Journal of Visualized Experiments

High-throughput identification of gene regulatory sequences using next-generation sequencing of circular chromosome conformation capture (4C-seq)
--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE58030R2
Full Title:	High-throughput identification of gene regulatory sequences using next-generation sequencing of circular chromosome conformation capture (4C-seq)
Keywords:	chromatin; looping; enhancers; regulatory; sequencing; bait; restriction enzyme; lysis
Corresponding Author:	Cristina de Guzman Strong Washington University in Saint Louis School of Medicine Saint Louis, MO UNITED STATES
Corresponding Author's Institution:	Washington University in Saint Louis School of Medicine
Corresponding Author E-Mail:	CSTRONG@dom.wustl.edu
Order of Authors:	Erin A Brettmann
	Inez Y Oh
	Cristina de Guzman Strong
Additional Information:	
Question	Response
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US$2,400)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	McDonnell Basic Sciences Building RM770 4566 Scott Avenue St. Louis MO 63110

1 **TITLE:**
2 High-throughput Identification of Gene Regulatory Sequences Using Next-Generation
3 Sequencing of Circular Chromosome Conformation Capture (4C-seq)
4
5 **AUTHORS AND AFFILIATIONS:**
6 Erin A. Brettmann[1], Inez Y. Oh[1], and Cristina de Guzman Strong[1]
7 [1]Division of Dermatology, Center for Pharmacogenomics, Center for the Study of Itch,
8 Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA
9
10 **Corresponding Author:**
11 Cristina de Guzman Strong
12 Email Address: cristinastrong@wustl.edu
13 Tel: (314)-362-7695
14
15 **KEYWORDS:**
16 Chromatin, looping, enhancers, regulatory, sequencing, bait, restriction enzyme, lysis
17
18 **SUMMARY:**
19 The identification of physical interactions between genes and regulatory elements is
20 challenging but has been facilitated by chromosome conformation capture methods. This
21 modification to the 4C-seq protocol mitigates PCR bias by minimizing over-amplification of PCR
22 templates and maximizes the mappability of reads by incorporating an addition restriction
23 enzyme digest step.
24
25 **ABSTRACT:**
26 The identification of regulatory elements for a given target gene poses a significant technical
27 challenge owing to the variability in the positioning and effect sizes of regulatory elements to a
28 target gene. Some progress has been made with the bioinformatic prediction of the existence
29 and function of proximal epigenetic modifications associated with activated gene expression
30 using conserved transcription factor binding sites. Chromatin conformation capture studies
31 have revolutionized our ability to discover physical chromatin contacts between sequences and
32 even within an entire genome. Circular chromatin conformation capture coupled with next-
33 generation sequencing (4C-seq), in particular, is designed to discover all possible physical
34 chromatin interactions for a given sequence of interest (viewpoint), such as a target gene or a
35 regulatory enhancer. Current 4C-seq strategies directly sequence from within the viewpoint but
36 require numerous and diverse viewpoints to be simultaneously sequenced to avoid the
37 technical challenges of uniform base calling (imaging) with next generation sequencing
38 platforms. This volume of experiments may not be practical for many laboratories. Here, we
39 report a modified approach to the 4C-seq protocol that incorporates both an additional
40 restriction enzyme digest and qPCR-based amplification steps that are designed to facilitate a
41 greater capture of diverse sequence reads and mitigate the potential for PCR bias, respectively.
42 Our modified 4C method is amenable to the standard molecular biology lab for assessing
43 chromatin architecture.
44

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[8–12]. First, it uses a qPCR-based method to empirically determine the optimal number of amplification cycles for 4C-seq library preparation steps and

89   thus mitigates the potential for PCR bias stemming from over-amplification of libraries. Second,
90   it uses an additional restriction digest step in an effort to reduce the uniformity of known "bait"
91   sequences that hinders accurate base-calling by the sequencing instrument and, hence,
92   maximizes the unique, informative sequence in each read. Other protocols circumvent this
93   issue by pooling many (12-15)[8] 4C-seq libraries with different bait sequences and/or restriction
94   sites, a volume of experiments which may not be achievable by other laboratories. The
95   modifications presented here allow a small number of experiments, samples, and/or replicates
96   to be indexed and pooled into a single lane.
97
98   **PROTOCOL:**
99
100   **1. Restriction Enzyme Selection**
101
102   1.1. Identify a region of interest (*e.g.*, gene promoter, single nucleotide polymorphism (SNP),
103   enhancer) and obtain the DNA sequence from repositories such as National Center for
104   Biotechnology Information (NCBI).
105
106   1.2. Identify the candidate restriction enzymes (REs) for the first restriction enzyme digestion
107   (RE1) that do not cut within the sequence of interest, which produce sticky ends (DNA termini
108   with overhangs) after RE digestion, and whose activities are not inhibited by CpG methylation.
109
110   Note: The resolution of the data is determined by RE1. An enzyme with a 6 base pair (bp)
111   recognition sequence produces, on average, fragments approximately 4 kilobases (kb) in length,
112   while an enzyme with a 4 bp recognition sequence produces fragments approximately 250 bp in
113   length, allowing more precise identification of interacting sequences.
114
115   1.3. Select a RE1 from the candidate enzymes identified in Step 1.2 that produces a "viewpoint"
116   restriction fragment at least 500 bp long that includes the region of interest or, alternatively, a
117   neighboring region.
118
119   Note: The enzyme selected must be amenable to primer design (see Step 2).
120
121   1.4. For the second digestion, identify a restriction enzyme (RE2) with a 4 bp recognition
122   sequence that produces sticky ends (DNA termini with overhangs) after RE digestion, whose
123   activity is not inhibited by CpG methylation, and that cuts within the restriction fragment
124   generated by RE1 to produce a 250-500 bp bait fragment (**Figure 1**).
125
126   Note: The enzyme selected must be amenable to primer design (see Step 2).
127
128   **2. Design and Test Primers for Inverse PCR and Digestion Efficiency qPCR**
129
130   2.1. Use a primer design tool such as Primer3 (http://primer3.ut.ee) to design inverse primers
131   that are directed outwards towards the ends of the bait fragment (*i.e.*, the 5' primer is a
132   "reverse" primer, complementary to the plus strand, while the 3' primer is a "forward" primer,

133 complementary to the minus strand) and as close to the restriction sites as possible to minimize
134 the amplification of non-informative bait sequence and maximize PCR efficiency (**Figure 2A**).

136 Note: Primers should be predicted to have minimal non-specific amplification from genomic
137 DNA by *in silico* PCR predictions and should not align elsewhere in the genome except their
138 intended target with more than 16/18 identity[9]. As sequencing adapters are not incorporated
139 into the inverse PCR primers, the site of primer binding is more flexible than in other 4C
140 preparation methods; primers should optimally anneal within 50 bp of the end of the
141 corresponding restriction site.

143 2.1.1. Confirm the specificity of inverse PCR primers by amplification using purified genomic
144 DNA (gDNA) as template.

146 2.1.2. Optimal primers should yield few products when using gDNA as a template (see Step 11.2
147 for the description of expected 4C products). If substantial amplification from gDNA occurs,
148 design new primers. If no acceptable primer sets are identified, return to Step 1 and select a
149 new bait by selecting new restriction enzymes.

151 2.2. Design qPCR primers to amplify the fragments 70-200 nucleotide (nt) in length to monitor
152 the restriction digestion efficiency (**Figure 2B**).

154 2.2.1. Design one pair of primers to amplify across each restriction site (selected in Step 1) that
155 define the bait sequence. Design the primers for both RE1 (see Step 6.5.6) and RE2 sites (see
156 Step 9.2).

158 2.2.2. Design a set of primers that amplifies a region not containing the sites for either
159 restriction enzyme as a normalization control (uncut DNA) for RE1 and RE2 (see also Step 6.5.6).

161 **3. Collection of Cells**

163 3.1. Using a method that is appropriate for the tissue or cell culture, obtain a single-cell
164 suspension. Pellet the cells for 5 min at 200 x g, discard the supernatant and resuspend the
165 pellet in 500 µL of 1x phosphate buffered saline (PBS) per 10^7 cells.

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

169 4.1. Per 10^7 cells, add 9.5 mL of 1% electron microscopy (EM)-grade formaldehyde (methanol-
170 free) in PBS and incubate, while tumbling on rocker (or similar), for 10 min at room
171 temperature (RT, 18–22 °C).

173 Notes: Fixation conditions may have to be optimized for each cell type. Prior published work in
174 mouse cells reported that optimal fixation results in at least 40% of mapped reads aligning to
175 the chromosome containing the viewpoint[9] assuming a non-telomeric region for an average
176 sized chromosome.

177

178 4.2. Transfer the reaction tubes to ice and add ice-cold 1 M glycine to a final concentration of
179 0.125 M (1.425 mL) to quench the crosslinking reaction. Mix by gentle inversion.

180

181 4.3. Centrifuge for 5 min at 200 x g, 4 °C and carefully remove all the supernatant. Store the cell
182 pellet at -80 °C or proceed immediately to cell lysis.

183

184 **5. Cell Lysis**

185

186 5.1. Resuspend the cell pellet from Step 4.3 in 500 µL of 5 mM ethylenediaminetetraacetic acid
187 (EDTA) to wash. Centrifuge at 200 x g for 5 min at RT. Discard the supernatant.

188

189 5.2. Resuspend the cell pellet in 125 µL of 5 mM EDTA with 0.5% sodium dodecyl sulfate (SDS)
190 and 1x protease inhibitors, such as a 100x cocktail containing 5 mg/mL antipain, 10 mg/mL
191 chymostatin, 10 mg/mL leupeptin, and 10 mg/mL pepstatin A.

192

193 Note: Detergent concentration may have to be optimized for each cell type. Insufficient
194 detergent concentrations will result in incomplete cell lysis, leaving chromatin inaccessible to
195 restriction enzymes. If restriction digestion is persistently low in Step 6.5.6 and the activity of
196 the enzyme has been confirmed, additional detergent may be required. Increase SDS
197 concentration in 0.1% increments.

198

199 5.3. Incubate the cell suspension on ice for 10 min.

200

201 Note: For primary human keratinocytes, optimal lysis was achieved using a 10 min incubation at
202 65 °C followed by a 37 °C overnight incubation in a shaking heating block with agitation (900
203 rpm).

204

205 5.4. Ensure that cell lysis is complete.

206

207 5.4.1. Mix 6 µL of the cells with 6 µL of Trypan Blue on a slide and cover with a coverslip. View
208 under a microscope.

209

210 Note: Upon successful lysis, the interior of the cells should appear blue and un-lysed cells will
211 appear white.

212

213 5.4.2 If the cell lysis appears insufficient, pellet the cells at 200 x g for 5 min and save the
214 supernatant. Resuspend the pellet and repeat Steps 5.2-5.4 and/or alternatively with different
215 incubation temperatures (see the Note in Step 5.3).

216

217 5.4.3. Combine the re-lysed pellet with the saved supernatant and proceed.

218

219 Note: Visual inspection is not a guarantee of sufficient lysis. Digestion efficiency should be
220 objectively determined as in Step 6.5.6.

6. First Restriction Digestion

6.1. 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.

6.2. Remove a 15 μL aliquot as an "Undigested control" and store at 4 °C.

6.3. 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.

6.4. Remove a 15 μL aliquot as a "Digested control" and store at 4 °C.

6.5. Determine digestion efficiency:

6.5.1. 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.

6.5.2. 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.

6.5.3. 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.

6.5.4. 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.

6.5.5. Remove the supernatant and dry the pellet at room temperature for 2 min. Resuspend the pellet in 50 μL of nuclease-free H_2O.

6.5.6. 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

263 7.1. Heat-inactivate the restriction enzyme by incubating 20 min at 65 °C. Alternatively, if the
264 enzyme cannot be heat inactivated, phenol-chloroform extract and ethanol precipitate the
265 sample.
266
267 Note: The higher inactivation temperatures recommended for some restriction enzymes
268 denature the proteins in the chromatin, and this may negatively affect the sample quality.
269 Additionally, phenol:chloroform extraction is not ideal, as it results in the loss of sample.
270
271 7.2. Transfer to a 50 mL conical tube and add 6 mL of nuclease-free H_2O, 700 µL of 10x ligase
272 buffer (660 mM Tris-HCl pH 7.5, 50 mM magnesium chloride ($MgCl_2$), 10 mM dithiothreitol
273 (DTT), 10 mM adenosine triphosphate (ATP)), and 50 U T4 DNA Ligase. Mix gently by swirling
274 and incubate overnight at 16 °C.
275
276 7.3. Remove a 100 µL aliquot of the sample as the ''Ligation control''.
277
278 7.4. Determine ligation efficiency:
279
280 7.4.1. Add 2.5 µL of Proteinase K (20mg/mL) and incubate 1 h at 65 °C.
281
282 7.4.2. Add 100 µL of phenol-chloroform and mix vigorously by inversion. Centrifuge for 5 min,
283 16,100 x g at RT.
284
285 7.4.3. Transfer the aqueous phase to a new tube. Add 6.66 µL of 3 M sodium acetate pH 5.2, 1
286 µL of glycogen, and 300 µL of 100% EtOH. Mix gently by inversion and place at -80 °C about 1 h.
287
288 7.4.4. Centrifuge for 20 min at 16,100 x g at 4 °C. Remove the supernatant, add 500 µL of 70%
289 EtOH, and centrifuge for 5 min at 16,100 x g at 4 °C.
290
291 7.4.5. Remove the supernatant and dry the pellet at room temperature. Resuspend the pellet in
292 20 µL of water and load on a 0.6% agarose gel next to the "digestion controls" from Step 6.5.
293
294 Note: A well-ligated sample should appear as a relatively tight, high molecular weight band
295 (**Figure 3**).
296
297 7.4.6. If the ligation is sufficient, proceed with Step 8. Otherwise, add fresh ATP (1 mM final
298 concentration) and new ligase; incubate overnight at 16 °C and repeat Steps 7.3-7.4.
299
300 **8. Reverse Cross-linking and Isolate Chromatin**
301
302 8.1. Add 15 µL of Proteinase K (20 mg/mL) and incubate overnight at 65 °C to reverse cross-
303 links.
304
305 8.2. Add 30 µL of RNase A (10 mg/mL) and incubate 45 min at 37 °C.
306

307 <mark>8.3. Add 7 mL of phenol-chloroform and mix vigorously by inversion. Centrifuge 15 min, 3300 x</mark>
308 <mark>g at RT.</mark>
309
310 <mark>8.4. Transfer the aqueous phase to a new 50 mL tube and add 7.5 mL of nuclease-free H_2O (to</mark>
311 <mark>dilute the DTT present in the ligase buffer, which otherwise precipitates with the DNA), 1 mL of</mark>
312 <mark>3 M sodium acetate pH 5.6, 7 μL of glycogen (20 mg/mL), and 35 mL of 100% EtOH. Mix and</mark>
313 <mark>incubate at -80 °C for 1 h.</mark>
314
315 <mark>8.5. Centrifuge 20 min, 3900 x g at 4 °C. Remove the supernatant (pellet may be difficult to</mark>
316 <mark>see), wash the pellet with 10 mL of ice-cold 70% EtOH, and centrifuge 15 min, 3300 x g at 4 °C.</mark>
317
318 <mark>8.6. Remove the supernatant and briefly dry the pellet at RT. Dissolve the pellet in 150 μL of 10</mark>
319 <mark>mM Tris-HCl pH 7.5 at 37 °C. Store at -20 °C or continue with Step 9.</mark>
320
321 **9. Second Restriction Digestion: Trimming the Circles**
322
323 Note: This step creates smaller circles to minimize the overrepresentation of smaller captured
324 fragments due to PCR bias in downstream amplification steps.
325
326 9.1. To the sample from Step 8.6, add 50 μL of 10x restriction enzyme buffer (specified by the
327 manufacturer), 300 μL of nuclease-free H_2O, and 50 U restriction enzyme RE2. Incubate
328 overnight at the temperature appropriate for the chosen enzyme.
329
330 9.2. Remove a 15 μL aliquot of the sample as the "Digestion control". Determine digestion
331 efficiency as described in Step 6.5.
332
333 **10. Second Ligation and DNA Purification**
334
335 10.1. Inactivate the restriction enzyme as recommended by the manufacturer. If the enzyme
336 cannot be heat-inactivated, remove the enzyme using a column-based purification kit.
337
338 Note: As this results in the loss of sample, column purification is not ideal.
339
340 10.2 Transfer the sample to a 50 mL tube and add 12.1 mL of nuclease-free H_2O, 1.4 mL of 10x
341 ligation buffer (660 mM Tris-HCl, pH 7.5; 50 mM $MgCl_2$; 10 mM DTT; 10 mM ATP), and 100 U T4
342 DNA ligase. Incubate overnight at 16 °C.
343
344 10.3 Add 467 μL of 3 M sodium acetate pH 5.6, 233 μL nuclease-free H_2O, 7 μL glycogen (20
345 mg/mL), and 35 mL 100% EtOH. Mix well and incubate at -80 °C 1 h.
346
347 10.4. Centrifuge 45 min, 3900 x g at 4 °C. Remove the supernatant, add 10 mL of cold 70%
348 EtOH, and centrifuge 15 min, 3300 x g at 4 °C.
349

350 10.5. Remove the supernatant and briefly dry the pellet at room temperature. Add 150 μL of 10
351 mM Tris-HCl pH 7.5 and incubate at 37 °C to dissolve the pellet.

352

353 10.6. Purify the samples with a silica column-based PCR purification kit, following the
354 manufacturer's instructions. Use 1 column per 3×10^6 cells, based on the initial number of the
355 cells. Elute the columns with 50 μL of 10 mM Tris-HCl, pH 7.5 and pool the samples.

356

357 10.7. Measure the concentration by fluorimetry or spectrophotometry using the absorbance at
358 260 nm. The 4C template is now ready for inverse PCR. Store at -20 °C or proceed directly to
359 Step 11.

360

361 **11. PCR Amplify Unknown Interacting Sequences by Inverse PCR**

362

363 11.1. Determine the linear range of amplification by performing a PCR using template dilutions
364 of 12.5, 25, 50, and 100 ng 4C template. If desired, amplify from gDNA in parallel to directly
365 compare products in order to identify non-specific amplification. Run PCR reactions using an
366 initial denaturation at 94 ˚C for 2 min; 30 cycles consisting of a denaturation step at 94 ˚C for 10
367 s, an annealing step at 55 ˚C for 1 min, and an extension step at 68 ˚C for 3 min; and a final
368 extension at 68 ˚C for 5 min.

369

370 11.2. Separate 15 μL of each PCR product on a 1.5% agarose gel to confirm linear amplification
371 and assess template quality (**Figure 4**). Amplification from 4C template should yield discrete
372 banding at low DNA concentrations and a smear at higher concentrations. The presence of the
373 smear indicates increased complexity of the amplified 4C ligations.

374

375 11.3. When satisfied with the quality and quantity of the inverse PCR product generated, set up
376 a qPCR to determine the optimal number of cycles to use for amplification:

377

378 11.3.1. Set up the reactions containing SYBR and ROX dyes using the reaction mixture in **Table**
379 **2**. Unless the amplification is not linear at high concentrations in Step 11.2, use 100 ng of the
380 template per reaction.

381

382 Note: The addition of ROX facilitates the normalization of fluorescent signal from well to well
383 and cycle to cycle.

384

385 11.3.2. Run the PCR reactions using an initial denaturation at 94 ˚C for 2 min; 40 cycles
386 consisting of a denaturation step at 94 ˚C for 10 s, an annealing step at 55 ˚C for 1 min, and an
387 extension step at 68 ˚C for 3 min; and a final extension at 68 ˚C for 5 min.

388

389 11.3.3. Determine the peak (endpoint) fluorescence of the reactions using the amplification
390 plot. Determine the cycle at which reactions reach 25% of peak fluorescence (**Figure 5**).

391

392 Note: This is the number of cycles that will be used to amplify the 4C libraries (see Step 11.4).

393

394 11.4. Set up inverse PCR as in **Table 3** to amplify unknown sequences ligated to the bait from
395 the 4C template. Divide into 16 reactions of 50 µL before running. Run PCR reactions using an
396 initial denaturation at 94 ˚C for 2 min and cycles consisting of a denaturation step at 94 ˚C for
397 10 s, an annealing step at 55 ˚C for 1 min, and an extension step at 68 ˚C for 3 min. Use the
398 number of cycles determined in Step 11.3.3.

400 11.5. Collect and pool the reactions. Purify using a silica column-based PCR purification kit. Use
401 at least 2 columns per 16 reactions. Pool the purified PCR products.

403 11.6. Determine the sample quantity and purity by spectrophotometry. Typical yields are
404 between 10 and 20 µg with A260/A280 ~1.85. If the absorption ratios are sub-optimal, re-purify
405 in order to prevent the problems during sequencing.

407 11.7. Assess the complexity of library by separating 300 ng of purified PCR product on a 1.5%
408 agarose gel.

410 Note: Amplified product should resemble that from Step 11.2.

412 **12. Third Restriction Digestion: Trim off Bait Sequences**

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

421 12.1. Obtain a RE digest monitor and test the digestion in a 50 µL reaction.

423 Note: This is a molecule of dsDNA (for example, a plasmid, PCR amplicon, or synthetic DNA)
424 that contains the RE site(s). The only requirement is that it should be easy to distinguish cut
425 from uncut monitor on an agarose gel. Optimize RE monitor mass and enzyme concentration if
426 needed.

428 12.2. Digest 1 µg of purified inverse PCR product from Step 11.6 and, in parallel, the RE monitor
429 from Step 12.1.

431 Note: DNA and enzyme concentration, as well as incubation time, should be the same as for the
432 test digest in Step 12.1 for both the inverse PCR product and the monitor digests. Adjust the
433 reaction volume as needed.

435 12.3. Run the digested RE monitor on an agarose gel of concentration appropriate for the
436 expected fragments. Digestion is considered sufficient when <10% of the DNA remains uncut
437 (**Figure 6**).

438

439 12.4. Purify the digested inverse PCR product on a silica column-based DNA purification kit. If
440 sequential digests are required, repeat Steps 12.1-12.4 with the second enzyme.

441

442 **13. Preparation of Sequencing Library**

443

444 13.1. Ligate the adapters compatible with the respective next generation sequencing platform.

445

446 13.1.1. Design the adapters for each RE1 and RE2 so that, after annealing the oligos, each RE
447 adaptor has the respective 1-sided overhang.

448

449 Note: For example, if HindIII is used, the annealed adapter should contain a 5' phosphorylated
450 "AGCT" overhang. Alternatively, if standard T-overhang adapters are used, end-repair and A-tail
451 the library prior to adapter ligation.

452

453 13.1.2. Anneal the respective adaptor RE oligos. Resuspend the oligos in annealing buffer (10
454 mM Tris, pH 7.5, 50 mM sodium chloride (NaCl), 1 mM EDTA), mix in equimolar quantities to 50
455 µM, heat to 95 ˚C, and allow to cool slowly to 25 ˚C.

456

457 13.1.3. Perform the ligation reaction. Mix RE-digested 4C library with a total 5- to 10-fold molar
458 excess of annealed adaptors (Step 13.1.2; 50:50 ratio for each RE adaptor used) in 1x ligase
459 buffer and add 6U DNA ligase. Incubate according to manufacturer instructions.

460

461 13.1.4. Remove excess adapters by purifying using a low-elution volume silica-based column kit.

462

463 13.2. Size-select.

464

465 Note: Size selection can also be performed using a bead-based cleanup kit and is recommended
466 even after column purification.

467

468 13.2.1. Cast a 2% agarose gel using a high-resolution agarose, adding a non-UV stain prior to
469 pouring. Ensure that any water lost due to evaporation during heating is replaced by weighing
470 the bottle or flask before and after heating and adding water to recover the lost mass.

471

472 13.2.2. Run the adapter-ligated libraries on the gel, leaving empty lanes between the samples.

473

474 13.2.3. Excise the gel slices corresponding to the size range 150 bp-1 kb using a clean scalpel or
475 razor blade.

476

477 13.2.4. Purify the libraries from the gel using a gel extraction kit. To minimize GC bias in the
478 recovery of DNA, dissolve the gel slices at room temperature with rotation.

479

480 13.3. Determine the number of cycles for PCR amplification by qPCR using the reaction mixture
481 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.

13.4. 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.

13.5. Purify amplified libraries using a low-elution volume silica-based column kit.

14. Sequencing and Analysis of Sequencing Data

14.1. 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).

14.2. 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.

14.3. 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

15.1. 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].

15.2. 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/).

524   15.3. Map demultiplexed reads to the appropriate genome of interest (*e.g.*, UCSC mm10, hg19)
525   using the Burrows-Wheeler Aligner (BWA)[17] or other alignment software resulting in SAM
526   output files. If needed, convert SAM files to BAM via samtools[18] (see Step 15.4).
527
528   15.4. Use a 4C-seq analysis pipeline such as Basic4Cseq[19], 4C-ker[20], FourCSeq[21], or fourSig[22] to
529   analyze the data and identify the interactions.
530
531   Notes: Data analysis and the assessment of the quality thereof should be conducted according
532   to the selected software package's reference manual. Packages generate BED output files
533   except where noted. Briefly, Basic4CSeq[19] uses an input SAM file (Step 15.3) to visualize the
534   interactions (txt, tiff, BED, and wig output files) and assesses the quality of data based on the
535   criteria set forth in van de Werken *et al*.[9] 4C-ker[20] (input trimmed FASTQ, step 15.2) and
536   FourCSeq[21] (input BAM, Step 15.3) identify differential interactions between conditions.
537   fourSig[22] (input SAM) also identifies significant interactions and prioritizes those that are likely
538   to be reproducible. Tools such as the Genomic Regulatory Enrichment of Annotations Tool
539   (GREAT)[23], or integration with ENCODE data sets may also be used to predict the biological
540   function of interactions discovered by 4C-seq.
541
542   **REPRESENTATIVE RESULTS:**
543   Primary human keratinocytes were isolated from 2-3 discarded neonatal foreskins, pooled, and
544   cultured in KSFM supplemented with 30 µg/mL bovine pituitary extract, 0.26 ng/mL
545   recombinant human epidermal growth factor, and 0.09 mM calcium chloride ($CaCl_2$) at 37 ˚C,
546   5% carbon dioxide. The cells were split into two flasks, and one flask was differentiated by the
547   addition of $CaCl_2$ to a final concentration of 1.2 mM for 72 h. 10^7 cells each from proliferating
548   and differentiated keratinocyte populations were fixed in 1% formaldehyde for 10 min at room
549   temperature. Separately, K562 cells were grown in RPMI supplemented with 10% fetal bovine
550   serum at 37 ˚C, 5% carbon dioxide. 10^7 cells were fixed in 1% formaldehyde for 10 min at room
551   temperature.
552
553   The cells were lysed, and the fixed chromatin was digested with HindIII and ligated as described
554   above. Cross-links were reversed, and the DNA is purified and digested with CviQI. DNA was
555   ligated and used as the template for inverse PCR amplification. Primers used were: 5'-
556   GATCAGGAGGGACTGGAACTTG/5'- CCTCCCTTCACATCTTAGAATG. 1 µg of purified inverse PCR
557   product was digested overnight with CviQI in parallel with 225 ng of RE digest monitor and
558   column-purified. Purified DNA was then digested overnight with HindIII in parallel with 125 ng
559   of RE digest monitor and column-purified. 250 ng of purified double-digested inverse PCR
560   product was used for library preparation. Briefly, the ends were repaired via conversions to 5'-
561   phosphorylated and blunt-ends and subsequent DNA column-purification. Ends were A-tailed
562   20 min at 72 ˚C in a 50 µL reaction with 1 U Taq polymerase and 200 µM dATP and column-
563   purified. Compatible sequencing library prep adapters were ligated at a 10:1 (adapter:DNA)
564   ratio using T4 DNA ligase in a 30 µL reaction at room temperature for 15 min and the DNA
565   column-purified. Libraries were run on a 2% agarose gel, gel slices were cut from 120 bp to the
566   top of the ladder to remove adapters, and libraries were purified. 200 pg of each library were
567   evaluated in 10 µL qPCR reactions containing indexing primers to determine optimal PCR

568   amplification cycles. Libraries were amplified to add indices using the number of cycles
569   determined by qPCR and sequenced on a HiSeq2500 to obtain 1x50 reads.
570
571   Reads were demultiplexed and trimmed. First, sequencing adapters were removed.
572   Subsequently, the sequence from the beginning of each primer to the restriction site was
573   trimmed. This enables the mapping of inverse PCR products that were not fully digested prior
574   to library preparation. Importantly, including sequence between the primer binding site and the
575   restriction site prevents mapping of non-specifically-amplified PCR product. Trimmed reads
576   were mapped to hg38 using BWA. **Figure 7** shows reads mapping to the region surrounding the
577   viewpoint.
578
579   **FIGURE AND TABLE LEGENDS:**
580   **Figure 1. Schematic of 4C-seq workflow.** Chromatin is cross-linked to preserve protein-DNA
581   contacts, digested with RE1, and ligated to link interacting loci. Cross-links are reversed, and
582   DNA is digested with RE2 and ligated. Unknown interacting sequences are amplified using
583   primers that bind to the region of interest, and PCR products are digested with RE1 and RE2 to
584   trim known sequences. Amplified DNA of unknown sequence is used in a sequencing library
585   prep, in which the adapters are ligated and the library is PCR-amplified and sequenced.
586
587   **Figure 2. Schematics of primer designs.** (A) Binding sites for inverse PCR primers. Primers are
588   oriented "outward" (*i.e.*, the 5' primer is a "reverse" primer, complementary to the plus strand,
589   while the 3' primer is a "forward" primer, complementary to the minus strand) and bind within
590   50 bp of each restriction site (indicated by red shading). (B) Binding sites for qPCR primers for
591   determining restriction enzyme digest efficiency. One primer pair flanks each restriction
592   enzyme site. A control primer set amplifies a sequence that does not contain a site for either
593   enzyme and is used to normalize Ct values for DNA input.
594
595   **Figure 3. Agarose gel electrophoresis to assess the ligation efficiency for RE1-digested**
596   **chromatin.** Uncut, HindIII-digested, and ligated samples were treated with proteinase K and
597   heated to reverse cross-links, phenol:chloroform extracted, EtOH precipitated, and
598   resuspended in H_2O. Purified DNA was run on a 0.6% agarose gel and bands visualized by
599   ethidium bromide staining with comparison to 1 kb plus ladder.
600
601   **Figure 4. 4C template titration for inverse PCR.** Serial dilutions were made of 4C templates and
602   used in inverse PCR reactions. PCR products were run on 1.5% agarose gel and visualized by
603   ethidium bromide staining alongside 1 kb plus ladder. NTC denotes no-template control
604   reaction. Note a correlative increase in amplification with an increase in template
605   concentration. This representative amplification was conducted on chromatin cleaved with
606   HindIII as RE1 and CviQI as RE2, and primer sequences used for inverse PCR amplification were
607   5'-GATCAGGAGGGACTGGAACTTG/5'- CCTCCCTTCACATCTTAGAATG.
608
609   **Figure 5. qPCR-mediated determination of amplification cycles needed to amplify template**
610   **for inverse PCR.** 4C templates were amplified in reactions containing 1x SYBR Green and 1x ROX
611   in a real-time thermocycler. Peak fluorescence was determined and the number of cycles

612   required to reach ¼ of peak fluorescence was calculated. This number of cycles was used to
613   amplify the 4C template for inverse PCR.

614

615   **Figure 6. Agarose gel electrophoresis of digested RE monitor indicating sufficient digestion.** RE
616   digest monitor was amplified from human genomic DNA using the primers F: 5'-
617   TCCTATCCCTGGTCTGTCTTAT and R: 5'-CCACATTGGTCCTTCTAGTCTTC and purified. 225 ng of
618   monitor was digested with 15 U CviQI in a 50 µL reaction at 25 ˚C overnight and run on a 1.5%
619   agarose gel alongside 1 kb plus ladder. Bands were visualized with ethidium bromide staining.
620   Uncut monitor is 2515 bp; expected fragment sizes are 132, 343, 488, 539, and 1013 bp.

621

622   **Figure 7. Representative genomic track of read coverage within 5 kb of the viewpoint region.**
623   Reads were trimmed and mapped to hg38 using BWA. The majority of reads (blue peaks) align
624   at HindIII or CviQI sites adjacent to HindIII sites, as expected. The viewpoint region is
625   highlighted in red.

626

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

628

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

631

632   **Table 3. PCR reaction mixture for the final inverse PCR amplification of 4C template (Step**
633   **11.4).**

634

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

637

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

639

640   **DISCUSSION:**
641   4C results have the potential to reveal chromatin interactions that can identify previously
642   unknown regulatory elements and/or target genes that are important in a specific biological
643   context[24–26]. However, technical hurdles may limit the data obtained from these experiments.
644   PCR bias stemming from over-amplification of template in 4C protocols is likely. This protocol
645   addresses this issue by utilizing qPCR to determine the optimal number of amplification cycles
646   in an objective manner. In addition, the removal of the bait sequences from amplified inverse
647   PCR products by restriction digest can facilitate the identification of chromatin interactions for
648   two reasons. First, it reduces the length of non-informative (bait) base pairs from the material
649   to be sequenced. Second, and as a result, it increases the likelihood that more reads will be
650   generated from diverse sequences (a property required for accurate base calling) and thus
651   more informative interacting sequences can be mapped. Other protocols require the pooling of
652   many libraries using different bait sequences and/or restriction enzymes or require increasing
653   the phiX concentration of the sequenced sample to circumvent the sequence uniformity issue
654   for accurate base calling. This method allows multiple samples with the same bait sequence to
655   be pooled into a single sequencing lane without occupying valuable sequencing capacity with

656   excess phiX.

657

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

669

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

678

681

682   **DISCLOSURES:**
683   The authors declare that they have no competing financial interests.

684

685   **REFERENCES:**
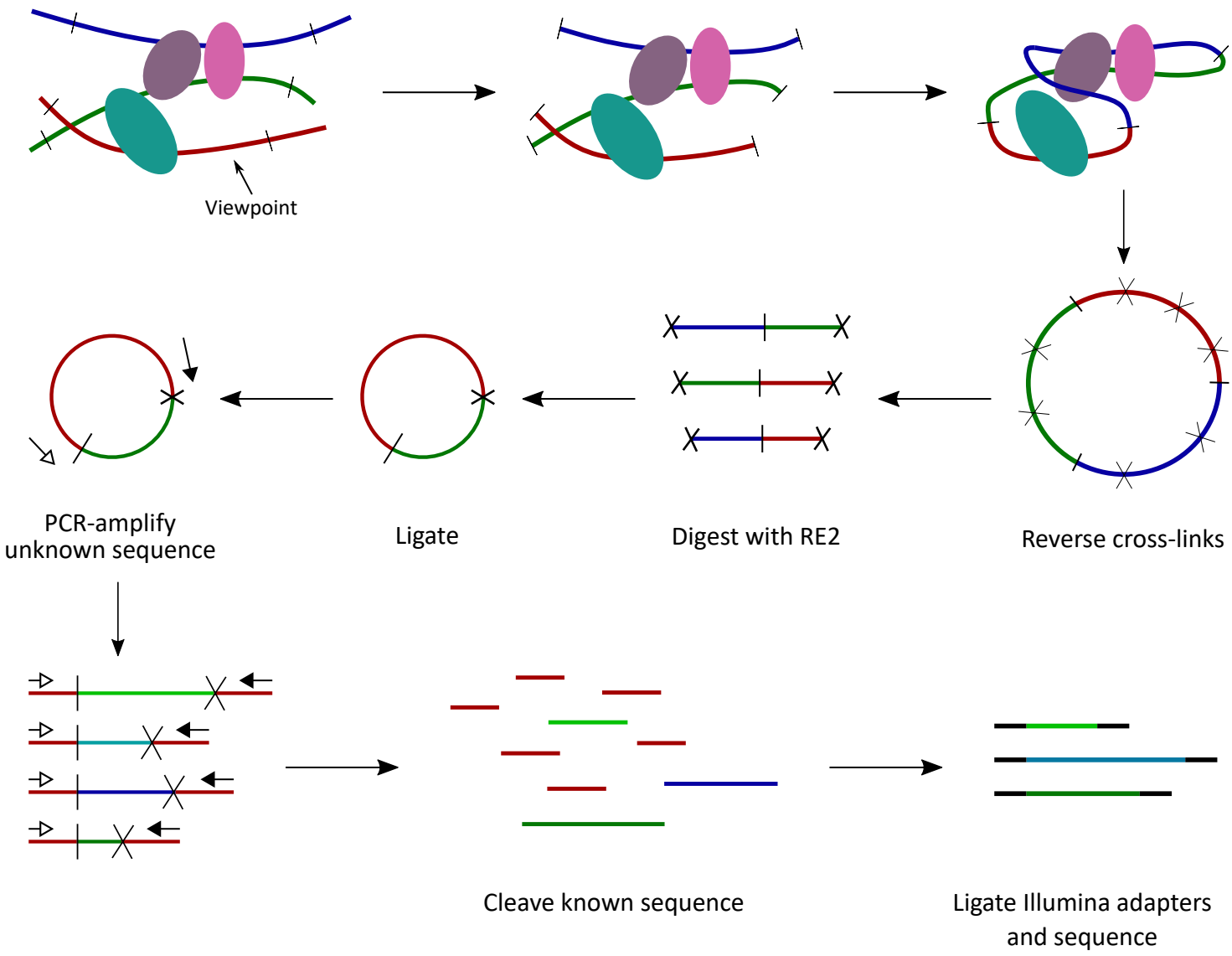686   1.   Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome.
687   *Nature*. **489** (7414), 57–74, doi: 10.1038/nature11247 (2012).
688   2.   Hoffman, M.M. *et al.* Integrative annotation of chromatin elements from ENCODE data.
689   *Nucleic Acids Research*. **41** (2), 827–841, doi: 10.1093/nar/gks1284 (2013).
690   3.   Dekker, J., Rippe, K., Dekker, M., Kleckner, N. Capturing Chromosome Conformation.
691   *Science*. **295** (5558), 1306–1311, doi: 10.1126/science.1067799 (2002).
692   4.   Sanyal, A., Lajoie, B.R., Jain, G., Dekker, J. The long-range interaction landscape of gene
693   promoters. *Nature*. **489** (7414), 109–113, doi: 10.1038/nature11279 (2012).
694   5.   Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological
695   basis for transcription regulation. *Cell*. **148** (1–2), 84–98, doi: 10.1016/j.cell.2011.12.014
696   (2012).
697   6.   Schmitt, A.D., Hu, M., Ren, B. Genome-wide mapping and analysis of chromosome
698   architecture. *Nature Reviews Molecular Cell Biology*. **17** (12), 743–755, doi:
699   10.1038/nrm.2016.104 (2016).

700   7.   Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive
701   networks of epigenetically regulated intra- and interchromosomal interactions. *Nature*
702   *genetics*. **38** (11), 1341–1347, doi: 10.1038/ng1891 (2006).
703   8.   Splinter, E., de Wit, E., van de Werken, H.J.G., Klous, P., de Laat, W. Determining long-
704   range chromatin interactions for selected genomic sites using 4C-seq technology: From
705   fixation to computation. *Methods*. **58** (3), 221–230, doi: 10.1016/j.ymeth.2012.04.009
706   (2012).
707   9.   Van De Werken, H.J.G. *et al. 4C technology: Protocols and data analysis*. *Methods in*
708   *Enzymology*. **513**, doi: 10.1016/B978-0-12-391938-0.00004-5. Elsevier Inc. (2012).
709   10.   Brouwer, R.W.W., van den Hout, M.C.G.N., van IJcken, W.F.J., Soler, E., Stadhouders, R.
710   Unbiased Interrogation of 3D Genome Topology Using Chromosome Conformation
711   Capture Coupled to High-Throughput Sequencing (4C-Seq). *Eukaryotic Transcriptional*
712   *and Post-Transcriptional Gene Expression Regulation*. 199–220, doi: 10.1007/978-1-4939-
713   6518-2_15 (2017).
714   11.   Matelot, M., Noordermeer, D. Determination of High-Resolution 3D Chromatin
715   Organization Using Circular Chromosome Conformation Capture (4C-seq). *Polycomb*
716   *Group Proteins: Methods and Protocols*. 223–241, doi: 10.1007/978-1-4939-6380-5_20
717   (2016).
718   12.   Gheldof, N., Leleu, M., Noordermeer, D., Rougemont, J., Reymond, A. Detecting Long-
719   Range Chromatin Interactions Using the Chromosome Conformation Capture Sequencing
720   (4C-seq) Method. *Gene Regulatory Networks: Methods and Protocols*. 211–225, doi:
721   10.1007/978-1-61779-292-2_13 (2012).
722   13.   Göndör, A., Rougier, C., Ohlsson, R. High-resolution circular chromosome conformation
723   capture assay. *Nature protocols*. **3** (2), 303–13, doi: 10.1038/nprot.2007.540 (2008).
724   14.   Hagège, H. *et al.* Quantitative analysis of chromosome conformation capture assays (3C-
725   qPCR). *Nature protocols*. **2** (7), 1722–1733, doi: 10.1038/nprot.2007.243 (2007).
726   15.   Anand, R.D., Sertil, O., Lowry, C. V. Restriction digestion monitors facilitate plasmid
727   construction and PCR cloning. *BioTechniques*. **36** (6), 982–985 (2004).
728   16.   Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative
729   biomedical analyses: 2016 update. *Nucleic acids research*. **44** (W1), W3–W10, doi:
730   10.1093/nar/gkw343 (2016).
731   17.   Li, H., Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler
732   transform. *Bioinformatics*. **26** (5), 589–595, doi: 10.1093/bioinformatics/btp698 (2010).
733   18.   Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25** (16),
734   2078–2079, doi: 10.1093/bioinformatics/btp352 (2009).
735   19.   Walter, C., Schuetzmann, D., Rosenbauer, F., Dugas, M. Basic4Cseq: An R/Bioconductor
736   package for analyzing 4C-seq data. *Bioinformatics*. **30** (22), 3268–3269, doi:
737   10.1093/bioinformatics/btu497 (2014).
738   20.   Raviram, R. *et al.* 4C-ker : A Method to Reproducibly Identify Genome-Wide Interactions
739   Captured by 4C-Seq Experiments. *PLoS Computational Biology*. **12** (3), doi:
740   10.1371/journal.pcbi.1004780 (2016).
741   21.   Klein, F.A., Pakozdi, T., Anders, S., Ghavi-helm, Y., Furlong, E.E.M., Huber, W. FourCSeq:
742   analysis of 4C sequencing data. *Bioinformatics*. **31** (19), 3085–3091, doi:
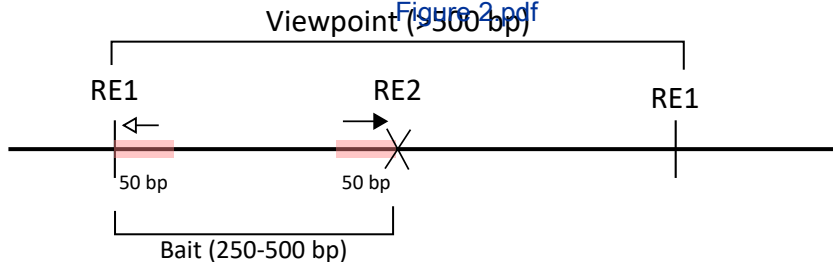743   10.1093/bioinformatics/btv335 (2015).

744 22. Williams Jr, R.L. *et al.* fourSig : a method for determining chromosomal interactions in
745   4C-Seq data. *Nucleic Acids Research*. **42** (8), doi: 10.1093/nar/gku156 (2014).
746 23. McLean, C.Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions.
747   *Nature Biotechnology*. **28** (5), 495–501, doi: 10.1038/nbt.1630 (2010).
748 24. Stolzenburg, L.R. *et al.* Regulatory dynamics of 11p13 suggest a role for EHF in modifying
749   CF lung disease severity. *Nucleic Acids Research*. **45** (15), 8773–8784, doi:
750   10.1093/nar/gkx482 (2017).
751 25. Meddens, C.A. *et al.* Systematic analysis of chromatin interactions at disease associated
752   loci links novel candidate genes to inflammatory bowel disease. *Genome Biology*. **17** (1),
753   1–15, doi: 10.1186/s13059-016-1100-3 (2016).
754 26. Yeung, J. *et al.* Transcription factor activity rhythms and tissue-specific chromatin
755   interactions explain circadian gene expression across organs. *Genome research*. (00 41),
756   207787, doi: 10.1101/207787 (2017).
757

Figure 1 Cross-link chromatin
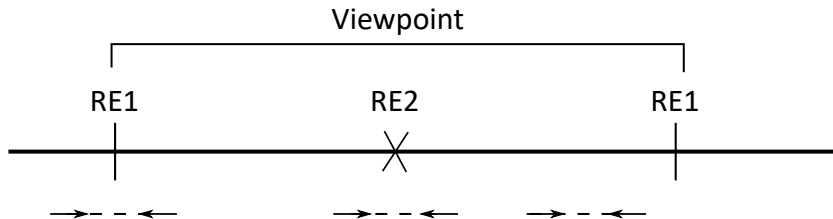
Digest with RE1

Ligate

Viewpoint

PCR-amplify
unknown sequence

Ligate

Digest with RE2

Reverse cross-links

Cleave known sequence

Ligate Illumina adapters
and sequence
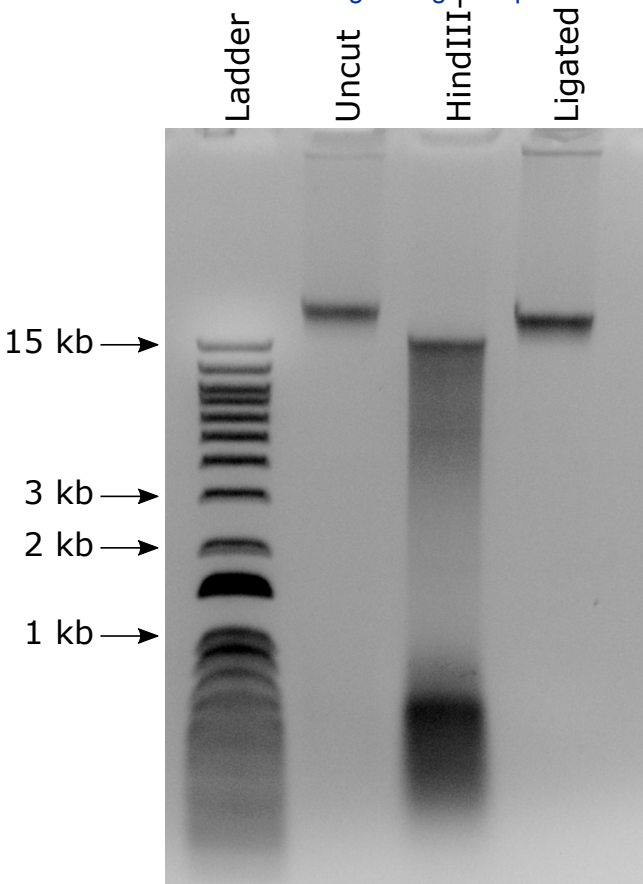
Figure 3

Figure 4

12.5 ng 25 ng 50 ng 100 ng NTC Ladder

← 2 kb

← 1 kb

Figure 5

Amplification Plot
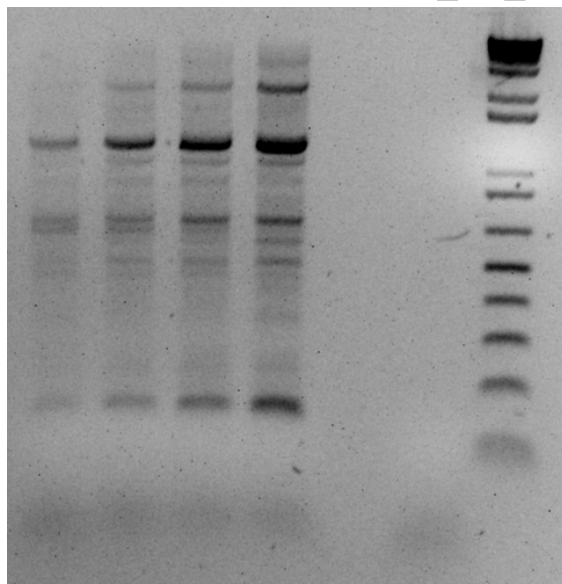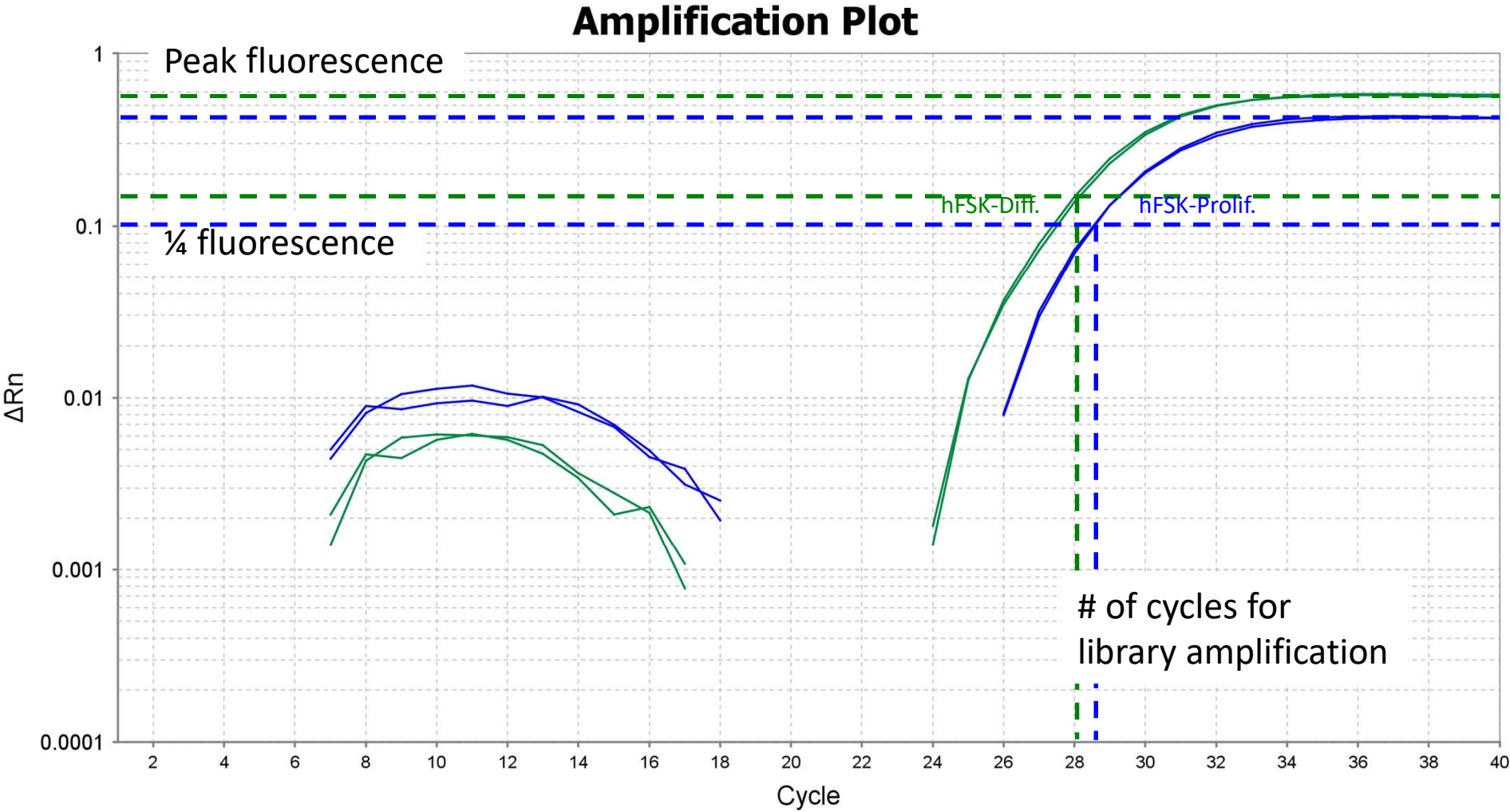
Figure 6

Uncut monitor

CviQI-digested monitor

Ladder
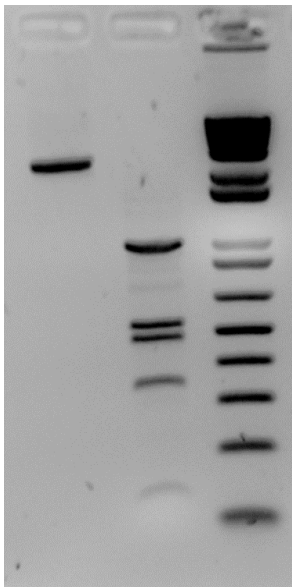
← 2 kb

← 1 kb

← 500 bp

← 300 bp

← 100 bp

Figure 7

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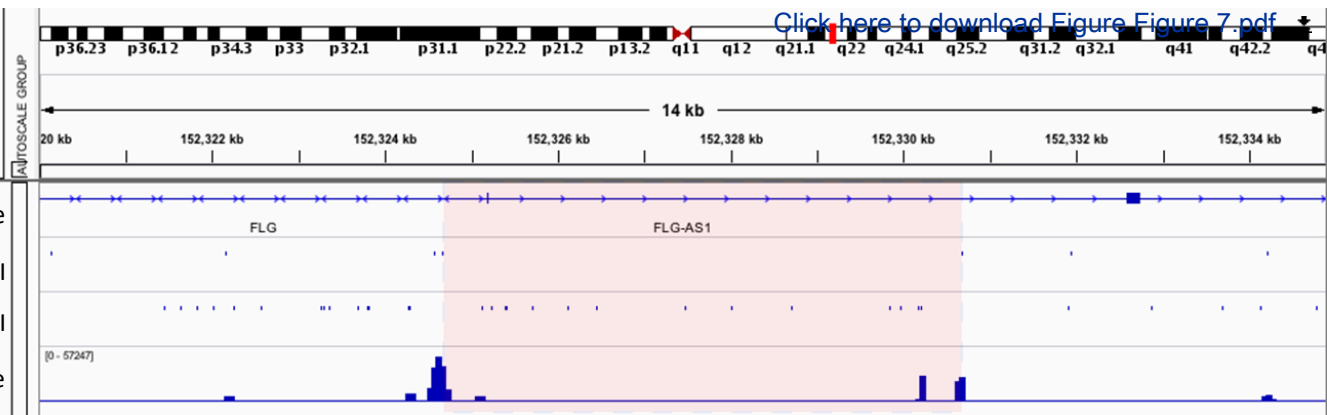
	Volume for 1x
10x Expand Long TemplateBuffer 1	2.5 µL
dNTPs (10 mM)	0.5 µL
Forward primer	35 pmol
Reverse primer	35 pmol
Expand Long Template Polymerase (5U/µL)	0.35 µL
DNA	
Nuclease-free water	to 25 µL

Table 2

	Volume for 1x
10x Expand Long TemplateBuffer 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 (5U/µL)	0.21 µL
DNA	100 ng
Nuclease-free water	to 25 µL

Table 3

	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 (5U/µL)	11.2 µL
DNA	3.2 µg
Nuclease-free water	to 800 µL

Table 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 5

	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

Name of Material/ Equipment	Company	Catalog Number	Comments/Description
HindIII	NEB	R0104S	
CviQI	NEB	R0639S	
DNA oligonucleotide primers	IDT		To be designed by the reader
50 mL conical centrifuge tubes	Fisher Scientific	06-443-19	
1.7 mL microcentrifuge tubes	MidSci	AVSS1700	
Phosphate buffered saline	Thermo Fisher	14190-136	
Formaldehyde, methanol free	Electron Microscopy Sciences	15710	
Nutator	VWR	15172-203	
Glycine	JT Baker	4059-00	
Benchtop centrifuge			
Refrigerated microcentrifuge			
Ethylenediaminetetraacetic acid (EDTA)	Sigma-Aldrich	ED2SS	
20% SDS solution	Sigma-Aldrich	05030	
Trypan Blue	Thermo Fisher	15250061	
Glass slides	Fisher Scientific	12-550-143	
Cover slips	VWR	16004-094	
Light microscope			
Triton X-100	Alfa Aesar	A16046	
Shaking heat block			
2M Tris-HCl	Quality Biological	351-048-101	
Proteinase K	NEB	P8107S	
Phenol:chloroform:isoamyl alcohol (25:24:1)	Sigma-Aldrich	P2069	
Sodium acetate	Sigma-Aldrich	M5661	
20 mg/mL glycogen	Thermo Fisher	R0561	
Ethanol	Fisher Scientific	04-355-223	
Nuclease-free water	Fisher Scientific	MT-46-000-CM	
qPCR cycler	Thermo Fisher	4453536	
qPCR plates	Thermo Fisher	4309849	
Thermocycler	Thermo Fisher	4375786	
PCR strip tubes	MidSci	AVSST-FL	

1M Magnesium chloride	Quality Biological	351-033-721
Dithiothreotol	Sigma-Aldrich	43815
Adenosine triphosphate	Sigma-Aldrich	A2383
T4 DNA Ligase	NEB	M0202S
Agarose	Sigma-Aldrich	A6013
RNase A	Thermo Fisher	EN0531
Qiaquick PCR purification kit	Qiagen	28104
MinElute PCR Purification kit	Qiagen	28004
Spectrophotometer		
Expand Long Template PCR System	Sigma-Aldrich	11681834001
dNTP mix	Thermo Fisher	R0191
SYBR Green I	Sigma-Aldrich	S9430
ROX	BioRad	172-5858
Sodium chloride	Sigma-Aldrich	S5886
End-It DNA End-Repair Kit	Lucigen	ER0720
LigaFast Rapid DNA Ligation System	Promega	M8221
SYBR Safe	Thermo Fisher	S33102
Taq Polymerase	NEB	M0267S
UltraSieve Agarose	IBI Scientific	IB70054
Qiaquick Gel Extraction Kit	Qiagen	28704

jove
JOURNAL OF
VISUALIZED EXPERIMENTS

1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article: High throughput identification of gene regulatory seqeunces by circular chromosome conformation capture coupled with next-generation sequencing

Author(s): Erin A. Brettmann, Inez Y. Oh, and Cristina de Guzman Strong

Item 1 (check one box): The Author elects to have the Materials be made available (as described at http://www.jove.com/author) via: [X] Standard Access ☐ Open Access

Item 2 (check one box):

[X] The Author is NOT a United States government employee.

☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. Defined Terms. As used in this Article and Video License Agreement, the following terms shall have the following meanings: "**Agreement**" means this Article and Video License Agreement; "**Article**" means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; "**Author**" means the author who is a signatory to this Agreement; "**Collective Work**" means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; "**CRC License**" means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode; "**Derivative Work**" means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; "**Institution**" means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; "**JoVE**" means MyJove Corporation, a Massachusetts corporation and the publisher of *The Journal of Visualized Experiments;* "**Materials**" means the Article and / or the Video; "**Parties**" means the Author and JoVE; "**Video**" means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. Background. The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. Grant of Rights in Article. In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4** and **7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

612542.6

4. Retention of Rights in Article. Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. Grant of Rights in Video – Standard Access. This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. Grant of Rights in Video – Open Access. This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this Section 6 is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. Government Employees. If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. Likeness, Privacy, Personality. The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

9. Author Warranties. The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

10. JoVE Discretion. If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have

ARTICLE AND VIDEO LICENSE AGREEMENT

full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

11. Indemnification. The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in *JoVE* or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's
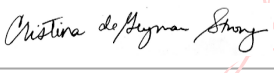
expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

12. Fees. To cover the cost incurred for publication, JoVE must receive payment before production and publication the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

13. Transfer, Governing Law. This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to me one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement required per submission.

CORRESPONDING AUTHOR:

Name: Cristina de Guzman Strong, PhD

Department: Medicine

Institution: Washington University School of Medicine

Article Title: High throughput identification of gene regulatory seqeunces by circular chromosome conformation capture coupled with next-generation sequencing

Signature: *Cristina de Guzman Strong*
Digitally signed by Cristina de Guzman Strong
DN: cn=Cristina de Guzman Strong, o, ou, email=cstrong@dom.wustl.edu, c=US
Date: 2018.02.08 17:19:55 -06'00'

Date: 2/8/2018

Please submit a signed and dated copy of this license by one of the following three methods:
1) Upload a scanned copy of the document as a pfd on the JoVE submission site;
2) Fax the document to +1.866.381.2236;
3) Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02139

For questions, please email submissions@jove.com or call +1.617.945.9051

612542.6

Editorial Comments:

1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues.

We have proofread the manuscript for spelling and grammar errors.

2. JoVE cannot publish manuscripts containing commercial language. This includes company names of an instrument or reagent. Please remove all commercial language from your manuscript and use generic terms instead. All commercial products should be sufficiently referenced in the Table of Materials and Reagents. Examples of commercial language in your manuscript include Qiagen, SyberSafe, AMPure XP, Qiaquick, TapeStation, BioAnalyzer, HiSeq, etc.

These were added at the request of reviewers. We have replaced specific products with more generic terms.

3. Please do not highlight notes for filming.

We removed highlights from notes.

4. Please define all abbreviations before use, e.g., EDTA, etc.

We have defined abbreviations at first use.

5. Step 6.1: What's the composition of restriction enzyme buffer?

The composition of the restriction enzyme buffer will depend on the enzyme selected by the reader. This has been clarified in the text (as well as in step 9.1).

6. 10.2: What's the composition of ligation buffer?

We added the composition of the ligation buffer to the text in step 10.2.

7. Line 361-366: Please use complete sentences instead of a chart to describe the reaction condition.

We have converted the chart to complete sentences to describe the PCR cycling conditions.

8. Line 384-389: Please use complete sentences instead of a chart to describe the reaction condition.

We have converted the chart to complete sentences to describe the PCR cycling conditions.

9. Line 399-403: Please use complete sentences instead of a chart to describe the reaction condition.

We have converted the chart to complete sentences to describe the PCR cycling conditions.

10. Where is step 12.1?

We combined steps 12.1.1 and 12.1.2 into a single step 12.1.

11. Line 492-495: Please use complete sentences instead of a chart to describe the reaction condition.

We have converted the chart to complete sentences to describe the PCR cycling conditions.

12. Line 500-504: Please use complete sentences instead of a chart to describe the reaction condition.

We have converted the chart to complete sentences to describe the PCR cycling conditions.