



Published in final edited form as:

Methods. 2012 November ; 58(3): . doi:10.1016/j.ymeth.2012.07.022.

Analysis of long-range chromatin interactions using Chromosome Conformation Capture

N. Naumova⁽¹⁾, E. M. Smith⁽¹⁾, Y. Zhan⁽¹⁾, and J. Dekker^{(2),*}

Programs in Systems Biology and Gene Function and Expression, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 364 Plantation Street, Worcester, MA, 01605-0103, USA

Abstract

Chromosome Conformation Capture, or 3C, is a pioneering method for investigating the three-dimensional structure of chromatin. 3C is used to analyze long-range looping interactions between any pair of selected genomic loci. Most 3C studies focus on defined genomic regions of interest that can be up to several hundred Kb in size. The method has become widely adopted and has been modified to increase throughput to allow unbiased genome-wide analysis. These large-scale adaptations are presented in other articles in this issue of *Methods*. Here we describe the 3C procedure in detail, including the appropriate use of the technology, the experimental set-up, an optimized protocol and troubleshooting guide, and considerations for data analysis. The protocol described here contains previously unpublished improvements, which save time and reduce labor. We pay special attention to primer design, appropriate controls and data analysis. We include notes and discussion based on our extensive experience to help researchers understand the principles of 3C-based techniques and to avoid common pitfalls and mistakes. This paper represents a complete resource and detailed guide for anyone who desires to perform 3C.

Introduction

Chromosomes form intricate three-dimensional structures inside the confined cell nucleus. This organization is thought to play roles in many, if not all, aspects of genome regulation, including gene expression, DNA replication, chromosome transmission and maintenance of genome stability [1–3]. Gene expression in particular is profoundly dependent on chromatin folding, where looping interactions facilitate long-range control by distant gene regulatory elements [2, 4, 5]. Furthermore, at the nuclear level, groups of active genes are found clustered around sub-nuclear structures enriched in transcription and splicing machineries [6]. Similarly, inactive regions of the genome are found in clusters, e.g. around polycomb bodies [7] and at the nuclear lamina [8].

Chromosome structure and nuclear organization have been studied extensively for over a century, using an expanding array of technologies that allow observation of chromosome folding at increasing resolution. Currently, two types of approaches are being used. First, microscopic studies allow study of chromosome structure and chromatin dynamics in single cells. Recent technologies use tagged DNA binding proteins allow the visualization of the positions and movements of defined loci inside living cells [9]. Recent developments in optics and image analysis have increased the resolution with which the relative sub-nuclear positions of loci can be determined. A second category of technologies employs molecular

⁽²⁾To whom correspondence should be addressed: job.dekker@umassmed.edu. Phone: 508-856-4371, FAX: 508-856-2105.

*Correspondence: job.dekker@umassmed.edu.

⁽¹⁾These authors contributed equally to this work

and genomic approaches to obtain information on average chromatin folding for large populations of cells [10]. This set of approaches is based on the Chromosome Conformation Capture technology (3C), developed over a decade ago [11]. 3C-based technologies allow the detection of the relative frequency of interaction between any pair of loci in the genome. From these interaction frequencies the folding of chromatin can be inferred. For instance, frequent interactions between two distant genomic loci point to the presence of a chromatin loop [12]. The resolution of 3C is determined by the choice of restriction enzyme, but is usually in the range of several Kb, significantly higher than achievable by light microscopy.

Application of 3C has identified direct and non-random looping interactions between distant parts of the linear genome, including physical contacts between enhancers and their distal target genes [2]. Further application of a variety of 3C derivatives has led to the notion that genomes are organized in complex spatial networks via looping interactions that often are cell-type and condition-dependent and directly related to long-range gene control [13].

3C and its offspring of variants including 4C [14, 15], 5C [16], ChIA-PET [17] and Hi-C [18] (described in separate papers in this issue of Methods), are all based on the same basic principle of capturing and detecting long-range chromatin interactions and have 4 common steps (Fig. 1A): 1) Chemical cross-linking of chromosomes to covalently link chromatin segments that are in close spatial proximity; 2) Fragmenting the solubilized genome into small pieces, usually by digesting it with a restriction enzyme; 3) ligation of linked DNA fragments under diluted conditions where intra-molecular ligation is strongly favored over inter-molecular events; and 4) detection and quantification of ligation products. The various 3C-based methods differ mostly in how ligation products are detected. In the case of 3C, ligation products are detected one at a time by PCR using locus specific primers (Fig. 1B). Other 3C-derived methods use a variety of approaches to increase the number of interactions (ligation products) that are detected in parallel, thereby increasing the throughput of the assay [10, 19, 20].

The 3C procedure produces a comprehensive library of ligation products representing chromatin interactions throughout the entire genome. However, because interactions in standard 3C experiments are detected one at a time, a typical 3C analysis is usually limited to interrogation of at most hundreds of pair-wise interactions and is focused on the detection of looping interactions in relatively small regions -from 10kb up to 1Mb [21–23]. 3C is mainly used in hypothesis-driven experiments, designed based on some prior knowledge such as the genomic locations of functional elements of interest.

Whereas subsequent 3C-based methods (4C, 5C and Hi-C) were designed to increase the throughput of interaction detection, 3C has remained a critical technique that is commonly used for fine scale analysis of genomic regions of interest. 3C has been used to study chromatin folding in a range of organisms, ranging from bacteria, yeast and plant to human. In the original study performed on *S. cerevisiae* [Dekker 2002] 3C was used to measure changes of inter-chromosomal contacts between centromeres and homologous chromosomes during meiosis, and to determine the overall population average three-dimensional conformation of chromosome III. Since then, 3C has been applied mostly to study the interactions between genes and distal regulatory elements such as enhancers. The first such study demonstrated physical contacts between the β -globin genes and the Locus Control Region (LCR), which is known to strongly activate these genes [21]. This was one of the first experimental demonstrations that long-range gene regulation involved looping interactions between widely spaced genomic elements. Many more examples of such looping interactions have now been described, indicating that chromatin looping is a general mechanism of gene control in higher eukaryotes [2, 13]. Chromatin looping interactions are driven by specific protein complexes that bind the two interacting loci. For instance, the

looping interactions between the LCR and the globin genes require several transcription factors, including GATA1 and EKLf1, and some require the CTCF protein [24–26].

3C has been used to identify interactions that occur between chromosomes as well, such as the interaction between the promoter region of the IFN-gamma gene on chromosome 10 and the regulatory regions of the T(H)2 cytokine locus on chromosome 11 [27], although the general relevance of such contacts for gene regulation is not well established.

Currently, 3C is mostly used for targeted analysis of loci of interest, to identify long-range interaction between candidate genes and regulatory elements, and to probe how these interactions change upon perturbations such as knockdown of specific chromatin factors thought to mediate chromatin folding. Another emerging application is to link regions identified as playing a role in disease by genome-wide association studies (GWAS) to other genomic loci. GWAS studies often identify regions devoid of genes, but containing putative gene regulatory elements. 3C is now being used to identify potential target genes located around the GWAS region that physically interact with the GWAS regions, or the regulatory elements located within it [28–30].

Here we discuss in detail the principles of the 3C method, paying special attention to experiment design and data analysis. We present an updated protocol for performing 3C analysis in mammalian cells and discuss potential pitfalls. The presented protocol can be easily adapted for any other organisms. Further, we build on our years of experience with 3C to describe troubleshooting solutions and to identify critical issues related to the planning, execution, and interpretation of the experiments.

Section 1: Technical Overview of the 3C method

In this section we provide a discussion of each of the 4 steps of 3C, paying particular attention to the molecular biology behind the method. The basic steps of 3C are: 1) Formaldehyde crosslinking, 2) Digestion with a restriction enzyme, 3) intra-molecular ligation, and 4) ligation product detection via PCR-based methods (Fig. 1A).

1.1 Formaldehyde concentration and fixation time

Formaldehyde concentration and time of fixation affect 3C-based experiments and need to be standardized to facilitate accurate comparison between samples. Fixation conditions can be different for different species and are defined by chromatin properties, presence of a cell wall etc. Generally, conditions used for chromatin immunoprecipitation experiments work for 3C. We crosslink mammalian cells with 1% formaldehyde for 10 minutes at room temperature. Other groups have reported using 2% formaldehyde for 10 minutes [21]. 3C experiments in *Drosophila* embryos were successfully performed with fixation with 3% formaldehyde for 30 minutes at 25°C [31]. Intact yeast cells should be fixed with 3% formaldehyde for 10 minutes (Belton, J-M. and Dekker, J., unpublished), but yeast spheroplasts are fixed with 1% formaldehyde for 10 minutes [11].

Changing fixation conditions affects the amount and density of protein-DNA cross-links, which in turn affects the efficiency of restriction digestion and thus the size of DNA-protein complexes. Inefficient formaldehyde cross-linking, caused by using a low concentration of formaldehyde or a too short of an incubation time, may lead to a failure to capture looping interactions. Over-fixed chromatin will make digestion very inefficient, leading to large DNA fragments and low PCR signals. Optimal conditions can be determined experimentally by performing 3C using a range of formaldehyde concentrations. An appropriate formaldehyde concentration will lead to readily detectable and abundant ligation products for fragments separated only a few Kb. When analyzing cells in tissues or organs, it is

recommended to first dissociate these materials, e.g. with collagenase, into single cells prior to fixation (e.g. [32]).

1.2 Digestion with Restriction Enzymes

The second step of 3C involves digesting the crosslinked chromatin with a restriction enzyme. We recommend using restriction enzymes that recognize and cut 6 bp sites when possible (“6-cutter”, see section 2.2) for overnight digestion. The amount of enzyme in the reaction can be increased if higher digestion efficiency is required (at least up to 5-fold).

Protein complexes crosslinked to DNA may block restriction sites and reduce efficiency of restriction digestion. For instance, efficiency of yeast and mammalian chromatin digestion is around 70–75% on average as was measured by PCR with primers located across restriction sites [13, 21, 22, 33].

Inactive and condensed chromatin is generally less accessible to nucleases than active and open chromatin, and this might confound 3C studies. Importantly, 3C analyses have been found to be not, or only to a very limited extent, affected by this intrinsic difference in chromatin compaction. Several studies directly determined the digestion efficiency in the context of a 3C experiment of actively transcribed accessible loci and of repressed, methylated and condensed loci. Digestion efficiency was found to be unaffected [21, 22, 34]. The explanation for this observation is that chromatin digestion in 3C is performed after chromatin is partly denatured in the presence of 0.1% SDS and brief incubation at 37°C, which removes proteins that are not cross-linked from DNA and partly denatures cross-linked proteins. This dramatically increases accessibility of DNA. Efficiency of restriction digestion can be easily determined by PCR using primers on either side of restriction sites (see supplemental Fig. 1). We recommend saving a small aliquot of chromatin directly after digestion (and before ligation) for this analysis if desired.

1.3 Ligation and Reversal of Crosslinking

The third step involves DNA ligation. This step is performed at low DNA concentrations to strongly favor intra-molecular ligation of cross-linked chromatin fragments over background intermolecular ligation between fragments that are not cross-linked. Intra-molecular ligation is kinetically fast, obviating the need for prolonged ligation times. The ligation time should be kept at a minimum, to avoid increasing the level of background ligations. After ligation, the cross-links are reversed by heating at 65°C in the presence of proteinase K. The 3C ligation product library is then purified and is ready for analysis.

1.4. Detection of Ligation Products

The final step is the detection and quantification of ligation products, representing long-range chromatin interactions. For this step, locus-specific primers are designed to specifically amplify ligation junctions. PCR amplicons are typically around 200 bp in size to facilitate efficient amplification. Both end-point PCR and quantitative real-time PCR (qPCR) have been employed to quantify the abundance of 3C ligation products, with very comparable results (e.g compare regular 3C data for the beta-globin locus described in Tolhuis et al. [21] with 3C-qPCR data for the same region described in Splinter et al. [25]). In both cases one needs to carefully titrate the amount of 3C ligation product library to ensure amplification and quantification is in the linear range. In addition, controls should be included to correct for any biases in PCR primer efficiency. These controls are described below in sections 2.1 and 2.4.

Section 2: Experimental design considerations

When embarking on a 3C analysis, one needs to carefully plan the design of the experiment. Here we describe important considerations related to selection of the genomic regions for analysis, the inclusion of a control region, the choice of restriction enzyme and the design of PCR primers.

2.1 Choosing the region of interest and a control region

The first step in setting up a 3C experiment is to identify the region(s) to investigate. Any unique region of a genome can be analyzed by 3C. The size of a region is limited firstly by the desired resolution, determined by the restriction enzyme, and secondly by the amount of PCR reactions one can perform, which is related to the amount of 3C library one can obtain for the cells of interest. The typical size of a region that can be comprehensively studied by 3C ranges from tens to hundreds of Kb, although longer-range interactions have been studied [29, 35], as well as interactions between chromosomes [27, 36, 37] (See Note 2.1 Signal to Noise).

When one intends to compare the folding of a locus of interest in different cell types, or under different conditions, it is important to choose a separate control region. This region needs to be selected based on prior knowledge that suggests that the region is similarly organized in the selected cell types or conditions. 3C interactions determined throughout this region are assumed to be the same and thus can be used as an internal data set to quantitatively normalize the 3C data obtained for the region of interest in the different cell types or conditions. We advise the use of gene-poor regions (or so called gene deserts), although a locus with house-keeping genes has also been successfully used [22].

2.2 Choosing an appropriate restriction enzymes

Restriction enzymes are used to digest crosslinked chromatin. Once the chromatin is digested, it is ligated to create a 3C template. The choice of restriction enzyme to use in a 3C experiment is highly dependent on the goal of the experiment and the region selected for analysis. Several points should be kept in mind when selecting a restriction enzyme.

2.2.1. Desired Resolution—The resolution at which interactions can be mapped is primarily determined by the size of the restriction fragments and thus the choice of the restriction enzyme. We recommend using a restriction enzyme that recognizes a 6 base-pair sequence cut site, such as EcoRI or HindIII. Such enzymes will cut the genome approximately once every 4 kb (although a wide variety of fragment sizes ranging from 100s of base pairs to 10s of kb will be obtained), resulting in around 1 million restriction fragments in the human genome. In some cases, higher resolution is desired. For instance, after initial 3C analysis with a “6-cutter” enzyme, one might want to map the location of an interacting element more precisely. “Fine-mapping” can be achieved by using a “4-cutter” restriction enzyme, which cuts on average every 256 base pairs, giving approximately 16,000,000 fragments of the human genome (see Note 2.2.1 for more on the desired resolution of restriction enzymes).

Note 2.1 **Signal to Noise:** 3C signals typically decay with genomic distance. As a result the signal to noise ratio decreases with increased distance between two interrogated loci, which usually limits 3C analysis to regions up to 1Mb. Other 3C-based techniques do not have this limitation because they employ binned data, where the signal is determined by all interactions within the bin and not by a single point, as in 3C.

Note 2.2.1 **Desired Resolution of Restriction Enzymes.** The complexity of the 3C library (i.e. number of potentially formed ligation products) is determined by the restriction enzyme, and can impact the reliability of PCR detection and quantification of individual ligation product. The complexity of the library obtained with a “4-cutter” restriction enzyme will be greatly increased as compared to a library generated with a “6-cutter”.

2.2.2. Proper Position of Cut Sites—It is desirable to choose a restriction enzyme which has a more or less equal spacing of cut sites across the analyzed region. Fragments that are too short or too long should be excluded from the primer design, as they can introduce biases to the data (see Note 2.2.2 Exclusion of Restriction Fragments). Thus, fragments less than 1 Kb or greater than 10 Kb should be excluded when using a “6-cutter.” In addition, when prior knowledge of positions of putative interacting elements is available, e.g. by the presence of histone modifications indicative of the presence of an enhancer or promoter, one can select a restriction enzyme that cuts the region in appropriate fragments that separate these elements from flanking regions, thus leaving elements of interest intact.

2.2.3. Digestion Efficiency—Not all restriction enzymes perform equally well in 3C. The reason for this is that digestion is performed in sub-optimal buffer conditions containing considerable concentrations of detergents. We have found that EcoRI, HindIII, BglII, XhoI, AclI, and BsrGI digest cross-linked chromatin efficiently, typically reaching 70% of digestion of each restriction site (although the region selected for analysis can change the digestion efficiency). Enzymes that produce staggered ends are recommended, as these ends are more efficiently ligated. Enzymes that generate blunt ends can be used as well, but ligation efficiency is somewhat reduced.

2.3. 3C Primer Design

In a 3C experiment any pair of interacting loci can lead to formation of six different ligation products (Figure 1B). Two of the resulting products are self-circles, which occur when a restriction fragment is ligated to itself. The other four combinations occur when two different restriction fragments are ligated to each other in various orientations. In a typical 3C experiment primers are designed to detect only one of the four ligation products between the two fragments. In order to detect a ligation product between two different fragments, PCR primers should be placed in an orientation indicated by the asterisk in Figure 1B. In this section we describe primer design and common physical properties of 3C primers.

2.3.1. Primer design in the region of interest—3C primers are designed for all restriction fragments of interest. For correct interpretation of the data it is important to not only interrogate interactions between pairs of loci of interest, but to obtain a more comprehensive interaction profile throughout the region. In general this profile will show an inverse relationship between interaction frequency and genomic distance (Figure 3). A looping interaction is then inferred when a peak on top of this overall profile is observed [12, 13].

An example 3C experiment determines whether a given genomic element, for instance a gene promoter, is engaged in a long-range interaction with one or more distally located elements, such as enhancers. When the positions of these distal elements are not known, one designs 3C primers for the promoter and all restriction fragments throughout the region under study. If the location of putative distal elements is known, one designs primers for the corresponding fragments, but also for a number of flanking fragments located in between the promoter and the distal elements to obtain a larger interaction profile. If flanking regions are excluded, one cannot conclude which fragment contains the actual point of looping contact.

2.3.2. Primer design in the control region—As described above a control region is included in 3C studies to allow for comparison of 3C data obtained in different cell types or

Note 2.2.2 **Exclusion of Restriction Fragments:** We have found that very large, and very small fragments can sometime yield aberrant interaction frequencies. This might be due to differences in intra-molecular ligation efficiency for very long DNA fragments. Therefore, we recommend avoiding, if at all possible, very long (>10Kb) restriction fragments.

conditions. We recommend designing several primers throughout the region, so a control 3C interaction profile can be obtained that covers a similar genomic distance as that obtained in the region of interest. 3C primers are designed for at least 10 different restriction fragments spaced at various distances.

2.3.3. Direction of the primers—We strongly encourage researchers to use a unidirectional primer design. In such a design all primers are oriented in the same direction, on the same DNA strand, along the chromosomal region of interest. All pairs of primers will amplify ligation products that are the result of head-to-head ligation of the corresponding restriction fragments. A unidirectional primer design is important because it avoids amplification of non-informative ligation products (shown in Supplementary Figure 2). First, when primers are used that point away from each other in the linear genome one runs the risk of accidentally amplifying a self-ligated partial digestion product. Second, primers for two directly adjacent restriction fragments that point towards each other in the linear genome sequence will produce a PCR product even when the restriction site between them was never cut in the 3C experiment.

2.3.4. Physical properties of the primers—To increase specificity of the primers we recommend designing long primers with high melting temperature (on average the T_m is 90°C); the length of 3C primers is 28–30bp with a GC content of ~50%, preferably carrying a single G or C nucleotide on the 3' end. We have found that the use of rather long primers is especially important for complex genomes, where short 20bp primers do not provide necessary specificity and efficiency. Primers are designed ~80–150bp away from the restriction cut site so that the predicted amplicon will be between 160 and 300bp in size. We recommend checking the uniqueness of each primer (See Note 2.3.4 Checking Primer Uniqueness).

2.4. Creating a Control Library

3C employs PCR with locus-specific primers, which may amplify their target ligation product with different efficiencies, even when great care is taken to design primers. It is therefore critical to correct for this differential primer efficiency. This can be done by PCR analysis of a control library that contains all interrogated ligation products in equimolar amounts. Any differences in PCR product formation obtained by pairs of 3C primers with this control library as template can then be used to estimate primer pair efficiency.

A control library is prepared by digesting and randomly ligating non-crosslinked purified DNA. For small genomes (yeast, bacteria, fly) the control library can be generated using purified genomic DNA. For larger genomes, such as mouse or human, a genome-wide random control library is too complex to allow reliable detection of individual ligation products. For these organisms the control library can be made from one or more bacterial artificial chromosomes (BACs) that span the genomic region(s) investigated by 3C, including the control region (e.g. [16]). When multiple BACs are used they should be selected so that they display minimal overlap while simultaneously keeping the number and size of gaps to a minimum. This ensures minimal over- or under- representation of genomic regions in the control library. If BACs are unavailable for a genomic region, they might be substituted with fosmids, cosmids or even plasmids. The control library is then generated by

Note 2.3.4 **Checking Primer Uniqueness:** We recommend using both BLAST and BLAT for checking uniqueness of 3C primers as those programs have different algorithms of searching for a match in a genome. BLAT works much quicker, however BLAST gives more comprehensive results. It is also possible to check if primers have strong secondary structures (hairpins) and form stable homodimers. Primers should also be checked for formation of heterodimers with anchor primer. Free online tools such as IDT oligo analyzer (<http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/>) can be used for this analysis.

mixing the clones in equimolar amounts and digesting the DNA followed by random intermolecular ligation.

2.5. Determination of Interaction Frequencies

It is important to first determine the optimal amount of 3C library to use in each PCR reaction. This amount has to be found experimentally for each 3C library in a titration experiment (Figure 2). To build a titration curve, a series of PCR reactions with a single pair of 3C primers and different amounts of input 3C template must be done. Supplementary Table 1 gives examples of pairs of 3C primers which were successfully used in our lab for the titration of human and mouse libraries. We recommend selecting the library concentration from the middle of a linear region of an amplification curve in order to avoid both saturation of a signal (at high library concentrations) and loss of a signal (at low library concentration).

Next PCR reactions are performed with each primer pair, using both the 3C ligation product library and the control ligation product library as a template. The relative interaction frequency of a pair of loci is then calculated by dividing the amount of PCR product obtained with the 3C ligation product library by the amount of PCR product obtained with the control library (see section 4 for data analysis). By calculating this ratio one effectively normalizes for differences in primer efficiency. Since the control library is used to normalize for primer efficiency, reactions for each primer pair should be performed in both templates simultaneously, to reduce PCR variation as much as possible. Given that the control library contains all ligation products in equimolar amounts, all primer pairs should yield similar, though not identical, amounts of PCR products. When a pair of primers fails to amplify any product, or a product with the wrong size, these primers should be discarded and new primers should be designed (Fig. 4).

Section 3: Experimental Protocol

3.1 Reagents required

Buffers

10X Ligation buffer

- 500 mM Tris-HCl pH7.5
- 100 mM MgCl₂
- 100 mM Dithiothreitol
- Autoclaved water
- Store 20ml and 1ml aliquots at -80°C

1X Lysis buffer

- 10 mM Tris-HCl pH8.0
- 10 mM NaCl
- 0.2% Igepal CA-630 (NP40)
- Autoclaved water
- Store at 4°C

10X PCR buffer

- 600 mM Tris, adjust to pH 8.9 with H₂SO₄,

180 mM (NH₄)₂SO₄

--- Store 1ml aliquots at -20°C

Reagents

Reagent	Company	Catalogue number
Saturated phenol, pH 6.6±0.2 *bring pH to 8.0	Fisher scientific	BP1750-400
Chloroform	Mallinckrodt CHEMICALS	4440-04
Formaldehyde, 37% by weight	Fisher scientific	BP531-25
ATP	Sigma	A9187-1G
Proteinase K (Fungal)	Invitrogen	25530-031
Protease inhibitors	Sigma	P8340-5ml
T4 DNA ligase	Invitrogen	15224
Amicon® Ultra – 0.5ml 30K	Millipore	UFC5030BK

3.2 Creating a BAC control library (for use with complex genomes)

Digestion of BAC genomic DNA

- 1 Digest the DNA in the following reaction overnight at 37°C with rotation
 - 10µg BAC DNA *
 - 10X restriction buffer
 - 10 mg/ml BSA (100X)
 - Restriction enzyme (up to 10% of total reaction volume)
 - Water (up to 500µl)
- 2 After digestion, add an equal volume of phenol (pH 8.0)/chloroform (1:1), vortex for 30 seconds
- 3 Spin at 18,000 x g in a microfuge at room temperature for 5 minutes
- 4 Transfer the upper (aqueous) phase to a new 2 ml tube and add 1/10 volume of 3M sodium acetate, pH 5.2, vortex briefly
- 5 Add 2.5 volume of ice-cold 100% ethanol and invert the tube gently
- 6 Incubate at -20°C for at least 15 minutes
- 7 Spin at 18,000 x g in a microfuge at 4°C for 20 minutes
- 8 Wash the pellet in 1ml of 70% ethanol
- 9 Spin at 18,000 x g in a microfuge at 4°C for 15 minutes
- 9 Remove all supernatant
- 10 Briefly air-dry the pellet
- 11 Resuspend the pellet in 44µl water and incubate at 37°C for 15 minutes. Reserve 1µl of digestion product to run as a control on the final analytical gel.

*If using more than one BAC, combine them in equimolar amounts before digestion. Concentration of each BAC clone should be determined using qPCR with primers that recognize the common BAC vector backbone.

BAC DNA ligation

- 12 Prepare the ligation reaction as follows and incubate at 16°C overnight

Digested BAC DNA	43µl
5X T4 ligase buffer	12µl
T4 DNA ligase	5µl
Final total volume	60µl

- 13 After overnight ligation, incubate at 65°C for 15 minutes to inactivate the ligase

Purification of BAC genomic DNA control template

- 14 Add 140µl of water to the ligation to make the final volume 200µl
- 15 Add an equal volume of phenol (pH 8.0)/chloroform (1:1), vortex 30 seconds and then spin at 18,000 x g in a microfuge at room temperature for 5 minutes
- 16 Transfer the upper (aqueous) phase to a fresh 1.7 ml tube
- 17 Repeat phenol/chloroform extraction once
- 18 Transfer the upper phase to a fresh 1.7 ml tube
- 19 Add equal volume of chloroform, vortex 30 seconds, then spin at 18,000 x g at room temperature for 5 minutes
- 20 Transfer the upper phase to a new 1.7 ml tube
- 21 Add 1/10 volume of 3M sodium acetate, pH 5.2, briefly vortex
- 22 Add 2.5 volumes of ice-cold 100% ethanol and invert the tube several times
- 23 Incubate at -20°C for at least 15 minutes
- 24 Spin at 18,000 x g in a microfuge at 4°C for 20 minutes
- 25 Wash the pellet in 1 ml of 70% ethanol, spin at 18,000 x g in a microfuge at 4°C for 15 minutes
- 26 Briefly air-dry the pellet
- 27 Resuspend the pellet in 100µl of 1X TE, pH 8.0
- 28 Incubate at 37°C for 15 minutes
- 29 Store the 3C control template at -20°C. This library can be stored for a few years.

3.3 Creating a 3C library from mammalian cells**Crosslinking of cells**

- 1 Centrifuge 1×10^8 cells at 300 x g for 10 minutes*
- 2 Resuspend the cell pellet in 45 ml of fresh medium, mix by pipetting
- 3 Add 1.25 ml of 37% formaldehyde to obtain 1% final concentration and pipette to mix
- 4 Incubate at room temperature for 10 minutes on a shaker

- 5 Add 2.5 ml of 2.5M glycine to stop crosslinking
- 6 Incubate at room temperature for 5 minutes and then on ice for 15 minutes
- 7 Spin down at 4°C 800 x g for 10 minutes and remove as much supernatant as possible**
- 8 Add 2 ml of cold Lysis buffer and 200 µl of 10X Protease Inhibitors to each pellet and resuspend it well on ice
- 9 Incubate on ice for at least 15 minutes to let the cells swell
- 10 Lyse the cells on ice using a Dounce homogenizer (pestle A). Use 10 slow strokes, then let the cells rest for 1 minute on ice. Complete lysis by applying 10 more strokes
- 11 Transfer to 2 microcentrifuge tubes, spin at 2,200 x g in a microfuge for 5 minutes
- 12 Wash each pellet with 1 ml of 1X NEB buffer 2 (or the buffer appropriate to your restriction enzyme, however do not use NEB4 as we have found it is not compatible with 3C buffers) and spin down at 2,200 x g in a microfuge for 5 minutes
- 13 Repeat Step 12 once
- 14 Resuspend all the pellets from the same sample in 1 ml of 1X NEB buffer

Digestion

- 15 Distribute the suspension over 20 1.7 ml microcentrifuge tubes, 50 µl per tube. (Optional) save 10µl of lysate for the chromatin integrity check.
- 16 (Optional) Chromatin integrity control. Take 10 µl of lysed cells from the previous step. Add 50 µl of 1x NEB and 10 µl of Proteinase K (10mg/ml). Incubate for 30 min at 65°C. Purify DNA by single phenol-chloroform extraction without ethanol precipitation. Add 5 µl of RNaseA (10mg/ml) to the aqueous phase and incubate for 15 min at 37°C. Check quality of the sample by running it on 0.7% agarose gel. The sample is good if DNA is either stuck in the well or runs as a single high molecular weight band (>23 kb).
- 17 Add 312 µl of corresponding 1X NEB buffer per tube.
- 18 Add 38 µl of 1% SDS per tube and mix well by pipetting up and down, avoiding air bubbles
- 19 Incubate at 65°C for exactly 10 minutes, then place on ice immediately*
- 20 Add 44 µl of 10% Triton X-100 to each tube to quench SDS, mix well, avoiding air bubbles

*In cases, when primary cells [38], [39], [40] or siRNA treated cells [38] are being analyzed by 3C, getting 10⁸ cells might be very difficult. Several groups has successfully applied 3C to 2–10 × 10⁶ cells and in some of those studies the original protocol has been modified and we advice to refer to the original studies. In general, while working with primary tissues it is necessary to break down tissues by applying collagenase to single-cell suspension before cross-linking [41]. Time of lysis can be increased for up to 2h as well as treatment with SDS prior restriction can be more severe [39]. At last, one might want to use qPCR to quantify 3C signal [41]. When number of cells is not an issue, we strongly recommend starting with large number of cells, so one would have enough material for making repeats in case it is needed.

**The experiment can stop at this point. Store the pellet at –80°C by incubating the pellet on dry ice for 20 minutes and then store at –80°C for up to at least two years.

*This step removes proteins that are not crosslinked to the DNA

- 21 Add 400 U of the restriction enzyme (up to 10% total reaction volume) of choice to each tube, mix well
- 22 Incubate at 37°C overnight with rotation

Ligation and reverse crosslink

- 23 Add 86 µl of 10% SDS to each tube, mix carefully to avoid making air bubbles
- 24 Incubate at 65°C for 30 min to inactivate the enzyme. Place tubes on ice immediately after incubation
- 25 Prepare the ligation cocktail master mix as follows (except ligase) and distribute 7.61 ml of cocktail to each of 20 pre-chilled 15 ml conical tube

Ligation cocktail	per reaction
10% Triton X-100	745 µl
10X Ligation buffer	745 µl
10 mg/ml BSA	80 µl
100 mM ATP	80 µl
Milli-Q water	5960 µl

- 26 Transfer each digestion product from Step 23 into each of the 15 ml conical tubes with ligation cocktail. Keep on ice.
- 27 Add 5000 U of T4 DNA ligase (Invitrogen) to each tube
- 28 Mix by inverting the tubes several times and spin briefly
- 29 Incubate at 16°C for 2 hours
- 30 Add 50 µl of fresh 10 mg/ml proteinase K solution per tube
- 31 Mix by inverting the tubes several times and spin briefly
- 32 Incubate at 65°C for 4 hours
- 33 Add 50 µl of 10 mg/ml fresh proteinase K per tube and continue incubating at 65°C overnight. *Alternatively, one can incubate overnight after the first proteinase K treatment, add the second proteinase K aliquot the next morning and incubate at 65°C for 2 hours*

DNA Purification

- 34 Transfer and combine reactions into 50 ml conical tubes, 20 ml per tube
- 35 Add equal volume Phenol per tube, vortex for 2 minutes and spin at 3,000 x g for 10 minutes
- 36 Transfer the aqueous phase to fresh 50 ml conical tubes
- 37 Add an equal volume of Phenol (pH 8.0)/Chloroform (1:1) to each tube, vortex for 1 minute and spin at 3,000 x g for 10 minutes
- 38 Pool aqueous phases and distribute equally in new 50 ml conical tubes
- 39 Add 1x TE buffer, pH 8.0 to a total volume of 50 ml per tube to prevent DTT from precipitating.

- 40 Transfer each 50 ml pool to a 250-ml screw-cap centrifuge bottle that is suitable for high-speed spinning
- 41 Add 1/10 volume of 3M sodium acetate, pH 5.2 per bottle, vortex briefly
- 42 Add 2.5 volumes of ice-cold 100% ethanol per bottle, mix well by inverting the bottles several times
- 43 Incubate at -80°C at least 60 minutes or overnight
- 44 Spin at $12,000 \times g$ for 20 minutes at 4°C (Make sure the mixture has thawed completely on ice before spinning down)
- 45 Dissolve each pellet in $450 \mu\text{l}$ 1x TE buffer, pH 8.0 and transfer to a 15 ml conical tube
- 46 Wash the bottle with $450 \mu\text{l}$ of 1x TE buffer, pH 8.0 and transfer to the same tubes as in step 44. *If dissolving is difficult, incubate samples at 65°C no more than 10 minutes*
- 47 Add an equal volume of phenol (pH 8.0) to each tube and vortex for 1 minute; then spin at room temperature at $3,000 \times g$ for 5 minutes
- 48 Transfer the upper aqueous phases to fresh 15 mL tubes
- 49 Add equal volume of phenol (pH 8.0)/chloroform (1:1), vortex 1 minute and spin at $3,000 \times g$ for 5 minutes
If the interfaces are clear, go to step 49; otherwise repeat step 48 once
- 50 Transfer the aqueous phases into clean 1.7 ml centrifuge tubes (about $400\mu\text{l}$ per tube)
- 51 Add 1/10 volume of 3M sodium acetate, pH 5.2, vortex briefly
- 52 Add 2.5 volumes of ice-cold 100% ethanol, invert gently
- 53 Incubate at -80°C at least 1 hour
- 54 Spin at 4°C at full speed in a microfuge for 20 minutes
- 55 Remove all supernatant and briefly air dry the pellets
- 56 Dissolve each pellet in $500 \mu\text{l}$ of 1X TE buffer, pH 8.0
- 57 Insert an Amicon ultra-0.5 30K device into one provided microcentrifuge tube
- 58 Transfer DNA sample to the Amicon ultra filter device (up to $500 \mu\text{l}$) and cap it
- 59 Place the capped devices into the rotor of a microfuge, aligning the cap strap towards the center of the rotor
- 60 Spin at $18,000 \times g$ for 5 minutes and discard the flow-through from the collecting tubes
- 61 Add $450 \mu\text{l}$ of 1X TE to each filter device, spin at $18,000 \times g$ for 5 minutes and discard the flow-through
- 62 Repeat wash step 3 times
- 63 After the final wash, place the filter device upside down in a new provided collecting tube
- 64 Place the tubes in centrifuge, aligning the cap strap towards the center of the rotor, spin at $1,000 \times g$ for 5 minutes

- 65 Transfer all collected DNA from the same sample to a 1.7ml centrifuge tube
- 66 Add 1X TE to each sample to make a total volume of 1ml
- 67 Add 1 μ l of 10 mg/ml of DNase-free, RNase A and incubate at 37°C for 15 minutes
- 68 Load 1 μ l and 2 μ l of 10X diluted template on 0.8% agarose/0.5X TBE gel, along with a sample of a 1Kb DNA ladder to check quality and quantity of the template
- 69 Store the 3C template up to 2 years at -20°C (kept in aliquots)

3.4 Determination of quantity and quality of 3C Libraries

Before embarking on quantitative analysis of the 3C library, one first has to determine the amount of 3C library to use in each PCR reaction. To do this, one can perform a titration experiment, as shown in Figure 2. Both the BAC control template and the experimental 3C library should be titrated using a serial two-fold dilution series beginning with 240 ng of 3C template and 25–50ng of (BAC) control template (Note 3.4 BAC Dilution Series). We routinely use two different primer pairs for the control region for this analysis. The first primer pair interrogates a short-range interaction (i.e. a pair of restriction fragments separated by only a few thousand bp in the genome). The second primer pair is chosen to interrogate a longer-range interaction (i.e. a pair of restriction fragments separated by tens of thousands of bp). We suggest performing each PCR reaction in duplicate. PCR products are run on a 2% agarose gel and quantified using a standard gel imaging set up. A water control should be included.

The amount of PCR product is then quantified and plotted versus the amount of input DNA. The resulting titration curve should plateau to a flat shoulder, as shown in Figure 2D and E. The concentration of 3C template to use in 3C experiments should be taken from the linear slope of the graph to ensure that one will not over-or under-saturate signals from the 3C library.

The PCR reaction is assembled as follows:

Titration Reaction

10x PCR Buffer	2.5 μ l
50 mM MgSO ₄	2 μ l
20 mM dNTPs	0.2 μ l
80 μ M Primer1	0.125 μ l
80 μ M Primer2	0.125 μ l
Taq Polymerase	0.2 μ l
Diluted template	4 μ l
dH ₂ O	<u>15.85 μl</u>
Total Volume	25 μ l

Note 3.4 **BAC Dilution Series:** The starting amount of a control template depends on the BAC composition of the template: the more complex the control library is, i.e. the more possible interactions it is covering, the more DNA is needed in a PCR reaction. In our experience, using BAC clones covering up to a megabase of genomic DNA, starting the dilution series with 50–70 ng of the template works well.

PCR conditions

1. 95 C 5 minutes
2. 95 C 30 seconds
3. 65 C 30 seconds
4. 72 C 30 seconds
5. 95 C 30 seconds
6. 65 C 30 seconds
7. 72 C 8 minutes
8. 10 C forever

Repeat steps 2–4 34 times.

3.4 Generating 3C Data

After choosing the appropriate concentration of both the control and the experimental 3C library from the titration analysis, one can start to determine interaction frequencies between pairs of loci. To do so, one uses pairs of primers for restriction fragments of interest to perform semi-quantitative PCR on each of the two templates. Each PCR reaction is performed in triplicate. The PCR conditions are identical to the ones used to titrate the 3C and control libraries (section 3.4). PCR products are run on a 2% agarose gel and the amount of PCR product is quantitated using a standard gel quantification set up. 3C products can also be quantified using qPCR, with very similar results [42].

Section 4: Analysis

A typical 3C experiment includes the analysis of three biological replicates of the 3C and control library. Further, each interaction frequency of interest is determined by three PCR reactions (technical replicates) using each of the three 3C and control libraries. Ideally, all 3C reactions should be prepared with the same PCR master mix and run simultaneously in the same PCR block, but practically it is not always possible if the experiment covers a large region. To minimize experimental noise we recommend that PCR replicates for the 3C library and the control library are performed in parallel and run side-by-side. We use LabWorks software (version 4.0, BioImaging Systems) to analyze the intensity of each band minus the background on an agarose gel. Then one calculates the average of the three technical replicates. Thus for each biological replicate one obtains an average value for the interaction frequency of each pair of loci. Finally, the three datasets obtained with the three biological replicates are normalized to each other so that they are all on the same scale. This allows the data from different replicates and different conditions to be directly compared. Below we present an illustration of how 3C data can be calculated and how 3C datasets obtained for different cells or conditions can be quantitatively compared.

4.1 Example of a 3C Analysis

We describe a 3C analysis of a gene in two cell lines, A and B. A BAC-based control library was also generated. In this example, cell line A expresses the gene of interest while cell line B does not. In this analysis interactions between a single anchor restriction fragment, containing the gene promoter, and 20 flanking restriction fragments were determined to generate a 3C interaction profile. In addition, a control region (ENCODE region Enr313) was analyzed to obtain a set of interactions that are assumed to be identical in cell line A and B.

4.2 Calculation of interaction frequencies for each biological replicate

The first step in analyzing a 3C experiment is to average the technical replicates. Simply average the technical replicates and find the standard deviation for each primer pair. In order to control for primer efficiency, divide each averaged technical replicate by its corresponding averaged control template value. In this example, for each pair of primers the averaged value of three PCR reactions performed on a given 3C library for each cell line (A or B) is divided by the corresponding average of three PCR reaction performed in the BAC control library. For example, the interaction frequency for a given primer pair is (Equation 1):

$$\frac{(A_1+A_2+A_3)}{(CL_1+CL_2+CL_3)}$$

where the values A1, A2 and A3 represent three technical replicates for that primer pair in cell line A, and CL1, CL2 and CL3 represent three technical replicates for the same primer pair in the control library. This value is the interaction frequency of a pair of loci for a given biological replicate. In order to calculate the standard deviation of each interaction frequency, use the following formula (Equation 2):

$$StDev = \text{sqrt} \left[\left(\frac{StDev_{Control\ Library}}{Avg_{Control\ Library}} \right)^2 + \left(\frac{StDev_{Exp. Library}}{Avg_{Exp. Library}} \right)^2 \right] * Interaction\ Frequency$$

where StDev is the standard deviation, Avg is the average, Exp. Library is the experimental 3C library value, and Interaction Frequency is the value calculated using Equation 1.

Using this approach the average interaction frequency, and the standard deviation, for each pair of loci is determined for each of the three biological replicate 3C libraries.

4.4 Combining biological replicates

The results from individual biological replicates can be directly compared in separate graphs (to determine whether the same peaks occur in each replicate) or the biological replicates can be averaged together. To calculate the combined standard deviation of all biological replicates, use the following formula (Equation 3):

$$StDev = \text{sqrt} \left[\left(\frac{StDev_{B.Rep1}}{Avg_{B.Rep1}} \right)^2 + \left(\frac{StDev_{B.Rep2}}{Avg_{B.Rep2}} \right)^2 + \left(\frac{StDev_{B.Rep3}}{Avg_{B.Rep3}} \right)^2 \right] * Avg_{All\ Reps}$$

where StDev is the standard deviation, Avg is the average, B is biological, and Rep is replicate. For more information on why to use the standard deviation and not the standard error of the mean when combining biological replicates, please see Note 4.4.

Note 4.4 **Standard Deviation versus Standard Error of the Mean.** We suggest that when plotting the average of biological replicates to also display the standard deviation (SD) of each data point and not the standard error of the mean (SEM). The standard error of the mean reflects the certainty with which the average can be estimated. The SEM incorporates the number of measurements taken because the more measurements made, the more likely it is that the correct value has been found. This is a valid error to plot while examining an individual biological replicate. A large SEM indicates that the value of a give data point is very uncertain, and it may be necessary to perform additional technical replicates to increase the precision of measuring that specific value. However, it is more informative to indicate the SD while comparing biological replicates, since the SD will better reflect potentially relevant variation between samples.

4.5 Normalizing 3C data obtained from different experiments

In order to allow direct quantitative comparison between two 3C datasets, e.g. two different biological replicates and/or data obtained with two different cell lines, they must be first normalized to each other. This normalization is done using the interaction frequencies measured within the control genomic region, which was selected based on the assumption that it has the same conformation in both cell lines. In our example we have analyzed two cell lines (A and B), and have three biological replicates. Here we provide an example of how these datasets can be compared. We will normalize the data for cell line B to the data for cell line A, for each of the three biological replicates separately, so that three independent cell line comparisons are obtained.

First, the interaction frequencies are calculated for each biological replicate for cell line A and B, as described above, for each of the pairs of loci in the control region. Next, a normalization factor is calculated to normalize the data for one experiment to the other, e.g. to normalize the data obtained with cell line B to the data obtained with cell line A. For this, determine the log ratio for each interaction frequency in the control region (Equation 4):

$$\log \frac{\sum A_1 + A_2 + A_3}{\sum B_1 + B_2 + B_3}$$

where A1, A2 and A3 are the normalized interaction frequencies of one primer pair in the first cell line, and B1, B2 and B3 are the normalized interaction frequencies of one primer pair in the second cell line. Next, the average of these log ratios is calculated. The normalization factor is then found by taking the inverse log of this average. Finally, the entire dataset that was in the denominator in the calculation, in this case, cell line B biological replicate 1, is multiplied by the normalization factor. Each normalized interaction frequency is multiplied by the normalization factor. This will bring the two datasets to the same scale, and they can now be plotted on the same graph. This analysis is then repeated for replicates 2 and 3 individually. One can subsequently average the normalized biological replicates together to obtain a final data set for each cell line, and these final data sets can be plotted on one graph.

4.6 Find Peaks of High Interaction Frequency

3C interaction frequencies are typically plotted versus genomic position with respect to the anchor point (Figure 3). In general interaction frequencies will decrease rapidly with increasing genomic distance. A specific looping interaction can be inferred when a peak is observed on top of this overall 3C profile. Visual inspection of 3C profiles has been used to identify such looping interactions. To obtain further support for a looping interaction additional analyses are essential, e.g. 3C analysis of cells or conditions where the looping interaction is absent. Figure 3 also illustrates the importance of obtaining a larger 3C profile so that a local background in non-specific 3C interactions is obtained. In the absence of this baseline estimation it is not possible to identify peaks in 3C interaction profiles and can lead to misinterpretation of individual 3C signals. In some cases it is possible to apply ANOVA statistic test for peak calling on the 3C profile, but usually there are not enough data points to perform in-depth analysis [43].

If a long-range interaction is inferred, it may be necessary to perform further experiments to validate the finding. For instance one can analyze the genomic region using a different restriction enzyme to confirm the looping interaction. Comparison of 3C data to other types of data sets, such as histone modification patterns or DNase hypersensitive sites, and analysis of the looping interaction across cell types that do or do not express the gene of

interest can help further define the functional elements involved and the role of the interaction in gene expression. Final experimental confirmation of the looping interaction and the DNA elements involved can be obtained by deleting or mutating the interacting regions and/or knocking down transcription factors that may mediate the interaction.

Section 5 Troubleshooting

5.1 Degraded template

Degradation of a 3C template can be observed when the DNA is run on an agarose gel (Figure 4A). In our experience this degradation occurs early in the 3C protocol, often at the step where cells are lysed. This may be due to contaminating nucleases. The quality of a cell pellet can be checked on the first steps of the 3C protocol (see pp. 15 and 16 (optional) of the protocol). If 3C template is degraded we recommended replacing all plastics and buffers before redoing the experiment.

5.2 Linear range of PCR amplification of 3C template cannot be determined with titration PCR

This problem is most likely the result of high salt concentrations in the 3C library preparation. The use of Amicon columns typically removes most salt. However, if this problem is observed, the 3C template can be re-purified with phenol/chloroform extraction, ethanol precipitated and washed again on Amicon columns.

5.3 Titration PCR on 3C template is very inefficient for all pairs of primers while primers works fine on the control template

For large genomes, e.g. from mammals, PCR amplification of ligation products using the control library is much more efficient than for the experimental 3C library. To optimize PCR amplification of ligation products with the 3C library it may be important to further optimize the PCR conditions, including the time and temperature of annealing and the concentration of magnesium ions in the PCR buffer. Several PCR primer pairs for human and mouse, which have been used in our lab and reproducibly given good titration curves, are listed in supplement Table 1.

Poor PCR amplification of ligation products with the 3C library can also be the result of inefficient digestion and/or ligation during the 3C procedure. Restriction efficiency can be estimated by taking an aliquot of chromatin right after digestion in the 3C protocol. DNA is then purified and analyzed by PCR with primers designed to amplify a genomic region containing a restriction site. An equal amount of DNA purified from an undigested chromatin sample should be used as a control. Digestion efficiency in the 3C protocol is then defined as the ratio of the amount of PCR product obtained with the 3C DNA divided by the amount obtained with genomic DNA. 3C digestion is considered successful when a more than 70% reduction of PCR product is observed for the digested 3C template compared to the undigested genomic DNA template.

5.4 PCR amplification of ligation products lead to multiple bands on the gel

There are might be several reasons for multiple bands in a 3C PCR (see Figure 4B–E). We recommend approaching this problem step by step, finding and eliminating each possible reason (see Figure 4B–E). First of all, multiple amplified DNA fragments can be the result of the incomplete digestion of the cross-linked chromatin, as is typical in 3C experiments (Figure 4C). This problem might be especially severe when frequently cutting enzymes are used, as the average size of restriction fragments (i.e. potential insert in the ligation product) is small and hence non-canonical ligation products can be amplified efficiently. If the extra bands are very prominent, one may consider re-doing 3C template with increased

concentration of restriction enzyme and extended incubation time. Multiple bands might also be the result of non-specific annealing of PCR primers. In that case, we recommend first to modify PCR condition (increase annealing temperature, Figure 4D) and if that does not help to redesign the primers (Figure 4E).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors wish to thank members of the Dekker lab for helpful discussions and criticisms of the manuscript.

References

1. Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet.* 2001; 2(3):292–301. [PubMed: 11283701]
2. Miele A, Dekker J. Long-range chromosomal interactions and gene regulation. *Mol BioSyst.* 2008; 4(11):1046–57. [PubMed: 18931780]
3. Misteli T, Soutoglou E. The emerging role of nuclear architecture in DNA repair and genome maintenance. *Nat Rev Mol Cell Biol.* 2009; 10(4):243–254. [PubMed: 19277046]
4. Dekker J. Gene regulation in the third dimension. *Science.* 2008; 319(5871):1793–1794. [PubMed: 18369139]
5. Dean A. In the loop: long range chromatin interactions and gene regulation. *Brief Funct Genomics.* 2011; 10(1):3–10. [PubMed: 21258045]
6. Fraser P, Bickmore W. Nuclear organization of the genome and the potential for gene regulation. *Nature.* 2007; 447(7143):413–417. [PubMed: 17522674]
7. Bantignies F, et al. Polycomb-dependent regulatory contacts between distant Hox loci in *Drosophila*. *Cell.* 2011; 144(2):214–226. [PubMed: 21241892]
8. Guelen L, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature.* 2008; 453(7197):948–951. [PubMed: 18463634]
9. Belmont AS. Visualizing chromosome dynamics with GFP. *Trends Cell Biol.* 2001; 11(6):250–257. [PubMed: 11356361]
10. van Steensel B, Dekker J. Genomics tools for unraveling chromosome architecture. *Nat Biotechnol.* 2010; 28(10):1089–1095. [PubMed: 20944601]
11. Dekker J, et al. Capturing Chromosome Conformation. *Science.* 2002; 295(5558):1306–1311. [PubMed: 11847345]
12. Dekker J. The 3 C's of Chromosome Conformation Capture: Controls, Controls, Controls. *Nat Methods.* 2006; 3(1):17–21. [PubMed: 16369547]
13. Sanyal A, et al. The long-range interaction landscape of gene promoters. *Nature.* 2012 in press.
14. Simonis M, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet.* 2006; 38(11):1348–1354. [PubMed: 17033623]
15. Zhao Z, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet.* 2006; 38(11):1341–1347. [PubMed: 17033624]
16. Dostie J, et al. Chromosome Conformation Capture Carbon Copy (5C): A Massively Parallel Solution for Mapping Interactions between Genomic Elements. *Genome Res.* 2006; 16(10):1299–1309. [PubMed: 16954542]
17. Fullwood MJ, et al. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature.* 2009; 462(7269):58–64. [PubMed: 19890323]
18. Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 2009; 326(5950):289–293. [PubMed: 19815776]

19. Sanyal A, et al. Chromatin globules: a common motif of higher order chromosome structure? *Curr Opin Cell Biol.* 2011; 23(3):325–331. [PubMed: 21489772]
20. Hakim O, Misteli T. SnapShot: Chromosome conformation capture. *Cell.* 2012; 148(5):1068.e1–2. [PubMed: 22385969]
21. Tolhuis B, et al. Looping and Interaction between Hypersensitive Sites in the Active beta-globin Locus. *Mol Cell.* 2002; 10(6):1453–1465. [PubMed: 12504019]
22. Gheldof N, Tabuchi TM, Dekker J. The active FMR1 promoter is associated with a large domain of altered chromatin conformation with embedded local histone modifications. *Proc Natl Acad Sci U S A.* 2006; 103(33):12463–12468. [PubMed: 16891414]
23. Gheldof N, et al. Cell-type-specific long-range looping interactions identify distant regulatory elements of the CFTR gene. *Nucleic Acids Res.* 2010; 38(13):4235–4336.
24. Drissen R, et al. The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev.* 2004; 18(20):2485–2490. [PubMed: 15489291]
25. Splinter E, et al. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* 2006; 20(17):2349–2354. [PubMed: 16951251]
26. Vakoc CR, et al. Proximity among Distant Regulatory Elements at the beta-Globin Locus Requires GATA-1 and FOG-1. *Mol Cell.* 2005; 17(3):453–462. [PubMed: 15694345]
27. Spilianakis CG, et al. Interchromosomal associations between alternatively expressed loci. *Nature.* 2005; 435(7042):637–645. [PubMed: 15880101]
28. Davison LJ, et al. Long-range DNA looping and gene expression analyses identify DEXI as an autoimmune disease candidate gene. *Hum Mol Genet.* 2012; 21(2):322–333. [PubMed: 21989056]
29. Wright JB, Brown SJ, Cole MD. Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. *Mol Cell Biol.* 2010; 30(6):1411–1420. [PubMed: 20065031]
30. Ahmadiyeh N, et al. 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. *Proc Natl Acad Sci U S A.* 2010; 107(21):9742–9746. [PubMed: 20453196]
31. Comet I, et al. A chromatin insulator driving three-dimensional Polycomb response element (PRE) contacts and Polycomb association with the chromatin fiber. *Proc Natl Acad Sci U S A.* 2011; 108(6):2294–2299. [PubMed: 21262819]
32. Montavon T, et al. A regulatory archipelago controls Hox genes transcription in digits. *Cell.* 2011; 147(5):1132–1145. [PubMed: 22118467]
33. Dekker J. GC- and AT-rich chromatin domains differ in conformation and histone modification status and are differentially modulated by Rpd3p. *Genome Biol.* 2007; 8(6):R116. [PubMed: 17577398]
34. Kim SI, et al. BRG1 requirement for long-range interaction of a locus control region with a downstream promoter. *Proc Natl Acad Sci U S A.* 2009; 106(7):2259–2264. [PubMed: 19171905]
35. Amano T, et al. Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev Cell.* 2009; 16(1):47–57. [PubMed: 19097946]
36. Bacher CP, et al. Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nat Cell Biol.* 2006; 8(3):293–299. [PubMed: 16434960]
37. Xu N, et al. Evidence that homologous X-chromosome pairing requires transcription and Ctfc protein. *Nat Genet.* 2007; 39(11):1390–1396. [PubMed: 17952071]
38. Smits BM, et al. An insulator loop resides between the synthetically interacting elements of the human/rat conserved breast cancer susceptibility locus MCS5A/Mcs5a. *Nucleic Acids Res.* 2012; 40(1):132–47. [PubMed: 21914726]
39. Vu TH, Nguyen AH, Hoffman AR. Loss of IGF2 imprinting is associated with abrogation of long-range intrachromosomal interactions in human cancer cells. *Hum Mol Genet.* 2010; 19(5):901–19. [PubMed: 20015958]
40. Montavon T, et al. A regulatory archipelago controls Hox genes transcription in digits. *Cell.* 2011; 147(5):1132–45. [PubMed: 22118467]
41. Hagege H, et al. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc.* 2007; 2(7):1722–33. [PubMed: 17641637]

42. Hagège H, et al. Quantitative analysis of Chromosome Conformation Capture assays (3C-qPCR). *Nat Protoc.* 2007; 2(7):1722–1733. [PubMed: 17641637]
43. McCord RP, et al. Distant cis-regulatory elements in human skeletal muscle differentiation. *Genomics.* 2011; 98(6):401–411. [PubMed: 21907276]

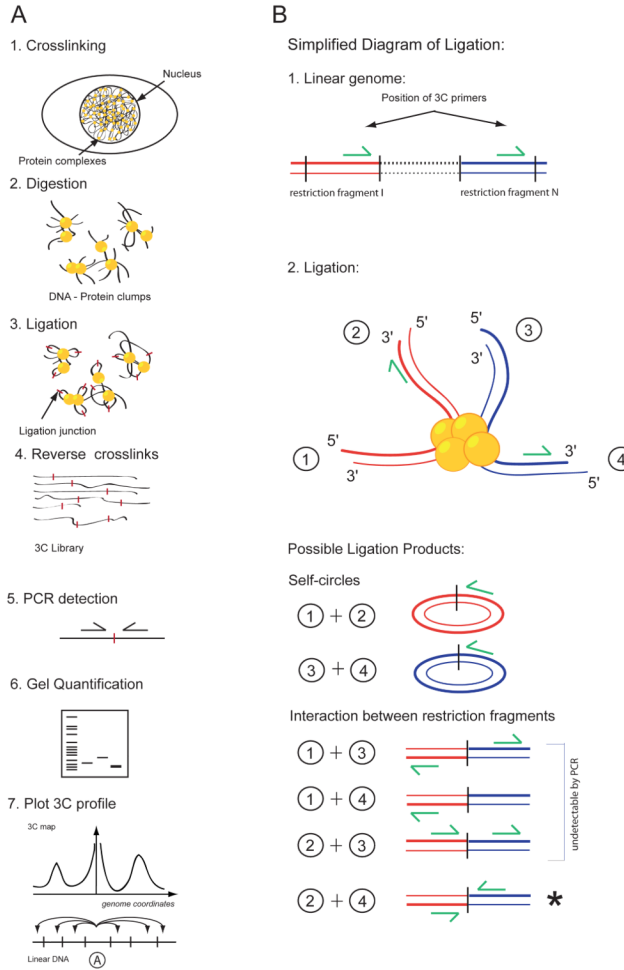


Figure 1. Method overview with details of ligation

A) An illustration of the 3C method. Genomic DNA is crosslinked (1), capturing three-dimensional interactions inside the cell. After cell lysis and removal of cell membranes, the captured chromatin is solubilized and digested (2), isolating protein-DNA complexes from the chromatin network. The free DNA ends are then ligated together (3) in dilute conditions favoring intra-molecular ligation, creating new DNA junctions representing the proximity of restriction fragments in the fixed sample. After ligation, the crosslinks are reversed (4) and 3C template is purified to eliminate cellular debris. Finally, the ligation products are detected (5) using PCR-based methods. After quantification (6) the results are plotted as a 3C profile (7), revealing interactions between anchor (labeled “A”) and all other fragments in the genomic regions, which mirrors 3D spacing in the nucleus.

B) Possible outcome of ligation reaction between two restriction fragments. As seen in part A, there are many restriction fragments contained within one complex. To further understand the ligation step, we have simplified the reality and show a view where one complex contains only two restriction fragments - red and blue. The 5' and 3' ends are indicated for each strand. Each digested end has been numbered 1–4. Also indicated are the locations where 3C primers have been designed. Note that all the primers are on the “forward” strand, located near the restriction site. There are six possible ligation products that result from this molecule. Two of these produce self circles, which are not of interest. Only one of the remaining four ligation products results in a detectable product - that is when end 2 and end 4 are ligated to each other. This ligation event will bring the primers

into the proper orientation to produce a PCR product. None of the other ligation products will be detected.

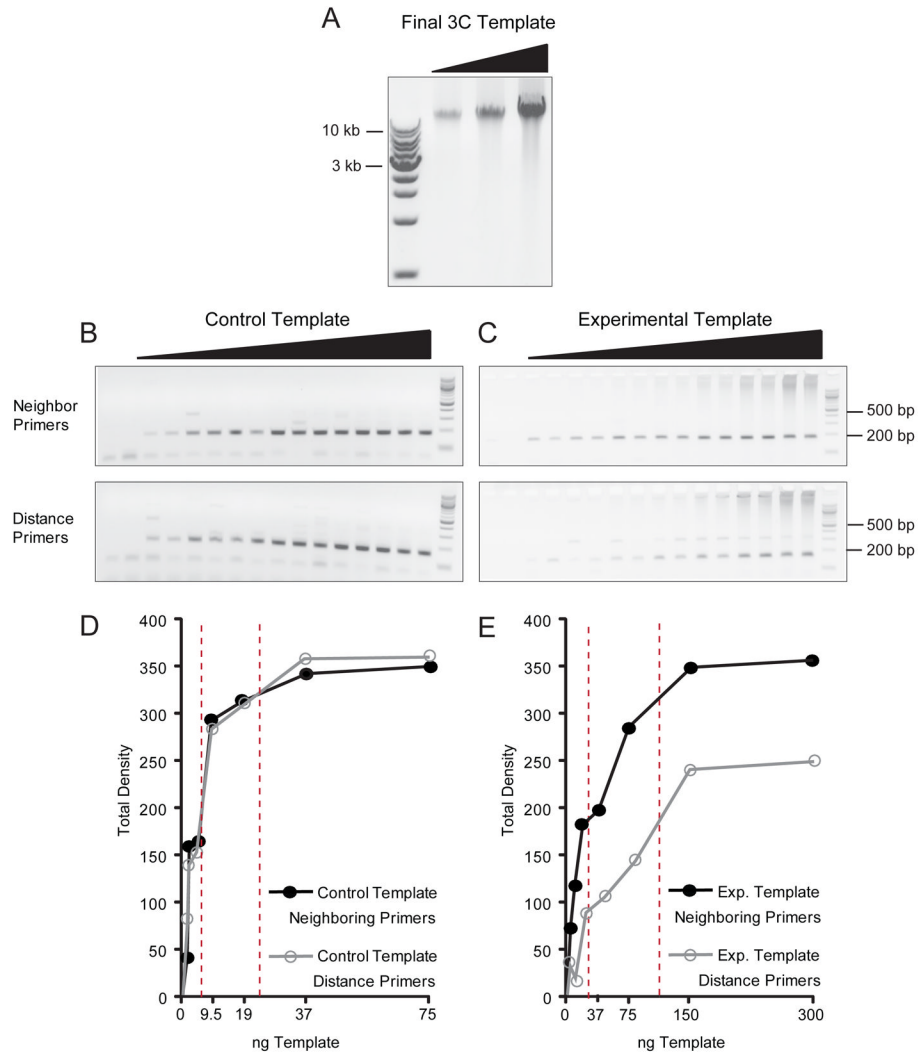


Figure 2. Examples of final 3C library and titrations

A) An example of a completed 3C template run on a 0.8% agarose gel in increasing concentration. The main band of the template should run above 10 kb. There should be a minimal amount of smear underneath the band. If no band is seen, the template cannot be used for downstream analysis. B, C) Titrations of a control template (BAC) and an experimental template (mammalian cell line). Each band is present in duplicate. The upper gel represents ligation products from neighboring restriction fragments. These fragments are close together in both linear and three-dimensional distance, thus giving a strong signal in both the control and experimental templates, even at low template concentration. The lower gel represents ligation products from distal restriction fragments. These fragments are far apart in the linear genome and far apart in three-dimensional distance as well. The control template shows no difference between the neighboring and distal fragments, indicating there is no distance bias. In contrast, the experimental library shows a marked decrease of interaction frequency using the distance primers, which indicates that those fragments are indeed far apart in both linear and three-dimensional space. D, E) Graphs showing the quantification of gel bands from B and C. As described above, there is no difference in gel band intensity in the control template, while there is a difference in the experimental template. The dashed red lines indicate the shoulder of the graphs. When conducting the final 3C experiment, the working concentration in each PCR reaction should be chosen

based on the concentrations in this area of the graph. Using too high or too low of a concentration will skew results.

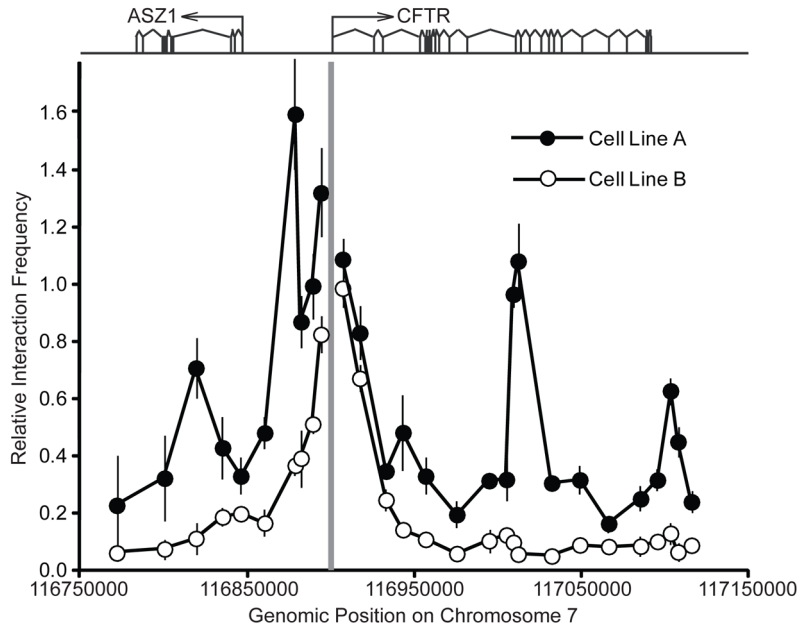


Figure 3. Representation of final 3C data

Figure adapted from Gheldof et al. Figure 1 [23]. Plotted here is sample 3C data from [23] for two cell lines. Above the graph is a schematic of the two genes covered in this data set. Each data point represents the average of three technical replicates and three normalized biological replicates. The cell lines were normalized to each other before they were plotted on the same graph. The data is plotted as relative interaction frequency versus genomic distance. The grey bar represents the anchor fragment. Note that as the genomic distance increases, the interaction frequency of both cell lines generally decreases. In cell line B, for example, this relationship is clear throughout the whole profile. Note four areas of high interaction frequency in cell line A. These four regions stand out as peaks, and represent locations in the genome that are close to the anchor point in three-dimensional space.

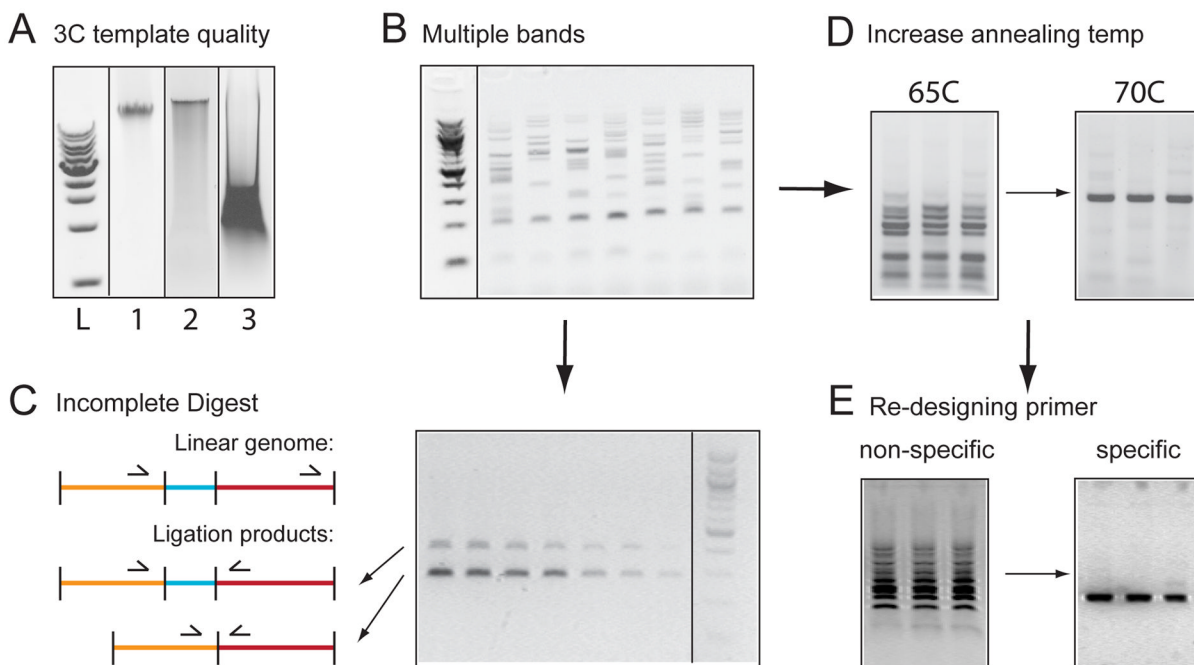


Figure 4. Examples of common 3C problems

A) Quality of 3C template. Typically, an excellent 3C template runs as one tight band above 10kb in size on 0.8% agarose gel (Lane 1). A mediocre 3C template (Lane 2) looks like a tight band with a noticeable smear running down the gel. Usually smeary templates like this are successfully used in 3C experiments. Lane 3 shows a completely degraded template, which should be excluded from downstream analysis. A 1kb ladder is shown in Lane L. (B–E): Multiple bands problem and possible solutions. (B) – Different pairs of 3C primers amplify several bands from a 3C template. There might be several reasons for that, which should be checked one by one. The first possibility is incomplete digestion of the chromatin (C). The figure represents the titration of a 3C template with same pair of primers (see the diagram on the left). The incomplete digestion results in two ligation products, with and without the blue restriction fragment, which are both detected by the same pair of 3C primers. If the size of the blue restriction fragment is relatively small, both ligation products could be amplified in the same PCR reaction. Usually severe incomplete digest, resulting in multiple extra bands of about same intensity as the “main” band (B), is seen in 3C templates made with a fine cutter restriction enzyme (one that recognized a 4-bp cut site). In this case one should consider remaking the 3C library with an increased amount of enzyme and elongated restriction time. The case presented in (C) is legitimate for further analysis. Next, if the size of the additional bands cannot be explained by incomplete digest, one should consider non-specific primer annealing. Note that even when a pair of primers amplifies one band of expected size from a (BAC) control template, the more complex 3C template, made from the whole genome, might have additional sites for primer annealing. Simply increasing the annealing temperature in the PCR reaction can resolve the issue (D). If that’s not the case, the primer pair (or at least the non-anchoring 3C primer) should be re-designed (E).