Video Article

High-throughput Identification of Gene Regulatory Sequences Using Next-generation Sequencing of Circular Chromosome Conformation Capture (4C-seq)

Erin A. Brettmann[1], Inez Y. Oh[1], Cristina de Guzman Strong[1]

[1]Division of Dermatology, Center for Pharmacogenomics, Center for the Study of Itch, Department of Medicine, Washington University School of Medicine

Correspondence to: Cristina de Guzman Strong at CSTRONG@dom.wustl.edu

Abstract

The identification of regulatory elements for a given target gene poses a significant technical challenge owing to the variability in the positioning and effect sizes of regulatory elements to a target gene. Some progress has been made with the bioinformatic prediction of the existence and function of proximal epigenetic modifications associated with activated gene expression using conserved transcription factor binding sites. Chromatin conformation capture studies have revolutionized our ability to discover physical chromatin contacts between sequences and even within an entire genome. Circular chromatin conformation capture coupled with next-generation sequencing (4C-seq), in particular, is designed to discover all possible physical chromatin interactions for a given sequence of interest (viewpoint), such as a target gene or a regulatory enhancer. Current 4C-seq strategies directly sequence from within the viewpoint but require numerous and diverse viewpoints to be simultaneously sequenced to avoid the technical challenges of uniform base calling (imaging) with next generation sequencing platforms. This volume of experiments may not be practical for many laboratories. Here, we report a modified approach to the 4C-seq protocol that incorporates both an additional restriction enzyme digest and qPCR-based amplification steps that are designed to facilitate a greater capture of diverse sequence reads and mitigate the potential for PCR bias, respectively. Our modified 4C method is amenable to the standard molecular biology lab for assessing chromatin architecture.

Video Link

The video component of this article can be found at https://www.jove.com/video/58030/

Introduction

The identification of regulatory elements for gene expression has been facilitated by the Encyclopedia of DNA Elements (ENCODE) Project that comprehensively annotated functional activity for 80% of the human genome[1,2]. The identification of the sites for *in vivo* transcription factor binding, DNaseI hypersensitivity, and epigenetic histone and DNA methylation modifications in individual cell types paved the way for the functional analyses of candidate regulatory elements for target gene expression. Armed with these findings, we are faced with the challenge of determining the functional interconnectivity between regulatory elements and genes. Specifically, what is the relationship between a given target gene and its enhancer(s)? The chromatin conformation capture (3C) method directly addresses this question by identifying physical, and likely functional, interactions between a region of interest and candidate interacting sequences through captured events in fixed chromatin[3]. As our understanding of chromatin interactions has increased, however, it is clear that the investigation of preselected candidate loci is insufficient to provide a complete understanding of gene-enhancer interactions. For example, ENCODE used the high-throughput chromosomal conformation capture carbon copy (5C) method to examine a small portion of the human genome (1%, pilot set of 44 loci) and reported complex interconnectivity of the loci. Genes and enhancers with identified interactions averaged 2–4 different interacting partners, many of which were hundreds of kilobases away in linear space[4]. Further, Li *et al*. used Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET) to analyze whole-genome promoter interactions and found that 65% of RNA polymerase II binding sites were involved in chromatin interactions. Some of these interactions resulted in large, multi-gene complexes spanning hundreds of kilobases of genomic distance and containing, on average, 8-9 genes each[5]. Together, these findings highlight the need for unbiased whole-genome methods for interrogating chromatin interactions. Some of these methods are reviewed in Schmitt *et al*.[6].

More recent methods for chromatin conformation capture studies coupled with next-generation sequencing (Hi-C and 4C-seq) enable the discovery of unknown sequences interacting with a region of interest[6]. Specifically, circular chromosome conformation capture with next-generation sequencing (4C-seq) was developed to identify loci interacting with a sequence of interest in an unbiased manner[7] by sequencing DNA from captured chromatin proximal to the region of interest in 3D space. Briefly, chromatin is fixed to preserve its native protein-DNA interactions, cleaved with a restriction enzyme, and subsequently ligated under dilute conditions to capture biologically relevant "tangles" of interacting loci (**Figure 1**). The cross links are reversed to remove the protein, thus leaving the DNA available for additional cleavage with a

second restriction enzyme. A final ligation generates smaller circles of interacting loci. Primers to the sequence of interest are then used to generate an amplified library of unknown sequences from the circularized fragments, followed by downstream next generation sequencing.

The protocol presented here, which focuses on sample preparation, makes two major alterations to existing 4C-seq methods[8,9,10,11,12]. First, it uses a qPCR-based method to empirically determine the optimal number of amplification cycles for 4C-seq library preparation steps and thus mitigates the potential for PCR bias stemming from over-amplification of libraries. Second, it uses an additional restriction digest step in an effort to reduce the uniformity of known "bait" sequences that hinders accurate base-calling by the sequencing instrument and, hence, maximizes the unique, informative sequence in each read. Other protocols circumvent this issue by pooling many (12-15)[8] 4C-seq libraries with different bait sequences and/or restriction sites, a volume of experiments which may not be achievable by other laboratories. The modifications presented here allow a small number of experiments, samples, and/or replicates to be indexed and pooled into a single lane.

Protocol

1. Restriction Enzyme Selection

1. Identify a region of interest (*e.g.*, gene promoter, single nucleotide polymorphism (SNP), enhancer) and obtain the DNA sequence from repositories such as National Center for Biotechnology Information (NCBI).
2. Identify the candidate restriction enzymes (REs) for the first restriction enzyme digestion (RE1) that do not cut within the sequence of interest, which produce sticky ends (DNA termini with overhangs) after RE digestion, and whose activities are not inhibited by CpG methylation.
   Note: The resolution of the data is determined by RE1. An enzyme with a 6 base pair (bp) recognition sequence produces, on average, fragments approximately 4 kilobases (kb) in length, while an enzyme with a 4 bp recognition sequence produces fragments approximately 250 bp in length, allowing more precise identification of interacting sequences.
3. Select a RE1 from the candidate enzymes identified in Step 1.2 that produces a "viewpoint" restriction fragment at least 500 bp long that includes the region of interest or, alternatively, a neighboring region.
   NOTE: The enzyme selected must be amenable to primer design (see Step 2).
4. For the second digestion, identify a restriction enzyme (RE2) with a 4 bp recognition sequence that produces sticky ends (DNA termini with overhangs) after RE digestion, whose activity is not inhibited by CpG methylation, and that cuts within the restriction fragment generated by RE1 to produce a 250–500 bp bait fragment (**Figure 1**).
   NOTE: The enzyme selected must be amenable to primer design (see Step 2).

2. Design and Test Primers for Inverse PCR and Digestion Efficiency qPCR

1. **Use a primer design tool such as Primer3 (http://primer3.ut.ee) to design inverse primers that are directed outwards towards the ends of the bait fragment (*i.e.*, the 5' primer is a "reverse" primer, complementary to the plus strand, while the 3' primer is a "forward" primer, complementary to the minus strand) and as close to the restriction sites as possible to minimize the amplification of non-informative bait sequence and maximize PCR efficiency (Figure 2A).**
   NOTE: Primers should be predicted to have minimal non-specific amplification from genomic DNA by *in silico* PCR predictions and should not align elsewhere in the genome except their intended target with more than 16/18 identity[9]. As sequencing adapters are not incorporated into the inverse PCR primers, the site of primer binding is more flexible than in other 4C preparation methods; primers should optimally anneal within 50 bp of the end of the corresponding restriction site.
   1. Confirm the specificity of inverse PCR primers by amplification using purified genomic DNA (gDNA) as template.
   2. Optimal primers should yield few products when using gDNA as a template (see Step 11.2 for the description of expected 4C products). If substantial amplification from gDNA occurs, design new primers. If no acceptable primer sets are identified, return to Step 1 and select a new bait by selecting new restriction enzymes.
2. **Design qPCR primers to amplify the fragments 70–200 nucleotide (nt) in length to monitor the restriction digestion efficiency (Figure 2B).**
   1. Design one pair of primers to amplify across each restriction site (selected in Step 1) that define the bait sequence. Design the primers for both RE1 (see Step 6.5.6) and RE2 sites (see Step 9.2).
   2. Design a set of primers that amplifies a region not containing the sites for either restriction enzyme as a normalization control (uncut DNA) for RE1 and RE2 (see also Step 6.5.6).

3. Collection of Cells

1. Using a method that is appropriate for the tissue or cell culture, obtain a single-cell suspension. Pellet the cells for 5 min at 200 x *g*, discard the supernatant and resuspend the pellet in 500 μL of 1x phosphate buffered saline (PBS) per 10^7 cells.

4. Formaldehyde Cross-linking of Cells to Preserve Chromatin Interactions

1. Per 10^7 cells, add 9.5 mL of 1% electron microscopy (EM)-grade formaldehyde (methanol-free) in PBS and incubate, while tumbling on rocker (or similar), for 10 min at room temperature (RT, 18–22 °C).
   NOTE: Fixation conditions may have to be optimized for each cell type. Prior published work in mouse cells reported that optimal fixation results in at least 40% of mapped reads aligning to the chromosome containing the viewpoint[9] assuming a non-telomeric region for an average sized chromosome.
2. Transfer the reaction tubes to ice and add ice-cold 1 M glycine to a final concentration of 0.125 M (1.425 mL) to quench the crosslinking reaction. Mix by gentle inversion.

3. Centrifuge for 5 min at 200 x *g*, 4 °C and carefully remove all the supernatant. Store the cell pellet at -80 °C or proceed immediately to cell lysis.

5. Cell Lysis

1. Resuspend the cell pellet from Step 4.3 in 500 µL of 5 mM ethylenediaminetetraacetic acid (EDTA) to wash. Centrifuge at 200 x *g* for 5 min at RT. Discard the supernatant.
2. Resuspend the cell pellet in 125 µL of 5 mM EDTA with 0.5% sodium dodecyl sulfate (SDS) and 1x protease inhibitors, such as a 100x cocktail containing 5 mg/mL antipain, 10 mg/mL chymostatin, 10 mg/mL leupeptin, and 10 mg/mL pepstatin A.
   NOTE: Detergent concentration may have to be optimized for each cell type. Insufficient detergent concentrations will result in incomplete cell lysis, leaving chromatin inaccessible to restriction enzymes. If restriction digestion is persistently low in Step 6.5.6 and the activity of the enzyme has been confirmed, additional detergent may be required. Increase SDS concentration in 0.1% increments.
3. Incubate the cell suspension on ice for 10 min.
   NOTE: For primary human keratinocytes, optimal lysis was achieved using a 10 min incubation at 65 °C followed by a 37 °C overnight incubation in a shaking heating block with agitation (900 rpm).
4. **Ensure that cell lysis is complete.**
   1. Mix 6 µL of the cells with 6 µL of Trypan Blue on a slide and cover with a coverslip. View under a microscope.
      Note: Upon successful lysis, the interior of the cells should appear blue and un-lysed cells will appear white.
   2. If the cell lysis appears insufficient, pellet the cells at 200 x *g* for 5 min and save the supernatant. Resuspend the pellet and repeat Steps 5.2–5.4 and/or alternatively with different incubation temperatures (see the Note in Step 5.3).
   3. Combine the re-lysed pellet with the saved supernatant and proceed.
      Note: Visual inspection is not a guarantee of sufficient lysis. Digestion efficiency should be objectively determined as in Step 6.5.6.

6. First Restriction Digestion

1. Add 30 µL of 10x restriction enzyme buffer (specified by the manufacturer) and 27 µL of 20% Triton X-100 (final 1.8%) to the suspension from Step 5.4. Bring the total volume to 300 µL with H_2O.
   NOTE: The composition of the restriction enzyme buffer will depend on the RE chosen in Step 1.
2. Remove a 15 µL aliquot as an "Undigested control" and store at 4 °C.
3. Add 200 U RE1 to the remaining reaction mixture and incubate overnight at the temperature appropriate to the enzyme in a shaking heating block with agitation (900 rpm). The next day, add an additional 200 U RE1 and continue the incubation overnight.
4. Remove a 15 µL aliquot as a "Digested control" and store at 4 °C.
5. **Determine digestion efficiency:**
   1. Add 82.5 µL of 10 mM Tris-HCl pH 7.5 to the 15 µL samples from Steps 6.2 and 6.4. Add 2.5 µL of proteinase K (20 mg/mL) and incubate for 1 h at 65 °C.
   2. Add 100 µL of phenol-chloroform and mix vigorously by inversion to remove residual protein contamination. Centrifuge for 5 min, 16,100 x *g* at room temperature.
   3. Transfer the aqueous phase to a new tube. Add 6.66 µL of 3 M sodium acetate pH 5.2, 1 µL of 20 mg/mL glycogen, and 300 µL of 100% ethanol (EtOH). Mix gently by inversion and place at -80 °C for 1 h.
   4. Centrifuge for 20 min at 16,100 x *g* at 4 °C. Remove the supernatant, add 500 µL of 70% EtOH, and centrifuge for 5 min at 16,100 x *g* at RT.
   5. Remove the supernatant and dry the pellet at room temperature for 2 min. Resuspend the pellet in 50 µL of nuclease-free H_2O.
   6. Determine digest efficiency by qPCR[13,14] using the ΔΔCt method. Use the primer set not flanking a restriction site (see Step 2.2.2) as the normalization control. Proceed if the digestion efficiency is >85%. Otherwise, pellet the cells and repeat Steps 5.2–6.5, omitting the incubation at 65 °C in Step 5.3.

7. First Ligation

1. Heat-inactivate the restriction enzyme by incubating 20 min at 65 °C. Alternatively, if the enzyme cannot be heat inactivated, phenol-chloroform extract and ethanol precipitate the sample.
   NOTE: The higher inactivation temperatures recommended for some restriction enzymes denature the proteins in the chromatin, and this may negatively affect the sample quality. Additionally, phenol:chloroform extraction is not ideal, as it results in the loss of sample.
2. Transfer to a 50 mL conical tube and add 6 mL of nuclease-free H_2O, 700 µL of 10x ligase buffer (660 mM Tris-HCl pH 7.5, 50 mM magnesium chloride ($MgCl_2$), 10 mM dithiothreitol (DTT), 10 mM adenosine triphosphate (ATP)), and 50 U T4 DNA Ligase. Mix gently by swirling and incubate overnight at 16 °C.
3. Remove a 100 µL aliquot of the sample as the "Ligation control".
4. **Determine ligation efficiency:**
   1. Add 2.5 µL of Proteinase K (20mg/mL) and incubate 1 h at 65 °C.
   2. Add 100 µL of phenol-chloroform and mix vigorously by inversion. Centrifuge for 5 min, 16,100 x *g* at RT.
   3. Transfer the aqueous phase to a new tube. Add 6.66 µL of 3 M sodium acetate pH 5.2, 1 µL of glycogen, and 300 µL of 100% EtOH. Mix gently by inversion and place at -80 °C about 1 h.
   4. Centrifuge for 20 min at 16,100 x *g* at 4 °C. Remove the supernatant, add 500 µL of 70% EtOH, and centrifuge for 5 min at 16,100 x *g* at 4 °C.
   5. Remove the supernatant and dry the pellet at room temperature. Resuspend the pellet in 20 µL of water and load on a 0.6% agarose gel next to the "digestion controls" from Step 6.5.
      NOTE: A well-ligated sample should appear as a relatively tight, high molecular weight band (**Figure 3**).

6. If the ligation is sufficient, proceed with Step 8. Otherwise, add fresh ATP (1 mM final concentration) and new ligase; incubate overnight at 16 °C and repeat Steps 7.3–7.4.

8. Reverse Cross-linking and Isolate Chromatin

1. Add 15 µL of Proteinase K (20 mg/mL) and incubate overnight at 65 °C to reverse cross-links.
2. Add 30 µL of RNase A (10 mg/mL) and incubate 45 min at 37 °C.
3. Add 7 mL of phenol-chloroform and mix vigorously by inversion. Centrifuge 15 min, 3,300 x g at RT.
4. Transfer the aqueous phase to a new 50 mL tube and add 7.5 mL of nuclease-free H_2O (to dilute the DTT present in the ligase buffer, which otherwise precipitates with the DNA), 1 mL of 3 M sodium acetate pH 5.6, 7 µL of glycogen (20 mg/mL), and 35 mL of 100% EtOH. Mix and incubate at -80 °C for 1 h.
5. Centrifuge 20 min, 3,900 x g at 4 °C. Remove the supernatant (pellet may be difficult to see), wash the pellet with 10 mL of ice-cold 70% EtOH, and centrifuge 15 min, 3,300 x g at 4 °C.
6. Remove the supernatant and briefly dry the pellet at RT. Dissolve the pellet in 150 µL of 10 mM Tris-HCl pH 7.5 at 37 °C. Store at -20 °C or continue with Step 9.

9. Second Restriction Digestion: Trimming the Circles

NOTE: This step creates smaller circles to minimize the overrepresentation of smaller captured fragments due to PCR bias in downstream amplification steps.

1. To the sample from Step 8.6, add 50 µL of 10x restriction enzyme buffer (specified by the manufacturer), 300 µL of nuclease-free H_2O, and 50 U restriction enzyme RE2. Incubate overnight at the temperature appropriate for the chosen enzyme.
2. Remove a 15 µL aliquot of the sample as the "Digestion control". Determine digestion efficiency as described in Step 6.5.

10. Second Ligation and DNA Purification

1. Inactivate the restriction enzyme as recommended by the manufacturer. If the enzyme cannot be heat-inactivated, remove the enzyme using a column-based purification kit.
    NOTE: As this results in the loss of sample, column purification is not ideal.
2. Transfer the sample to a 50 mL tube and add 12.1 mL of nuclease-free H_2O, 1.4 mL of 10x ligation buffer (660 mM Tris-HCl, pH 7.5; 50 mM $MgCl_2$; 10 mM DTT; 10 mM ATP), and 100 U T4 DNA ligase. Incubate overnight at 16 °C.
3. Add 467 µL of 3 M sodium acetate pH 5.6, 233 µL nuclease-free H_2O, 7 µL glycogen (20 mg/mL), and 35 mL 100% EtOH. Mix well and incubate at -80 °C 1 h.
4. Centrifuge 45 min, 3,900 x g at 4 °C. Remove the supernatant, add 10 mL of cold 70% EtOH, and centrifuge 15 min, 3,300 x g at 4 °C.
5. Remove the supernatant and briefly dry the pellet at room temperature. Add 150 µL of 10 mM Tris-HCl pH 7.5 and incubate at 37 °C to dissolve the pellet.
6. Purify the samples with a silica column-based PCR purification kit, following the manufacturer's instructions. Use 1 column per 3 x 10^6 cells, based on the initial number of the cells. Elute the columns with 50 µL of 10 mM Tris-HCl, pH 7.5 and pool the samples.
7. Measure the concentration by fluorimetry or spectrophotometry using the absorbance at 260 nm. The 4C template is now ready for inverse PCR. Store at -20 °C or proceed directly to Step 11.

11. PCR Amplify Unknown Interacting Sequences by Inverse PCR

1. Determine the linear range of amplification by performing a PCR using template dilutions of 12.5, 25, 50, and 100 ng 4C template. If desired, amplify from gDNA in parallel to directly compare products in order to identify non-specific amplification. Run PCR reactions using an initial denaturation at 94 °C for 2 min; 30 cycles consisting of a denaturation step at 94 °C for 10 s, an annealing step at 55 °C for 1 min, and an extension step at 68 °C for 3 min; and a final extension at 68 °C for 5 min.
2. Separate 15 µL of each PCR product on a 1.5% agarose gel to confirm linear amplification and assess template quality (**Figure 4**). Amplification from 4C template should yield discrete banding at low DNA concentrations and a smear at higher concentrations. The presence of the smear indicates increased complexity of the amplified 4C ligations.
3. **When satisfied with the quality and quantity of the inverse PCR product generated, set up a qPCR to determine the optimal number of cycles to use for amplification:**
    1. Set up the reactions containing SYBR and ROX dyes using the reaction mixture in **Table 2**. Unless the amplification is not linear at high concentrations in Step 11.2, use 100 ng of the template per reaction.
        NOTE: The addition of ROX facilitates the normalization of fluorescent signal from well to well and cycle to cycle.
    2. Run the PCR reactions using an initial denaturation at 94 °C for 2 min; 40 cycles consisting of a denaturation step at 94 °C for 10 s, an annealing step at 55 °C for 1 min, and an extension step at 68 °C for 3 min; and a final extension at 68 °C for 5 min.
    3. Determine the peak (endpoint) fluorescence of the reactions using the amplification plot. Determine the cycle at which reactions reach 25% of peak fluorescence (**Figure 5**).
        NOTE: This is the number of cycles that will be used to amplify the 4C libraries (see Step 11.4).
4. Set up inverse PCR as in **Table 3** to amplify unknown sequences ligated to the bait from the 4C template. Divide into 16 reactions of 50 µL before running. Run PCR reactions using an initial denaturation at 94 °C for 2 min and cycles consisting of a denaturation step at 94 °C for 10 s, an annealing step at 55 °C for 1 min, and an extension step at 68 °C for 3 min. Use the number of cycles determined in Step 11.3.3.
5. Collect and pool the reactions. Purify using a silica column-based PCR purification kit. Use at least 2 columns per 16 reactions. Pool the purified PCR products.

6. Determine the sample quantity and purity by spectrophotometry. Typical yields are between 10 and 20 μg with A260/A280 ~1.85. If the absorption ratios are sub-optimal, re-purify in order to prevent the problems during sequencing.
7. Assess the complexity of library by separating 300 ng of purified PCR product on a 1.5% agarose gel.
   NOTE: Amplified product should resemble that from Step 11.2.

12. Third Restriction Digestion: Trim off Bait Sequences

NOTE: This step removes non-informative bait sequences from the inverse PCR products to maximize informative captured sequences in the downstream sequencing steps. To monitor digest efficiency, a "digest monitor"[15] is digested in parallel using equivalent DNA and enzyme concentration. If RE1 and RE2 are incompatible for simultaneous digestion, for example, due to different optimal incubation temperatures or reaction buffers, this must be done as a sequential digest (this is not ideal).

1. Obtain a RE digest monitor and test the digestion in a 50 μL reaction.
   NOTE: This is a molecule of dsDNA (for example, a plasmid, PCR amplicon, or synthetic DNA) that contains the RE site(s). The only requirement is that it should be easy to distinguish cut from uncut monitor on an agarose gel. Optimize RE monitor mass and enzyme concentration if needed.
2. Digest 1 μg of purified inverse PCR product from Step 11.6 and, in parallel, the RE monitor from Step 12.1.
   NOTE: DNA and enzyme concentration, as well as incubation time, should be the same as for the test digest in Step 12.1 for both the inverse PCR product and the monitor digests. Adjust the reaction volume as needed.
3. Run the digested RE monitor on an agarose gel of concentration appropriate for the expected fragments. Digestion is considered sufficient when <10% of the DNA remains uncut (**Figure 6**).
4. Purify the digested inverse PCR product on a silica column-based DNA purification kit. If sequential digests are required, repeat Steps 12.1–12.4 with the second enzyme.

13. Preparation of Sequencing Library

1. **Ligate the adapters compatible with the respective next generation sequencing platform.**
   1. Design the adapters for each RE1 and RE2 so that, after annealing the oligos, each RE adaptor has the respective 1-sided overhang.
      NOTE: For example, if HindIII is used, the annealed adapter should contain a 5' phosphorylated "AGCT" overhang. Alternatively, if standard T-overhang adapters are used, end-repair and A-tail the library prior to adapter ligation.
   2. Anneal the respective adaptor RE oligos. Resuspend the oligos in annealing buffer (10 mM Tris, pH 7.5, 50 mM sodium chloride (NaCl), 1 mM EDTA), mix in equimolar quantities to 50 μM, heat to 95 °C, and allow to cool slowly to 25 °C.
   3. Perform the ligation reaction. Mix RE-digested 4C library with a total 5- to 10-fold molar excess of annealed adaptors (Step 13.1.2; 50:50 ratio for each RE adaptor used) in 1x ligase buffer and add 6U DNA ligase. Incubate according to manufacturer instructions.
   4. Remove excess adapters by purifying using a low-elution volume silica-based column kit.

2. **Size-select.**
   NOTE: Size selection can also be performed using a bead-based cleanup kit and is recommended even after column purification.
   1. Cast a 2% agarose gel using a high-resolution agarose, adding a non-UV stain prior to pouring. Ensure that any water lost due to evaporation during heating is replaced by weighing the bottle or flask before and after heating and adding water to recover the lost mass.
   2. Run the adapter-ligated libraries on the gel, leaving empty lanes between the samples.
   3. Excise the gel slices corresponding to the size range 150 bp to 1 kb using a clean scalpel or razor blade.
   4. Purify the libraries from the gel using a gel extraction kit. To minimize GC bias in the recovery of DNA, dissolve the gel slices at room temperature with rotation.

3. Determine the number of cycles for PCR amplification by qPCR using the reaction mixture in **Table 4**. Run PCR reactions using an initial denaturation at 98 °C for 2 min and 40 cycles consisting of a denaturation step at 98 °C for 10 s, an annealing step at 60 °C for 1 min, and an extension step at 72 °C for 3 min.
   NOTE: The cycle at which fluorescence reaches 25% of the maximum is the number of cycles that will be used to amplify the library, as in Step 11.3.
4. Amplify the libraries using the reaction mixture in **Table 5**. Run PCR reactions using an initial denaturation at 98 °C for 2 min and cycles consisting of a denaturation step at 98 °C for 10 s, an annealing step at 60 °C for 1 min, and an extension step at 72 °C for 3 min. The number of amplification cycles to be used for each library was determined in Step 13.3.
5. Purify amplified libraries using a low-elution volume silica-based column kit.

14. Sequencing and Analysis of Sequencing Data

1. Determine the concentrations of amplified libraries using a fluorimetric assay. Dilute the libraries to the appropriate concentration for sequencing (check with the sequencing service or core for their recommendation).
2. Determine the quality of the libraries using a microfluidic nucleic acid analysis platform.
   NOTE: The size profile of the library should mirror that seen on the size selection gel in Step 13.2, and the concentration of the sample determined in this step is to be used for sequencing.
3. Pool indexed libraries to obtain at least 3 million reads (at least 50 bp read; single [1X50] or paired end [2X50]) per sample. A high-quality library requires at least 1 million mapped reads[9], but there will be some attrition during mapping. For example, a sequencing platform that yields approximately 150 million reads per lane can accommodate up to 50 samples in a single lane.

15. Analysis of Sequence Data

1. Demultiplex the data by using index sequences to assign the reads to their appropriate samples.
   NOTE: Depending on their familiarity, users may perform downstream computational analyses of raw FASTA/FASTQ sequence files using basic command line interface or, alternatively, the user-friendly, web-based Galaxy graphical user interface (GUI)[16].
2. Trim adapter sequences and any bait sequence arising from incomplete RE digestion in Step 12, for example, using the FAST-X Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/).
3. Map demultiplexed reads to the appropriate genome of interest (*e.g.*, UCSC mm10, hg19) using the Burrows-Wheeler Aligner (BWA)[17] or other alignment software resulting in SAM output files. If needed, convert SAM files to BAM via samtools[18] (see Step 15.4).
4. Use a 4C-seq analysis pipeline such as Basic4Cseq[19], 4C-ker[20], FourCSeq[21], or fourSig[22] to analyze the data and identify the interactions.
   NOTE: Data analysis and the assessment of the quality thereof should be conducted according to the selected software package's reference manual. Packages generate BED output files except where noted. Briefly, Basic4CSeq[19] uses an input SAM file (Step 15.3) to visualize the interactions (txt, tiff, BED, and wig output files) and assesses the quality of data based on the criteria set forth in van de Werken *et al*.[9] 4C-ker[20] (input trimmed FASTQ, step 15.2) and FourCSeq[21] (input BAM, Step 15.3) identify differential interactions between conditions. fourSig[22] (input SAM) also identifies significant interactions and prioritizes those that are likely to be reproducible. Tools such as the Genomic Regulatory Enrichment of Annotations Tool (GREAT)[23], or integration with ENCODE data sets may also be used to predict the biological function of interactions discovered by 4C-seq.

Representative Results

Primary human keratinocytes were isolated from 2–3 discarded neonatal foreskins, pooled, and cultured in KSFM supplemented with 30 µg/mL bovine pituitary extract, 0.26 ng/mL recombinant human epidermal growth factor, and 0.09 mM calcium chloride ($CaCl_2$) at 37 °C, 5% carbon dioxide. The cells were split into two flasks, and one flask was differentiated by the addition of $CaCl_2$ to a final concentration of 1.2 mM for 72 h. 10^7 cells each from proliferating and differentiated keratinocyte populations were fixed in 1% formaldehyde for 10 min at room temperature. Separately, K562 cells were grown in RPMI supplemented with 10% fetal bovine serum at 37 °C, 5% carbon dioxide. 10^7 cells were fixed in 1% formaldehyde for 10 min at room temperature.

The cells were lysed, and the fixed chromatin was digested with HindIII and ligated as described above. Cross-links were reversed, and the DNA is purified and digested with CviQI. DNA was ligated and used as the template for inverse PCR amplification. Primers used were: 5'-GATCAGGAGGGACTGGAACTTG/5'- CCTCCCTTCACATCTTAGAATG. 1 µg of purified inverse PCR product was digested overnight with CviQI in parallel with 225 ng of RE digest monitor and column-purified. Purified DNA was then digested overnight with HindIII in parallel with 125 ng of RE digest monitor and column-purified. 250 ng of purified double-digested inverse PCR product was used for library preparation. Briefly, the ends were repaired via conversions to 5'-phosphorylated and blunt-ends and subsequent DNA column-purification. Ends were A-tailed 20 min at 72 °C in a 50 µL reaction with 1 U Taq polymerase and 200 µM dATP and column-purified. Compatible sequencing library prep adapters were ligated at a 10:1 (adapter:DNA) ratio using T4 DNA ligase in a 30 µL reaction at room temperature for 15 min and the DNA column-purified. Libraries were run on a 2% agarose gel, gel slices were cut from 120 bp to the top of the ladder to remove adapters, and libraries were purified. 200 pg of each library were evaluated in 10 µL qPCR reactions containing indexing primers to determine optimal PCR amplification cycles. Libraries were amplified to add indices using the number of cycles determined by qPCR and sequenced on a HiSeq2500 to obtain 1x50 reads.

Reads were demultiplexed and trimmed. First, sequencing adapters were removed. Subsequently, the sequence from the beginning of each primer to the restriction site was trimmed. This enables the mapping of inverse PCR products that were not fully digested prior to library preparation. Importantly, including sequence between the primer binding site and the restriction site prevents mapping of non-specifically-amplified PCR product. Trimmed reads were mapped to hg38 using BWA. **Figure 7** shows reads mapping to the region surrounding the viewpoint.
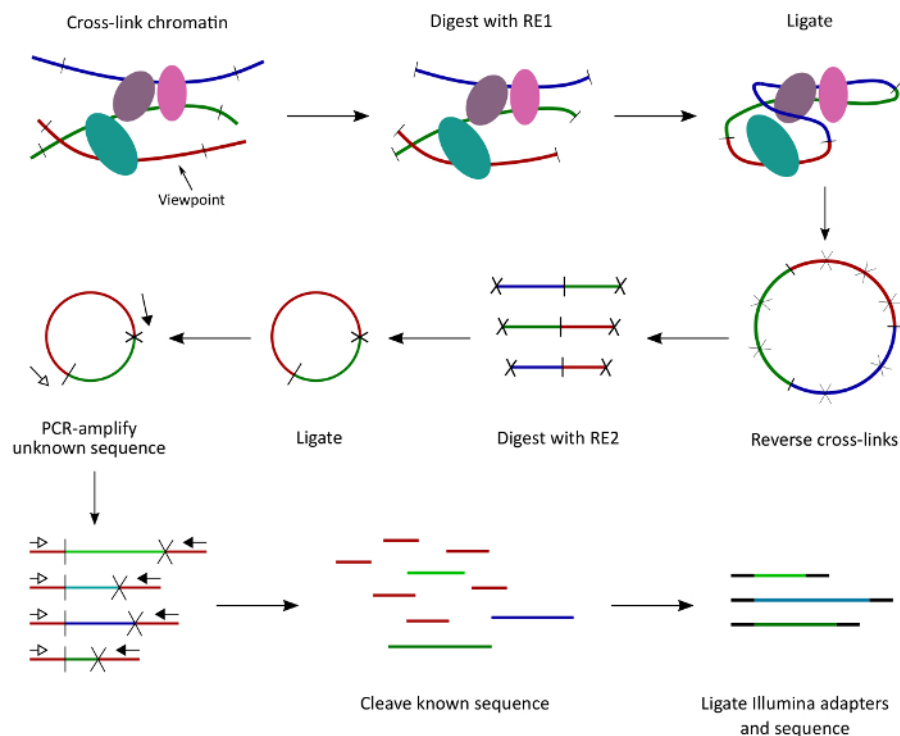
Figure 1: Schematic of 4C-seq workflow. Chromatin is cross-linked to preserve protein-DNA contacts, digested with RE1, and ligated to link interacting loci. Cross-links are reversed, and DNA is digested with RE2 and ligated. Unknown interacting sequences are amplified using primers that bind to the region of interest, and PCR products are digested with RE1 and RE2 to trim known sequences. Amplified DNA of unknown sequence is used in a sequencing library prep, in which the adapters are ligated and the library is PCR-amplified and sequenced. Please click here to view a larger version of this figure.
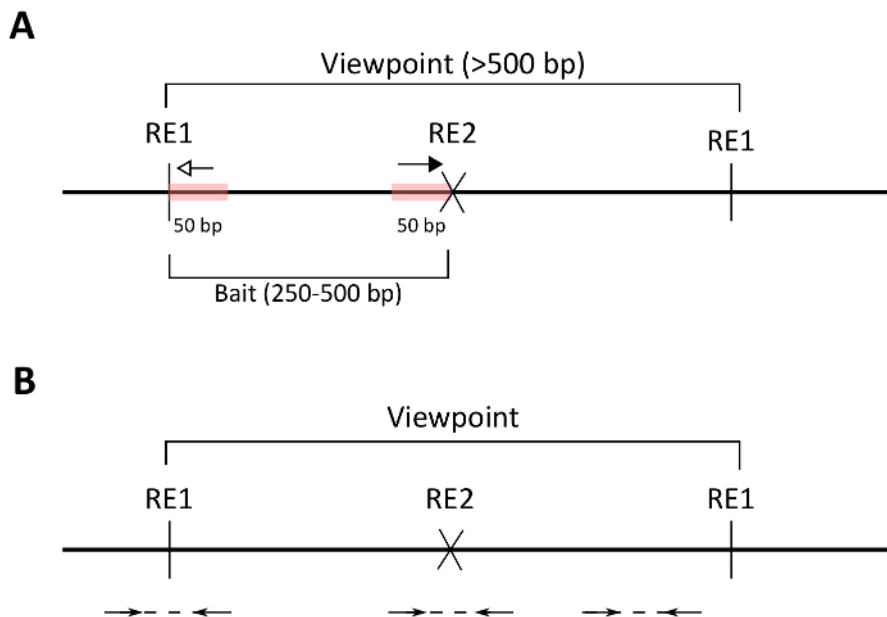

Figure 2: Schematics of primer designs. (A) Binding sites for inverse PCR primers. Primers are oriented "outward" (*i.e.*, the 5' primer is a "reverse" primer, complementary to the plus strand, while the 3' primer is a "forward" primer, complementary to the minus strand) and bind within 50 bp of each restriction site (indicated by red shading). **(B)** Binding sites for qPCR primers for determining restriction enzyme digest efficiency. One primer pair flanks each restriction enzyme site. A control primer set amplifies a sequence that does not contain a site for either enzyme and is used to normalize Ct values for DNA input. Please click here to view a larger version of this figure.
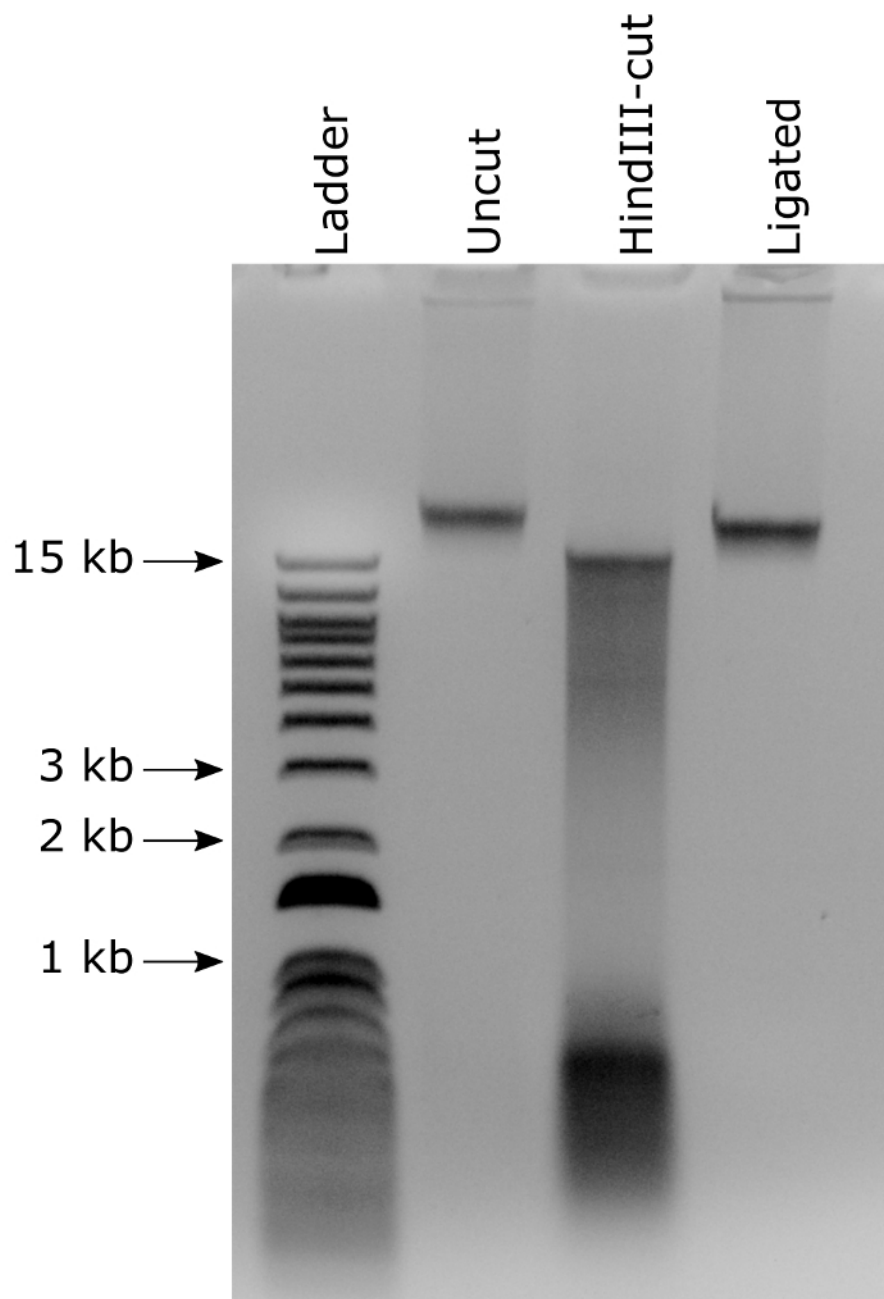
Figure 3: Agarose gel electrophoresis to assess the ligation efficiency for RE1-digested chromatin. Uncut, HindIII-digested, and ligated samples were treated with proteinase K and heated to reverse cross-links, phenol:chloroform extracted, EtOH precipitated, and resuspended in H_2O. Purified DNA was run on a 0.6% agarose gel and bands visualized by ethidium bromide staining with comparison to 1 kb plus ladder. Please click here to view a larger version of this figure.
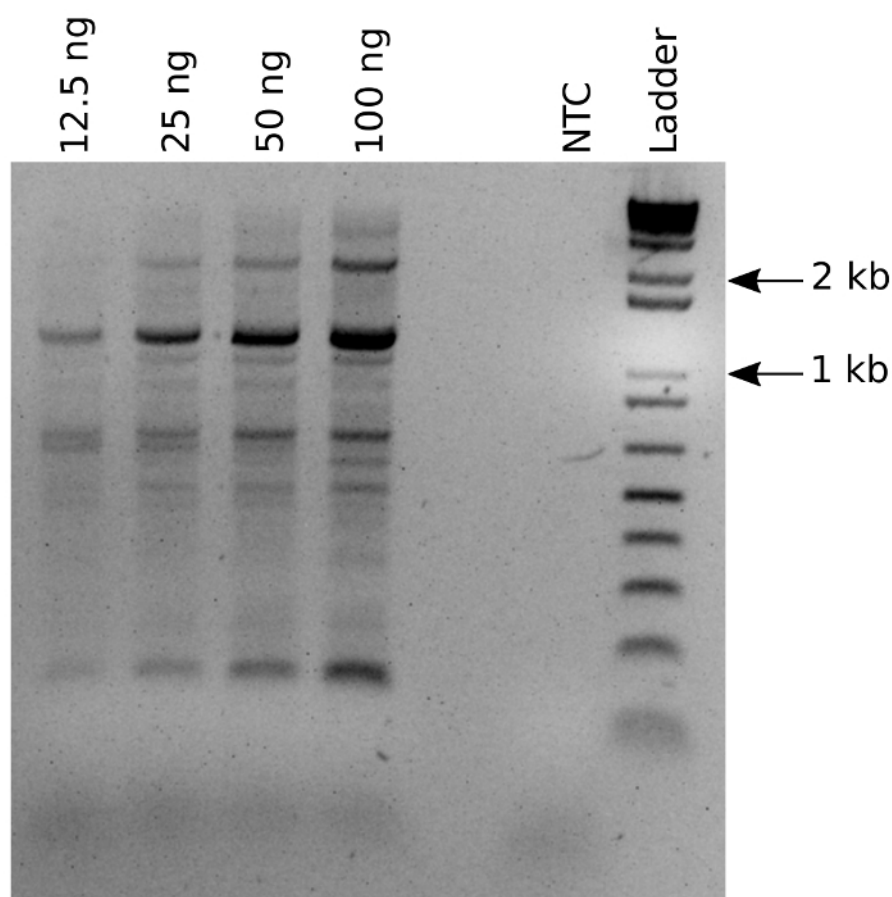
Figure 4: 4C template titration for inverse PCR. Serial dilutions were made of 4C templates and used in inverse PCR reactions. PCR products were run on 1.5% agarose gel and visualized by ethidium bromide staining alongside 1 kb plus ladder. NTC denotes no-template control reaction. Note a correlative increase in amplification with an increase in template concentration. This representative amplification was conducted on chromatin cleaved with HindIII as RE1 and CviQI as RE2, and primer sequences used for inverse PCR amplification were 5'-GATCAGGAGGGACTGGAACTTG/5'- CCTCCCTTCACATCTTAGAATG. Please click here to view a larger version of this figure.
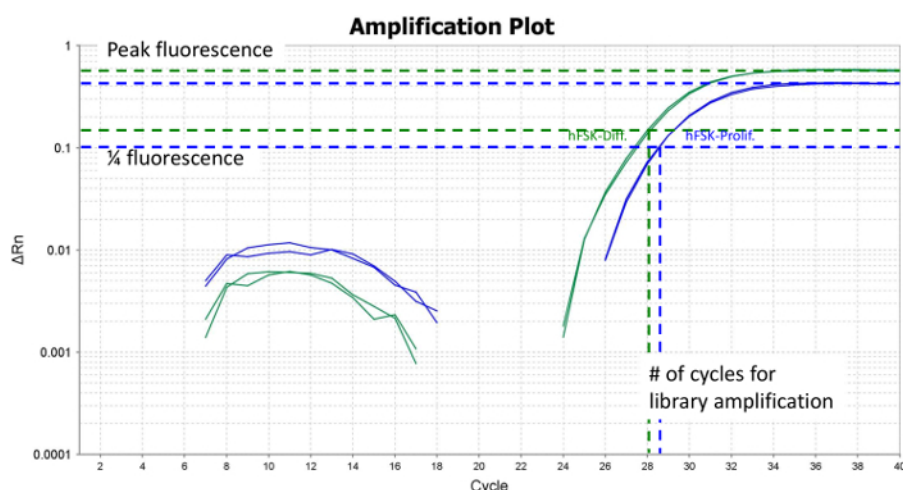

Figure 5: qPCR-mediated determination of amplification cycles needed to amplify template for inverse PCR. 4C templates were amplified in reactions containing 1x SYBR Green and 1x ROX in a real-time thermocycler. Peak fluorescence was determined and the number of cycles required to reach ¼ of peak fluorescence was calculated. This number of cycles was used to amplify the 4C template for inverse PCR. Please click here to view a larger version of this figure.
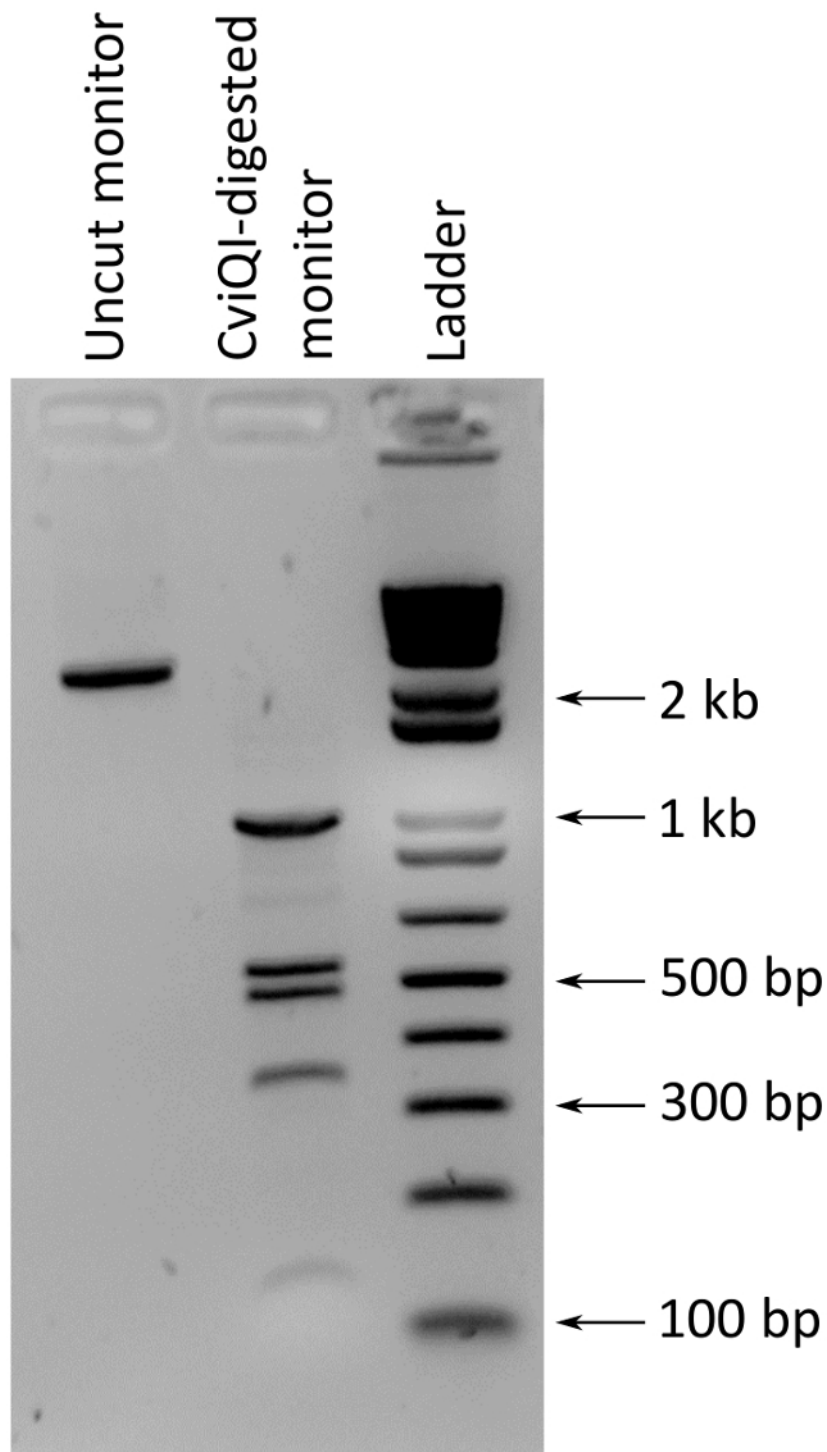
Figure 6: Agarose gel electrophoresis of digested RE monitor indicating sufficient digestion. RE digest monitor was amplified from human genomic DNA using the primers F: 5'-TCCTATCCCTGGTCTGTCTTAT and R: 5'-CCACATTGGTCCTTCTAGTCTTC and purified. 225 ng of monitor was digested with 15 U CviQI in a 50 μL reaction at 25 °C overnight and run on a 1.5% agarose gel alongside 1 kb plus ladder. Bands were visualized with ethidium bromide staining. Uncut monitor is 2515 bp; expected fragment sizes are 132, 343, 488, 539, and 1013 bp. Please click here to view a larger version of this figure.

Figure 7: Representative genomic track of read coverage within 5 kb of the viewpoint region. Reads were trimmed and mapped to hg38 using BWA. The majority of reads (blue peaks) align at HindIII or CviQI sites adjacent to HindIII sites, as expected. The viewpoint region is highlighted in red. Please click here to view a larger version of this figure.

	Volume for 1x
10x Expand Long Template Buffer 1	2.5 µL
dNTPs (10 mM)	0.5 µL
Forward primer	35 pmol
Reverse primer	35 pmol
Expand Long Template Polymerase (5 U/µL)	0.35 µL
DNA	
Nuclease-free water	to 25 µL

Table 1: PCR reaction mixture for the amplification of 4C template (Step 11.1).

	Volume for 1x
10x Expand Long Template Buffer 1	1.5 µL
dNTPs (10 mM)	0.3 µL
Forward primer	21 pmol
Reverse primer	21 pmol
100x SYBR Green I	0.15 µL
50x ROX	0.3 µL
Expand Long Template Polymerase (5 U/µL)	0.21 µL
DNA	100 ng
Nuclease-free water	to 25 µL

Table 2: qPCR reaction mixture for the determination of the number of amplification cycles for inverse PCR (Step 11.3.1).

	Volume for 1x
10x Expand Long Template Buffer 1	80 µL
dNTPs (10 mM)	16 µL
Forward primer	1.12 nmol
Reverse primer	1.12 nmol
Expand Long Template Polymerase (5 U/µL)	11.2 µL
DNA	3.2 µg
Nuclease-free water	to 800 µL

Table 3. PCR reaction mixture for the final inverse PCR amplification of 4C template (Step 11.4).

	Volume for 1x
5x Phusion HF buffer	2 μL
dNTPs (10 mM)	0.2 μL
Miltiplexing Primer 1.0	5 pmol
Miltiplexing Primer 2.0	0.1 pmol
Index primer	5 pmol
100x SYBR Green I	0.1 μL
50x ROX	0.2 μL
Phusion polymerase	0.1 μL
DNA	2 μL
Nuclease-free water	to 10 μL

Table 4: qPCR reaction mixture for the determination of the number of amplification cycles for sequencing library prep (Step 13.3).

	Volume for 1x
5x Phusion HF buffer	10 μL
dNTPs (10 mM)	1 μL
Miltiplexing Primer 1.0	25 pmol
Miltiplexing Primer 2.0	0.5 pmol
Index primer	25 pmol
Phusion polymerase	0.5 μL
DNA	10 μL
Nuclease-free water	to 50 μL

Table 5: PCR reaction mixture for the amplification of libraries for sequencing (Step 13.4).

Discussion

4C results have the potential to reveal chromatin interactions that can identify previously unknown regulatory elements and/or target genes that are important in a specific biological context[24,25,26]. However, technical hurdles may limit the data obtained from these experiments. PCR bias stemming from over-amplification of template in 4C protocols is likely. This protocol addresses this issue by utilizing qPCR to determine the optimal number of amplification cycles in an objective manner. In addition, the removal of the bait sequences from amplified inverse PCR products by restriction digest can facilitate the identification of chromatin interactions for two reasons. First, it reduces the length of non-informative (bait) base pairs from the material to be sequenced. Second, and as a result, it increases the likelihood that more reads will be generated from diverse sequences (a property required for accurate base calling) and thus more informative interacting sequences can be mapped. Other protocols require the pooling of many libraries using different bait sequences and/or restriction enzymes or require increasing the phiX concentration of the sequenced sample to circumvent the sequence uniformity issue for accurate base calling. This method allows multiple samples with the same bait sequence to be pooled into a single sequencing lane without occupying valuable sequencing capacity with excess phiX.

Cell fixation and lysis are critical early steps, both of which may require optimization for particular cell types. Insufficient fixation will fail to preserve specific contacts between a region of interest and its interacting sequences, yielding uninformative data dominated by noise. In contrast, over-fixation will decrease the ability of restriction enzymes to cleave chromatin, resulting in fewer informative ligation events. Both fixative concentration and incubation time can be altered to optimize this variable. Similarly, insufficient cell lysis reduces access of restriction enzymes to chromatin, again reducing the number of informative ligation events. In our hands, human primary keratinocytes lysed most effectively using a combination of hypotonic conditions and detergent. Other cell types might require different lysis conditions, which will have to be determined empirically. Efficient lysis can be identified by microscopy dye exclusion methods such as Trypan Blue staining.

One limitation of the 4C method is that the results can only represent a population average. With a heterogeneous cell population, it can be difficult to determine true interactions vs. noise due to biological variability. While the use of a cell line or the sorting of cells to produce a homogeneous cell population is predicted to generate clearer signals, variability between individual cells is still a possible source of noise. Recent advances in single-cell sequencing technologies have the potential to overcome this problem. Additionally, the validation of population-based 4C results may be performed using methods such as digital droplet PCR or FISH to determine if these results are reflected at the single-cell level.

Disclosures

The authors declare that they have no competing financial interests.

Acknowledgements

References

1. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature.* **489** (7414), 57-74 (2012).
2. Hoffman, M.M. *et al.* Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Research.* **41** (2), 827-841 (2013).
3. Dekker, J., Rippe, K., Dekker, M., Kleckner, N. Capturing Chromosome Conformation. *Science.* **295** (5558), 1306-1311 (2002).
4. Sanyal, A., Lajoie, B.R., Jain, G., Dekker, J. The long-range interaction landscape of gene promoters. *Nature.* **489** (7414), 109-113 (2012).
5. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell.* **148** (1-2), 84-98 (2012).
6. Schmitt, A.D., Hu, M., Ren, B. Genome-wide mapping and analysis of chromosome architecture. *Nature Reviews Molecular Cell Biology.* **17** (12), 743-755 (2016).
7. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature genetics.* **38** (11), 1341-1347 (2006).
8. Splinter, E., de Wit, E., van de Werken, H.J.G., Klous, P., de Laat, W. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: From fixation to computation. *Methods.* **58** (3), 221-230 (2012).
9. Van De Werken, H.J.G. *et al. 4C technology: Protocols and data analysis. Methods in Enzymology.* **513** Elsevier Inc. (2012).
10. Brouwer, R.W.W., van den Hout, M.C.G.N., van IJcken, W.F.J., Soler, E., Stadhouders, R. Unbiased Interrogation of 3D Genome Topology Using Chromosome Conformation Capture Coupled to High-Throughput Sequencing (4C-Seq). *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation.* 199-220 (2017).
11. Matelot, M., Noordermeer, D. Determination of High-Resolution 3D Chromatin Organization Using Circular Chromosome Conformation Capture (4C-seq). *Polycomb Group Proteins: Methods and Protocols.* 223-241 (2016).
12. Gheldof, N., Leleu, M., Noordermeer, D., Rougemont, J., Reymond, A. Detecting Long-Range Chromatin Interactions Using the Chromosome Conformation Capture Sequencing (4C-seq) Method. *Gene Regulatory Networks: Methods and Protocols.* 211-225 (2012).
13. Göndör, A., Rougier, C., Ohlsson, R. High-resolution circular chromosome conformation capture assay. *Nature protocols.* **3** (2), 303-13 (2008).
14. Hagège, H. *et al.* Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nature protocols.* **2** (7), 1722-1733 (2007).
15. Anand, R.D., Sertil, O., Lowry, C. V. Restriction digestion monitors facilitate plasmid construction and PCR cloning. *BioTechniques.* **36** (6), 982-985 (2004).
16. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic acids research.* **44** (W1), W3-W10 (2016).
17. Li, H., Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* **26** (5), 589-595 (2010).
18. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* **25** (16), 2078-2079 (2009).
19. Walter, C., Schuetzmann, D., Rosenbauer, F., Dugas, M. Basic4Cseq: An R/Bioconductor package for analyzing 4C-seq data. *Bioinformatics.* **30** (22), 3268-3269 (2014).
20. Raviram, R. *et al.* 4C-ker: A Method to Reproducibly Identify Genome-Wide Interactions Captured by 4C-Seq Experiments. *PLoS Computational Biology.* **12** (3) (2016).
21. Klein, F.A., Pakozdi, T., Anders, S., Ghavi-helm, Y., Furlong, E.E.M., Huber, W. FourCSeq: analysis of 4C sequencing data. *Bioinformatics.* **31** (19), 3085-3091 (2015).
22. Williams Jr, R.L. *et al.* fourSig : a method for determining chromosomal interactions in 4C-Seq data. *Nucleic Acids Research.* **42** (8) (2014).
23. McLean, C.Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology.* **28** (5), 495-501 (2010).
24. Stolzenburg, L.R. *et al.* Regulatory dynamics of 11p13 suggest a role for EHF in modifying CF lung disease severity. *Nucleic Acids Research.* **45** (15), 8773-8784 (2017).
25. Meddens, C.A. *et al.* Systematic analysis of chromatin interactions at disease associated loci links novel candidate genes to inflammatory bowel disease. *Genome Biology.* **17** (1), 1-15 (2016).
26. Yeung, J. *et al.* Transcription factor activity rhythms and tissue-specific chromatin interactions explain circadian gene expression across organs. *Genome research.* (00 41), 207787 (2017).