5′-AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′;P7: 5′-CAAGCAGAAGACGGCATACGA-3′.
11. Formaldehyde crosslinking efficiency is influenced by temperature. Ensure that solutions in which cells are crosslinked are at room temperature.
12. Published 3C-based protocols typically employ a 1–2% range of final formaldehyde concentration when crosslinking mammalian cells. Although only few systematic comparisons have been published (especially when using mammalian cells), variations in % formaldehyde within this range appear to have little influence on experimental outcome. It should be noted that higher formaldehyde concentrations can reduce digestion efficiencies [28].
13. Using more than 10×10^6 cells as starting material can promote the formation of nuclear aggregates. This phenomenon is very cell type specific and seems to correlate with cell size. Some aggregation is normal and not detrimental, but excessive clumping can severely reduce digestion efficiencies and should be avoided. In our hands, the best solution to reduce aggregation is to start with lower numbers of cells.
14. Increasing Triton X-100 concentrations (as suggested by Splinter et al. [10] and van De Werken et al. [19]) might augment SDS quenching and further improve digestion efficiency.
15. Restriction enzymes are known to vary with respect to their ability to remain active during prolonged incubation times. In case a primary restriction enzyme is used with reported low survivability, multiple separate and/or extra enzyme additions can improve digestion efficiency.

16. Digestion efficiencies can also be quantified more accurately using quantitative PCR analysis (qPCR) employing primer sets that span restriction sites. One should strive for >70% overall digestion efficiencies.

17. This protocol still contains the traditional extreme dilution of the first 3C ligation initially implemented to favor intramolecular ligation and to minimize random ligation events. Although not detrimental, the benefit of ligating crosslinked chromatin under such dilute conditions is now considered obsolete [26, 27]. Therefore, primary 3C ligations can also be performed in smaller volumes (e.g., as described in Rao et al. [27]).

18. Bear in mind that average DNA fragment sizes after primary ligation differ depending on whether a 6-bp- or 4-bp-recognizing restriction enzyme was used: ligations after digestion with the latter type of enzyme tend to produce a less sharp high molecular weight band.

19. At this point, the 3C library can be subjected to quantitative interaction analysis using qPCR. Both SYBR Green and TaqMan probe approaches can be used to quantify interactions in a one-versus-one manner. This can be extremely useful when verifying key interactions identified using more high-throughput methodologies such as 3C-/4C-Seq. Important considerations for such 3C-(q)PCR assays have been described elsewhere [29].

20. In contrast to the primary ligation step (also *see* **Note 17**), the second ligation (specific for 3C-/4C-Seq protocols) needs to be performed under extremely diluted conditions to strictly promote self-circularization of individual fragments and to prevent random ligations between DNA molecules in the solution.

21. Traditionally, 3C-based protocols include several phenol/chloroform purifications steps. In the original 4C-/3C-Seq protocols [11, 30], ligation products at this stage were first purified by a phenol/chloroform/isoamyl alcohol extraction, immediately followed by ethanol precipitation and spin column-based purification. To prevent additional loss of material when working with low numbers of cells, we now omit the phenol/chloroform extraction and immediately continue with the ethanol precipitation.

22. A standard setup for testing 3C-Seq PCR primers involves running a duplicate series of PCR reactions (using the reaction setup and program described in **step 11** of Subheading 3.6) using 25, 50, and 100 ng of 3C-Seq PCR library input DNA. DNA smears are visualized on a 1.5% standard agarose gel to determine reproducibility and linearity. Another hallmark of a successful 3C-Seq PCR is the appearance of two "prominent" bands representing self-circularization of the viewpoint fragment and ligation of the viewpoint fragment to

AU5

the adjacent fragment when digestion of the primary restriction site was not achieved (Figs. 2 and 3). Both these events are much more prominent than other interactions, hence their abundance in 3C-/4C-Seq libraries. The size of the expected prominent bands can be easily estimated (Fig. 2). Note that these fragments can be very short and therefore not compatible with self-circularization, or they can simply not be visualized on a 1.5 % agarose gel.

23. One column is sufficient when using the High Pure PCR Product Purification kit.

24. The P5/P7 single-read Illumina sequencing adapters used in this protocol (*see* **Note 10** and Stadhouders et al. [11]) are not compatible with MiSeq and NextSeq sequencing machines and should be sequenced on HiSeq or GA instruments. If sequencing on MiSeq/NextSeq machines is desired, the P5/P7 sequences described here should be swapped for adapter sequences appropriate for these instruments.

25. Important to realize is that Illumina sequencing instruments use the first four bases to recognize DNA clusters on the flowcell. Especially when reading these first four bases, nucleotide complexity needs to be high—too little variation and base calling will be compromised—leading to reduced sequence yields. As 3C-/4C-Seq libraries consist of amplicons all starting with the same sequence (the viewpoint-specific primer), pools of different (>6) 3C-Seq libraries (either using different viewpoint PCR primers or different barcodes placed upstream of the viewpoint-specific reading primer sequence) have to be sequenced together in a single lane to create enough sequence diversity. Alternatively, one could combine 3C-Seq libraries with other samples not suffering from these diversity issues (e.g., RNA-Seq, ChIP-Seq samples, or a PhiX sequencing control sample).

26. All analysis software has been tested on Red Hat Enterprise Linux Server release 6.4 (Santiago), but should work on any Linux distribution.

27. In most situations, reads can be assigned to specific samples based on the first ten bases. Using more bases may cause data yield to suffer as sequence errors accumulate across the reads. Using fewer bases (down to six) is possible when the number of viewpoints is small and their sequences differ at the 5′-end.

28. Bowtie does not align reads with insertions or deletions, but this is generally not an issue in 3C-/4C-Seq data analysis. If insertions and deletions need to be taken into account, BWA [31] can be used.

29. BAM files can in principle be viewed in genome browsers directly, but the large numbers of reads present near the viewpoint location prevent efficient visualization on most personal

computers. BEDgraph files only contain the read–depth per genome position and thus allow 3C-Seq data to be rapidly viewed on any standard personal computer. We regularly use the standalone application IGV (*see* software list) or the web-based UCSC genome browser (https://genome.ucsc.edu/), although any genome viewer can be used.

Acknowledgments

The authors would like to thank the members of the Soler and Van IJcken laboratories for insightful discussions. R.S. is supported by an EMBO Long-term Fellowship (ALTF 1201-2014) and a Marie Curie Individual Fellowship (H2020-MSCA-IF-2014, proposal #654933). E.S. is supported by grants from the FSER (Schlumberger Foundation for Education and Research), ARC Foundation ("project ARC" n°SFI20121205625), and the Atip–Avenir program.

References

1. de Wit E, de Laat W (2012) A decade of 3C technologies: insights into nuclear organization. Genes Dev 26(1):11–24. doi:10.1101/gad.179804.111
2. Pombo A, Dillon N (2015) Three-dimensional genome architecture: players and mechanisms. Nat Rev Mol Cell Biol 16(4):245–257. doi:10.1038/nrm3965
3. de Laat W, Duboule D (2013) Topology of mammalian developmental enhancers and their regulatory landscapes. Nature 502(7472):499–506. doi:10.1038/nature12753
4. Visser M, Kayser M, Palstra RJ (2012) HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. Genome Res 22(3):446–455. doi:10.1101/gr.128652.111
5. Stadhouders R, Aktuna S, Thongjuea S, Aghajanirefah A, Pourfarzad F, van Ijcken W, Lenhard B, Rooks H, Best S, Menzel S, Grosveld F, Thein SL, Soler E (2014) HBS1L-MYB intergenic variants modulate fetal hemoglobin via long-range MYB enhancers. J Clin Invest 124(4):1699–1710. doi:10.1172/JCI71520
6. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C, Aneas I, Credidio FL, Sobreira DR, Wasserman NF, Lee JH, Puviindran 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
7. Pomerantz MM, Ahmadiyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H, Beckwith CA, Chan JA, Hills A, Davis M, Yao K, Kehoe SM, Lenz HJ, Haiman CA, Yan C, Henderson BE, Frenkel B, Barretina J, Bass A, Tabernero J, Baselga J, Regan MM, Manak JR, Shivdasani R, Coetzee GA, Freedman ML (2009) The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nat Genet 41(8):882–884. doi:10.1038/ng.403
8. Cowper-Sal lari R, Zhang X, Wright JB, Bailey SD, Cole MD, Eeckhoute J, Moore JH, Lupien M (2012) Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. Nat Genet 44(11):1191–1198. doi:10.1038/ng.2416
9. Bauer DE, Kamran SC, Lessard S, Xu J, Fujiwara Y, Lin C, Shao Z, Canver MC, Smith EC, Pinello L, Sabo PJ, Vierstra J, Voit RA, Yuan GC, Porteus MH, Stamatoyannopoulos JA, Lettre G, Orkin SH (2013) An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. Science 342(6155):253–257. doi:10.1126/science.1242088
10. Splinter E, de Wit E, van de Werken HJ, Klous P, de Laat W (2012) Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. Methods 58(3):221–230. doi:10.1016/j.ymeth.2012.04.009
11. Stadhouders R, Kolovos P, Brouwer R, Zuin J, van den Heuvel A, Kockx C, Palstra RJ, Wendt KS, Grosveld F, van Ijcken W, Soler E (2013) Multiplexed chromosome conformation capture sequencing for rapid genome-scale

high-resolution detection of long-range chromatin interactions. Nat Protoc 8(3):509–524. doi:10.1038/nprot.2013.018

12. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079. doi:10.1093/bioinformatics/btp352
13. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(6):841–842. doi:10.1093/bioinformatics/btq033
14. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10(3):R25. doi:10.1186/gb-2009-10-3-r25
15. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. Nat Biotechnol 29(1):24–26. doi:10.1038/nbt.1754
16. Stadhouders R, Thongjuea S, Andrieu-Soler C, Palstra RJ, Bryne JC, van den Heuvel A, Stevens M, de Boer E, Kockx C, van der Sloot A, van den Hout M, van Ijcken W, Eick D, Lenhard B, Grosveld F, Soler E (2012) Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development. EMBO J 31(4):986–999. doi:10.1038/emboj.2011.450
17. Ribeiro de Almeida C, Stadhouders R, Thongjuea S, Soler E, Hendriks RW (2012) DNA-binding factor CTCF and long-range gene interactions in V(D)J recombination and oncogene activation. Blood 119(26):6209–6218. doi:10.1182/blood-2012-03-402586
18. Stadhouders R, de Bruijn MJ, Rother MB, Yuvaraj S, Ribeiro de Almeida C, Kolovos P, Van Zelm MC, van Ijcken W, Grosveld F, Soler E, Hendriks RW (2014) Pre-B cell receptor signaling induces immunoglobulin kappa locus accessibility by functional redistribution of enhancer-mediated chromatin interactions. PLoS Biol 12(2), e1001791. doi:10.1371/journal.pbio.1001791
19. 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
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. Walter C, Schuetzmann D, Rosenbauer F, Dugas M (2014) Basic4Cseq: an R/Bioconductor package for analyzing 4C-seq data. Bioinformatics 30(22):3268–3269. doi:10.1093/bioinformatics/btu497
22. Williams RL Jr, Starmer J, Mugford JW, Calabrese JM, Mieczkowski P, Yee D, Magnuson T (2014) FourSig: a method for determining chromosomal interactions in 4C-Seq data. Nucleic Acids Res 42(8):e68. doi:10.1093/nar/gku156
23. 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
24. Klein FA, Pakozdi T, Anders S, Ghavi-Helm Y, Furlong EE, Huber W (2015) FourCSeq: analysis of 4C sequencing data. Bioinformatics. doi:10.1093/bioinformatics/btv335
25. Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295(5558):1306–1311. doi:10.1126/science.1067799
26. Gavrilov AA, Gushchanskaya ES, Strelkova O, Zhironkina O, Kireev II, Iarovaia OV, Razin SV (2013) Disclosure of a structural milieu for the proximity ligation reveals the elusive nature of an active chromatin hub. Nucleic Acids Res 41(6):3563–3575. doi:10.1093/nar/gkt067
27. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159(7):1665–1680. doi:10.1016/j.cell.2014.11.021
28. Palstra RJ, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W (2003) The beta-globin nuclear compartment in development and erythroid differentiation. Nat Genet 35(2):190–194. doi:10.1038/ng1244
29. Hagege H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, de Laat W, Forne T (2007) Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat Protoc 2(7):1722–1733. doi:10.1038/nprot.2007.243
30. 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
31. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760. doi:10.1093/bioinformatics/btp324