

# Protocol: Sonication-based Circular Chromosome Conformation Capture with next-generation sequencing analysis for the detection of chromatin interactions

Melissa Jane Fullwood (✉ [mjfullwood@gmail.com](mailto:mjfullwood@gmail.com))

Genome Technology and Biology, Genome Institute of Singapore

Yijun Ruan (✉ [ruanyj@gis.a-star.edu.sg](mailto:ruanyj@gis.a-star.edu.sg))

Genome Technology and Biology, Genome Institute of Singapore

Phillips Y.H. Huang

Genome Technology and Biology, Genome Institute of Singapore

Yuyuan Han

Genome Technology and Biology, Genome Institute of Singapore

Lusy Handoko

Genome Technology and Biology, Genome Institute of Singapore

Stoyan Velkov

Genome Technology and Biology, Genome Institute of Singapore

Eleanor Wong

Genome Technology and Biology, Genome Institute of Singapore

Edwin Cheung

Cancer Biology and Pharmacology, Genome Institute of Singapore

Xiaoan Ruan

Genome Technology and Biology, Genome Institute of Singapore

Chia-Lin Wei

Genome Technology and Biology, Genome Institute of Singapore

---

## Method Article

**Keywords:** Chromatin interactions, DNA sequencing

**Posted Date:** December 14th, 2010

**DOI:** <https://doi.org/10.1038/protex.2010.207>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)



# Abstract

The three-dimensional organization of chromatin in the nuclear space is involved in regulation of gene expression. Circular Chromosome Conformation Capture (4C) is an established method for genome-wide screening of chromatin interactions associated with a given locus of interest, without prior knowledge of the identity of interacting partners. Briefly, 4C involves the cross-linking of chromatin material, followed by restriction enzyme digestion of the chromatin, proximity-based ligation of interacting DNA fragments within the same DNA-protein complex, and amplification of interacting sequences by inverse PCR. The use of restriction enzyme digestion together with PCR could lead to biases in detection as well as false positives through repeated identification of clonal products. Here, we present a modification of the original 4C method, in which sonication is used to randomly fragment chromatin fibers instead of restriction enzymes at specific sites, to eliminate these biases, thus enabling high-throughput analysis by next-generation sequencing to detect interacting sequences. P.Y.H.H. and Y.H. made equal and critical contributions to this manuscript.

## Introduction

The three-dimensional organization of chromatin in the nuclear space is considered to be a critical contributor to the regulation of gene expression. The formation of chromatin loops, for instance, allows distal cis-enhancer elements to communicate with gene promoters and thereby up-regulate or down-regulate transcription [1][2]. High-resolution analysis of chromatin organization *in vivo* was made possible by the invention of the Chromosome Conformation Capture (3C) method [3], which has been widely used to detect and quantify interactions between genomic loci of interest. In this assay, intact cells are treated with formaldehyde to cross-link chromatin segments that are in close proximity. Cross-linked chromatin is then digested by a restriction enzyme and ligated under dilute conditions that favor the formation of junctions between interacting DNA fragments. Finally, the cross-links are reversed and site-specific PCR primers are used to quantitatively detect ligation events between selected pairs of restriction fragments. However, as the 3C technique assumes prior knowledge of interacting sequences, it cannot be used to screen for unknown interactions. To address this limitation, the Circular Chromosome Conformation Capture (4C) approach was developed [4][5][6][7], enabling genome-wide interrogation of DNA segments associated with a given locus of interest, also known as the “bait” region. A typical 4C protocol operates on the same principles pioneered by 3C: formaldehyde cross-linking is carried out to capture DNA-protein interactions, and restriction digest generates DNA fragments which are then subjected to proximity ligation. Unlike in 3C, however, the generation of circular DNA during ligation is central to the 4C strategy. These circularized molecules serve as the template for inverse PCR using bait-specific primers that are strategically positioned to amplify interacting DNA sequences flanking the target region. The amplified sequences are subsequently identified either by large-scale sequencing or by microarray analysis. In 2009, Chromatin Interaction Analysis with Paired-End Tags (ChIA-PET) was developed for high-throughput, global, *de novo* detection of chromatin interactions without the requirement for a specific bait region [8][9]. Briefly, chromatin immunoprecipitation is performed on

sonicated, cross-linked chromatin to reduce the complexity of the library and allow investigation of specific transcription factors. The chromatin is subjected to proximity ligation, reverse cross-linked, and sequenced using next-generation sequencing methods. In addition, Hi-C was developed for global detection of chromatin interactions [10]. Briefly, cross-linked chromatin is digested by restriction enzymes, subjected to proximity ligation, reverse cross-linked, and sequenced using next-generation sequencing methods. Blocks of DNA are then examined to determine the proximity of these regions of the DNA. As these methods are high-throughput, 4C methods can complement them through serving as validation studies, and in-depth analyses of specific regions. While the 4C protocol in the general form described above has been used in several studies [4][5][6][7], there are nonetheless certain technical limitations. A serious shortcoming of the original 4C technique lies in its inability to provide an accurate assessment of interaction frequencies and the possibility of false positives. It is important to recognize that the use of multiple rounds of PCR, both to select for bait-associated interactions and to generate sufficient DNA for sequencing or microarray hybridization, inevitably leads to massive clonal amplification of interacting DNA sequences. This could lead to false positives or negatives that would complicate the interpretation of sequencing or microarray data, making it difficult to determine true interaction frequencies. Generally, multiple independent rounds of 4C on different biological samples which show repeated PCR products has been accepted to indicate that the chromatin interactions are *bona fide*; however, if a particular DNA sequence is always enriched because of PCR preferential clonal amplification biases, the chromatin interaction could nonetheless be a false positive. Another major issue is the potential for bias associated with the use of restriction digest to fragment chromatin material, as performed in many 3C, 4C, and Hi-C protocols [10][11][12]. Certain chromatin regions – for example, transcriptionally active sites containing fewer histone proteins – may be relatively more accessible to endonucleases and hence be preferentially digested. Cross-linking stringency has been shown to be inversely related to restriction digest efficiency [13]; furthermore, long fragments arising from incomplete digestion are selected against during PCR amplification [14]. As a consequence, interacting sequences in regions with high cross-linking efficiency may be under-represented. The uneven distribution of restriction enzyme recognition sites may also contribute to bias, as different interacting segments may be over- or under-represented depending on the frequency of restriction sites in their respective genomic regions [14][15]. Moreover, some restriction enzymes may perform poorly in the presence of Sodium dodecyl sulphate (SDS) and Triton X-100, both of which are used in the 4C technique to prevent aggregation of nuclei and to open up chromatin for restriction enzyme digestion [14]. The clonal amplification issue which could lead to high false positives has to a certain extent, been ameliorated in the microarray-based detection method. 4C microarray approaches work by normalizing PCR-amplified 4C data against PCR-amplified genomic background, thus normalizing clonal amplifications; followed by applying a running mean algorithm to define clusters of increased hybridization signals relative to the surrounding genomic area [4][16], hence enabling the detection of clusters indicating chromatin interactions. However, microarrays have a limited dynamic range and poor coverage of repetitive regions. Sequencing offers many advantages over microarrays: within reasonable cost limits, the dynamic range can be expanded to suit the needs of each experiment simply by increasing the total number of sequencing reads. Also, sequencing allows for direct counting. However, previous 4C studies that made use of the sequencing

approach failed to exclude the possibility of sampling error as their analyses were focused on relatively small numbers of clones; moreover such studies would not have been able to eliminate the clonal amplification biases [5][6][7]. To overcome these issues and develop an unbiased 4C sequencing-based assay, we have developed a modified 4C protocol which uses sonication instead of restriction digest to fragment chromatin DNA. Sonication has the benefit of eliminating any potential problems with restriction enzyme bias, as its acoustic-based physical shearing mechanism disrupts DNA in a random manner, generating fragments a few hundred base pairs in length with a random distribution of breakpoints. Hence, when two interacting DNA fragments are joined together, their breakpoints form a ligation junction that is defined by a unique set of genomic coordinates. Sequencing across ligation junctions is thus a potentially powerful means of identifying unique sequences from independent ligation events. The presence of multiple unique sequences clustered around the same locus would then indicate a likely interaction. Coupled with high-throughput next-generation sequencing, this strategy enables the identification of multiple unique sequences, indicating possible chromatin interactions. **Overview of the sonication-based 4C technique** The 4C protocol described here is a genome-wide and unbiased approach for the *de novo* detection of chromatin interaction targets with a particular bait. 4C makes use of the proximity ligation concept, pioneered by the 3C method, to capture interacting DNA segments within DNA-protein complexes. As opposed to the original 4C strategy, our new 4C strategy uses sonication, as opposed to restriction enzyme digestion, to fragment chromatin fibers, allowing for the identification of partners in an unbiased manner by next generation sequencing (Figure 1). **Sonication-based 4C experimental design** *Site selection and inverse primer design.* Sites for 4C analysis were selected based on ChIA-PET data, but any non-repetitive site may be used for analysis. Primers were designed using Primer3 software (<http://frodo.wi.mit.edu/primer3/>) [17]. The RepeatMasker track in the UCSC Genome Browser was used to ensure that the primers did not lie in repeat regions (<http://genome.ucsc.edu/>) [18]. To ensure specificity, primer sequences were analyzed by BLAT (<http://genome.ucsc.edu/cgi-bin/hgBlat>) [19], and only unique primer sequences were used. Flanking inverse primers should be around 25 bp long for high specificity, which is critical for the first round of inverse PCR as it selects for the products that would be amplified by nested PCR. Nested inverse primers will need to be designed with the 454 adaptor sequences attached onto the 5' ends, such that the 454 adaptors can be incorporated into the PCR products. Hence, nested primers can be approximately 20 bp in size to reduce the cost and length of the final oligonucleotides used for nested PCR. Also, flanking inverse primers should be no more than 100 bp apart, and nested primers should be designed to be as close to the inverse primers as possible, because the probability of a randomly sheared DNA fragment containing the sequences necessary for successful priming and PCR amplification is inversely related to the genomic distance spanned by the primers. Hence, to amplify a larger and more diverse population of DNA fragments, the primers should be designed relatively close together. *Fragmentation of chromatin fibers* Sonication-based 4C uses sonication instead of restriction enzyme digestion to fragment chromatin fibers. The advantages are that more regions of the genome can be interrogated than with restriction enzyme digestion, and also, unique end sequences are generated as opposed to non-unique restriction enzyme digestion ends. Unique end sequences allow clonal amplifications to be identified and removed. The unique tags can then be clustered to identify *bona fide* chromatin interactions.

**\_Sequencing of 4C material\_** 4C material may be analyzed by 454 Titanium next-generation sequencing. 4C material consists of bait-target-bait structures, and long read lengths will enable read-through past the bait into the target sequences. While we used the 250 bp read length in the experiments described here, the use of 400 bp in the new Titanium system would allow for better read-throughs of the sequences. In its present form, sonication-based 4C cannot be used with Illumina or ABI SoLiD next-generation sequencing methods.

**\_Data analysis\_** Data analysis is then required to identify putative interactions. First, the sequences are mapped to the genome. BLAT or BLAST may be used. As the chromatin is sonicated, the probability of generating exactly identical DNA fragments is low; hence any redundant sequences are considered to be copies amplified during the cloning and/or PCR amplification processes. Therefore, only nonredundant distinct sequences are used for further analysis. Next, the “multiple overlaps” concept is used to distinguish true signals from noise. The principle of this concept is that we expect PETs derived from nonspecific fragments to be randomly distributed in the genome as background sequences, whereas interacting sequences derived from the same *\_bona fide\_* interactions will overlap with each other to form a cluster of interacting sequences.

**\_Validations\_** These can be performed using the 3C method, as well as FISH, to confirm whether there is an interaction. FISH is particularly useful as an orthogonal validation method that employs very different techniques from 4C to visualize the interaction. FISH probes may also then be used to study the interaction in clinical samples that generally involve very small amounts of cells. It should be noted that because FISH is limited by low resolution, it can only be performed on interactions that exceed 1 Mb.

**\*\*Applications of the 4C method\*\*** This protocol may be used to interrogate chromatin interactions that interact with a genomic region of interest, and can serve as a validation method for ChIA-PET, 5C, and other genome-wide analyses. With 4C, targeted questions may be asked in specific genome regions, for example, in analyses of the keratin gene cluster. Our analyses of the keratin gene cluster suggest that keratin genes may be brought together by chromatin interactions for coordination of transcription. Moreover, 4C may be combined with ChIP in order to interrogate chromatin interactions bound by particular proteins.

## Reagents

- 454 Titanium sequencing kit \ (Roche)
- Designed, HPLC-purified primers \ (IDT) \ (Sequence listing in Appendix V)
- End-It kit \ (Epicentre, ER81050)
- Plasmid-Safe DNase \ (Epicentre, E3101K)
- Formaldehyde, 37% \ (Sigma) \ **! CAUTION:** Formaldehyde is toxic and corrosive. It may be harmful by inhalation, ingestion, or skin absorption. Wear gloves and eye protection; avoid skin contact and the inhalation of fumes. Use in a chemical fume hood. Dispose of formaldehyde according to regulations on hazardous chemical waste.
- Protease Inhibitor Complete EDTA-free \ (Roche, 04 693 132 001)
- Trypsin \ (Invitrogen)
- DMEM phenol-red medium, includes 4500mg/L Glucose, 110mg/L sodium pyruvate, \ (pH6.8 ±0.3 ) \ (Invitrogen)
- Clear, phenol-red free DMEM/F12 media, includes L-Glutamine and 15mM HEPES \ (Invitrogen)
- Non-heat-inactivated Fetal Bovine Serum \ (FBS) \ (Invitrogen)
- Charcoal-Dextran-stripped Fetal Bovine Serum \ (CD-FBS) \ (Hyclone)
- Glycine \ (Invitrogen/Gibco)
- 17 beta-estradiol \ (E2) \ (Sigma)
- Penicillin/Streptomycin \ (Invitrogen)
- Gentamycin \ (Invitrogen)
- L-Glutamine \ (Invitrogen)
- 10x Phosphate Buffered Saline \ (PBS) \ (1st Base)
- 10x T4 DNA Ligase Buffer \ (NEB, B0202S) \ **! CAUTION:**

Ligase buffer contains dithiothreitol (DTT), a strong reducing agent that emits a foul odor. It may be harmful by inhalation, ingestion, or skin absorption. **! CRITICAL:** DTT may be oxidised over time. Use new ligase buffer if reagent is too old.

- T4 DNA ligase (30U/μl) (Fermentas, EL0017)
- Proteinase K Solution (~20 mg/ml), 1 ml (Fermentas, E00491)
- 6x Loading Dye (Fermentas, R0611)
- 25 bp DNA ladder (Invitrogen, 10597-011)
- Low Mass DNA ladder (Invitrogen, 10068-013)
- 6% Tris-Borate-EDTA (TBE) Polyacrylamide gel electrophoresis (PAGE) gel (5 wells) (Invitrogen, EC6264BOX)
- SYBR® Green I (Molecular Probes) (Invitrogen, S-7585) **! CAUTION:** Although SYBR Green is considered to be a safer replacement for ethidium bromide, it binds to DNA with high affinity and thus is potentially carcinogenic. Wear gloves and avoid skin contact; dispose of SYBR Green-containing solutions in appropriate waste containers.
- Phusion™ High-Fidelity PCR Master Mix with HF buffer (Finnzymes, F-531)
- Buffer EB (250 ml) (Qiagen, 19086)
- 10% Sodium dodecyl sulphate (SDS) (500 ml) (1st Base, 2051-500ml) **! CAUTION:** SDS may be harmful by inhalation or ingestion.
- Triton X-100 (Sigma, T8787)
- Tris-HCl (pH8.1, Fisher Scientific)
- 0.5M EDTA pH 8.0 (100 ml) (Ambion, AM9260G)
- 3M Sodium acetate pH 5.2 (100 ml) (Ambion, AM9740)
- 5M Sodium chloride (100 ml) (Ambion, AM9760G)
- Nuclease-free water (50 ml) (Ambion, AM9937)
- GlycoBlue (15 mg/ml), (Ambion, AM9516)
- Phenol:Chloroform:IAA, 25:24:1, pH 7.9, 100 ml (Ambion, AM9730) **! CAUTION:** Phenol-chloroform is highly toxic and corrosive. Wear gloves and eye protection; avoid skin contact and the inhalation of fumes. Use in a chemical fume hood. Dispose of phenol-chloroform according to regulations on hazardous chemical waste. **! CRITICAL:** DNA partitioning is pH dependent; at pH 7.0 or higher, both DNA and RNA partition into the aqueous phase. At an acidic pH, below pH 7.0, DNA will be denatured and partition into the organic phase and interphase, leaving the RNA alone in the aqueous phase. Hence, it is necessary to adjust the pH of all new phenol-chloroform bottles by adding the included buffer according to manufacturer's instructions. **! CRITICAL:** Store phenol-chloroform at 4°C in the dark. Storage at room temperature and exposure to light may cause oxidation of phenol-chloroform. In such cases, phenol-chloroform samples must be discarded as they can lead to cleavage of DNA in solution.
- Buffer Tris-EDTA (TE) (pH8.0), 500 ml (Ambion, AM9849)
- 10x Tris-Borate-EDTA (TBE) buffer (1st Base, 3010-10x1L)
- Isopropanol (Sigma, I-9516-500ml) **! CAUTION:** Highly flammable. Handle absolute ethanol in a chemical fume hood.
- Analytical reagent grade absolute ethanol (Fisher Scientific) **! CAUTION:** Highly flammable. Handle absolute ethanol in a chemical fume hood.
- Quant-iT™ PicoGreen® dsDNA reagent (10 x 100μl) (Invitrogen, P11495)
- DNA from calf thymus (genomic, unsharded) (Sigma, D4764)
- Agilent DNA 1000 Reagents (Agilent Technologies, 5607-1505)
- Agilent DNA 1000 Kit (Agilent Technologies, 5607-1504)

**OTHER CONSUMABLES**

- IWAKI flat bottom polystyrene 96-well microtiter plate
- Microseal 96-well PCR plate (Bio-Rad)
- Eppendorf Combitips plus 0.2 ml
- Eppendorf Combitips plus 0.1 ml
- BD Falcon polypropylene conical tubes (15-ml) (Becton Dickinson, 352097)
- BD Falcon polypropylene conical tubes (50-ml) (Becton Dickinson, 352070)
- 1.7-ml tubes (500) (Axygen)
- DNA LoBind Tubes, 1.5-ml PCR clean (Eppendorf, 0030 108.051)
- 0.6-ml tubes (500) (Axygen)
- 0.2-ml PCR tubes (Axygen)
- Spin X columns (Corning)
- 50-ml Oak Ridge PPCO centrifuge tubes (Nalgene, 3119-0050)
- Gel handler, 10sheets/pk (Sigma, Z376957-1PAK)
- Phase Lock Gel light tubes (Eppendorf)
- MaXtract High Density, 25x50ml (Qiagen, 129073)
- MaXtract High Density, 100x15ml (Qiagen, 129065)
- T175 Cell culture flasks (Biomed Diagnostics)
- 150 mm diameter cell culture plates (Biomed

Diagnostics) • Blade scrapers \ (Corning) • 21G needle \ (Becton Dickinson) • Stainless steel sterile surgical blades \ (Myco Medical Supplies Inc.) \! CAUTION: Handle needles and blades with care, and dispose of them in the sharps bin. REAGENT SETUP • Growth media for MCF-7 cells Phenol-Red DMEM media supplemented with 5% FBS, 1% penicillin/streptomycin, and 0.3% gentamycin \! CRITICAL: Ensure that growth media is not exposed to any bacterial or fungal contamination. Store at 4°C. • Starvation media for MCF-7 cells Clear, phenol-red free DMEM/F12 media supplemented with 5% CD-FBS, 1% penicillin/streptomycin, and 0.3% gentamycin \! CRITICAL: Ensure that growth media is not exposed to any bacterial or fungal contamination. Store at 4°C. • Triton-X lysis buffer Final concentration of 0.25% Triton X-100, 10 mM EDTA, 10 mM Tris HCl \ (pH 8.1), 100 mM NaCl, supplemented with 1x EDTA-free Protease Inhibitor. \! CRITICAL: Add protease inhibitor immediately prior to use. • SDS lysis buffer Final concentration: 1% SDS, 5 mM EDTA, 50 mM Tris HCl, \ (pH 8.1), supplemented with 1x EDTA-free Protease Inhibitor \! CRITICAL: Add protease inhibitor immediately prior to use.

## Equipment

• Centrifuges - Eppendorf 5415 R microcentrifuge \ (for 4°C) - Eppendorf 5415 D microcentrifuge \ (for room temperature) - Eppendorf 5810 R - Sorvall RC 5C Plus; SS-34 rotor • Belly Dancer \ (Stovall) • Stuart Roller Mixer SRT9 • Agilent 2100 Bioanalyzer \ (Agilent Technologies) • Bioruptor \ (Diagenode) \! CAUTION: Sonication produces high-frequency sounds which can damage hearing. Wear ear protection when operating sonicator. • Novex Mini-Cell \ (Invitrogen) • TECAN GENios Automated Microplate reader • Mastercycler EP for PCR reactions \ (Eppendorf) • MJ Thermocycler \ (MJ Research) • Incubator \ (Mettler) • Cell culture humidified CO<sub>2</sub> incubator \ (Sanyo Air jacket CO<sub>2</sub> incubator) • Cell culture hood \ (Nuaire) • Laminar flow hood \ (Nuaire) • Chemical Fume hood \ (Kewaunee) • ND-1000 Spectrophotometer \ (Nanodrop) • DarkReader Transilluminator \ (Clare Chemical Research) • Roche 454 GS-FLX \ (454 Life Sciences) • Microscope \ (Nikon Eclipse TS100)

## Procedure

**\*\*General points to note:\*\*** - Chromatin material and enzymes are sensitive to both proteases and nucleases. Care should be taken to avoid introducing these contaminants. Precautionary procedures such as swabbing down lab benches with 70% ethanol are advised, as well as the use of nuclease-free water. - The use of DNA Lo-Bind tubes \ (Eppendorf) is recommended to reduce loss of sample. - A general note for all enzymatic reactions: because excessive glycerol from the enzyme stock may interfere with the reaction, ensure the volume of enzyme is <10% of the final reaction mixture. The starting material is cells. The chromatin should be isolated from about 10<sup>6</sup> - 10<sup>7</sup> cells such that there will be sufficient chromatin material for library construction and the resulting library will be of high complexity. The cells may be treated with drugs as appropriate \ (Appendix I). The results shown came from a 4C library prepared from estrogen-treated MCF-7 cells. **\*\*A. Chromatin preparation\*\*** Timing: ~ 1-2 days 1. To treated MCF-7 human breast adenocarcinoma cells growing in a 150 mm diameter plate in 20 ml of media \ (Appendix I), add 540 µl of fresh formaldehyde \ (37% stock) to obtain a final concentration of 1% formaldehyde for



cross-linking. Rotate on a belly dancer for 10 min at 22°C. \! CRITICAL STEP: Fresh formaldehyde should be used. The amount and timing of formaldehyde cross-linking should be controlled and optimized, as too much formaldehyde cross-linking will make sonication difficult as well as increase the level of non-specific interactions, while too little formaldehyde would result in insufficient capture of chromatin interactions.

2. Add 2 ml of 2.5 M filter-sterilized glycine to stop the cross-linking. Rotate on a belly dancer for 5 min at 22°C. Pour away the medium into a waste flask.
3. Wash 2x with 10 ml of cold (4°C) 1x PBS with gentle shaking. Pour away the PBS into a waste flask.
4. Add 2 ml of 1x PBS and 2 ml of trypsin to each plate. Mix well and let the plates sit for 10 min at 22°C.
5. Add 4 ml of growth media to stop the trypsinization.
6. Use a blade scraper to scrape off the cells. Collect cells into a 15 ml falcon tube.
7. Wash the plates with 4 ml of 1x PBS and place these cells into the same 15 ml falcon tube.
8. Pellet cells at 3,000 rpm (800xg, Eppendorf) for 15 min at 4°C and remove the supernatant.
9. Wash the pellet with 5 ml of 1x PBS, then pellet cells at 3,000 rpm (800xg, Eppendorf) for 5 min at 4°C and remove the supernatant.
10. Resuspend and wash each pellet with 5 ml of Triton X-100 lysis buffer. Incubate at 4°C with gentle agitation (Stuart Roller Mixer SRT9) for 30 minutes.
11. Spin at 3,000 rpm (800xg, Eppendorf) for 5 min at 4°C and pour away supernatant. \! PAUSE POINT: The pellet may be stored for several months at -80°C.
12. Resuspend the pellet (nuclear lysate) with 600 µl SDS lysis buffer.
13. Sonicate the nuclear lysate for 8 min using the Bioruptor (Diagenode) operated in a 4°C cold room, 30 seconds on, 30 seconds off, at high power.
14. Pellet cell debris at 13,000 rpm (15700xg, Eppendorf) for 10 min at 4°C and transfer the supernatant (chromatin) into a fresh 1.7 ml tube. Store at -80°C. \! CRITICAL STEP: Ensure that the sonicator is balanced. Clean probes with 70% ethanol before and after sonication. Sonication of the chromatin should be done to a DNA size of approximately 500 bp with a range from 100-2000 bp, and be checked by reverse cross-linking an aliquot of 5 µl of nuclear lysate with 1 µl proteinase K (20 mg/ml) at 37°C for 1h and running on an 1% agarose gel to check the sizes of the DNA. Also, quantitate the reverse cross-linked DNA concentration with Picogreen Fluorimetry (Appendix II). \! PAUSE POINT: The chromatin may be stored for several months at -80°C.

**\*\*B. 4C library preparation\*\***

Timing: ~ 3-4 days

1. To a 125 µl aliquot (corresponding to approximately 10<sup>6</sup> cells) of sonicated, cross-linked chromatin containing 0.1% SDS, add 6.5 µl of 20% Triton-X. \! CRITICAL STEP: If taking aliquots out of -80°C, thaw chromatin samples gently on ice, and let them stand at 37°C for 1h to equilibrate. Clarify chromatin to remove any remaining cellular debris by spinning down chromatin in a microcentrifuge at maximum speed for 20 min at 4°C, and keeping the supernatant. Chromatin samples must contain 0.1% SDS – if samples contain no SDS, add SDS, or if samples contain too much SDS, then dilute down the sample in Buffer EB, mix well, and let chromatin samples stand at 37°C for 1h to equilibrate. The volume of the aliquot here is given for reference; the actual volume of the reference to be used should be calculated based on the DNA concentration. A maximum of 100 µg of DNA can be used.
2. For each sample, prepare the following End-blunting reaction mix: DNA 263 µl (this comes from two 125 µl aliquots of chromatin) 10x End-Repair Buffer 33 µl 2.5mM dNTP Mix 15 µl 10mM ATP 15 µl End-Repair Enzyme Mix 5 µl Total volume: 330 µl Incubate at room temperature for 45 min.
3. Set up the following ligation reaction: DNA 330 µl 10x T4 DNA Ligase Buffer (NEB) 5 ml T4 DNA Ligase (30u/µl, Fermentas) 166 µl Water (with 1x protease inhibitor) 44.5 ml Total volume: 50 ml Incubate at 16°C overnight (at least 16h). \! CRITICAL STEP: Dissolve protease inhibitor tablets in autoclaved water just

before use. For proximity ligation, the concentration of the DNA should not be higher than 2 ng/μl). 4. Perform reverse cross-linking by adding 375 μl of Proteinase K (20mg/ml, Invitrogen), mixing by inverting the tube several times and standing at 65°C overnight (at least 16h). 5. Split each ligation reaction into three 16.7 ml portions, and transfer each portion to a separate Maxtract High Density (50 ml) tube. Add an equal volume of phenol-chloroform and mix by inverting for about 2 min. Centrifuge at 1800 × g (3000 rpm), 5 min, room temperature (22-25°C). **CRITICAL STEP:** Pre-pellet the phase lock gel in the Maxtract tubes by centrifugation at 1800 × g (3000 rpm), 5 min, room temperature (22-25°C). 6. Transfer the upper aqueous phase to a 50-ml Nalgene PPCO tube. Perform isopropanol precipitation by adding: Sodium acetate (3 M, pH 5.2) 1.7 ml Glycoblu 6 μl Isopropanol 17 ml Total volume: 17.7 ml Incubate at -80°C for at least 1 h. 7. Pre-chill the Sorvall centrifuge to 4°C and pellet the DNA at 38730 × g (18000 rpm), 30 min, 4°C, using the SS-34 fixed rotor. 8. Carefully decant the supernatant and wash the pellet twice with 30 ml 75% ethanol. Allow the pellet to air dry in a laminar flow hood and re-suspend in 100 μl EB buffer. 9. Quantitate the DNA concentration using Picogreen Fluorimetry (Appendix II). **PAUSE POINT:** DNA may be stored for several months at -20°C. We recommend aliquots be stored at -20°C in case the downstream procedures fail for some reason. In this protocol, we assume half was stored. 10. Further purify the DNA by removing RNAs. DNA 49.5 μl RNase A (Qiagen, diluted to 2 μg/μl in EB buffer) 1 μl EB buffer 49.5 μl Total Volume: 100 μl Incubate at 37°C for 30 min. 11. Purify using phenol-chloroform extraction (Appendix III) followed by isopropanol precipitation (Appendix IV). Re-suspend each sample in 50 μl of EB buffer and quantitate using Picogreen (Appendix II). 12. Remove non-circular DNA by setting up the following reaction: DNA 49 μl 25 μM ATP 2 μl 10 × reaction buffer (provided with the plasmid-safe DNase kit) 6 μl Plasmid-safe DNase 1 μl Water 2 μl Total volume: 60 μl Incubate overnight at 37°C. 13. Inactivate plasmid-safe DNase by incubating the samples at 70°C for 20 min. Purify using phenol-chloroform extraction (Appendix III) followed by isopropanol precipitation (Appendix IV). Resuspend in 100 μl of EB buffer and quantitate using Picogreen (Appendix II). **\*\*C. 4C library amplification\*\*** Timing: ~ 2-3 days 1. Amplify the samples by nested inverse PCR (All PCR primers and adapters are listed in Appendix VI). For the first round of inverse PCR, prepare the following reaction mix: DNA 100 ng Inverse primer 1 (25 μM) 1 μl Inverse primer 2 (25 μM) 1 μl 2 × Phusion Master Mix 25 μl Water To 50 μl Total volume: 50 μl Cycle conditions are: 98 °C, 30 s 98 °C, 10 s 70 °C, 30 s 20 cycles 72 °C, 45 s 72 °C, 10 min; hold at 4 °C 2. To carry out nested PCR, prepare the following reaction mix: Undiluted product from inverse PCR 1 μl Nested, adaptor B-tailed primer 1 (biotinylated) (25 μM) 1 μl Nested, adaptor A-tailed primer 2 (25 μM) 1 μl 2 × Phusion Master Mix 25 μl Nuclease-free H<sub>2</sub>O 22 μl 50 μl Total volume: 50 μl Cycle conditions are: 98 °C, 30 s 98 °C, 10 s 72 °C, 30 s 20 cycles 72 °C, 1 min 72 °C, 10 min; hold at 4 °C **CRITICAL STEP:** Because the PCR step selects sequences for analysis, it is critical that PCR cycle conditions be optimized for every 4C experiment. The use of 2-step PCR (the first few cycles is done with reference to the melting temperature of the primer without the adaptor, and the next few cycles is done with reference to the melting temperature of the primer with the adaptor as the proportion of PCR products containing flanking adaptor sequences increases) might be useful for certain primers. 3. Run the PCR products in a 4-20% TBE PAGE gel at 200 V, 40 min in a Novex mini-cell. Stain with SYBR Green for 15 min in the dark and visualize on a DarkReader. We should observe a smear band across a range of approximately 200bp to >600bp. 4. Carry out additional nested PCR reactions using the products

obtained from the first round of inverse PCR. At least four 50  $\mu$ l reactions are recommended. 5. Run the PCR products in a 6% TBE PAGE gel for 30 min at 200V in a Novex mini-cell, together with 500 ng 25bp DNA ladder. Stain with SYBR Green for 15 min in the dark and visualize on a DarkReader. \! CRITICAL STEP: We have observed that certain DNA ladders that work well on agarose gel do not work on PAGE gels. We have tested Invitrogen's 25 bp DNA ladder and Low Mass ladder and found these to work well on PAGE gels. 6. Excise the smear band in the region of 400bp to >600bp, using a sharp scalpel and gel-handler sheets to protect the DarkReader. Take a photo of the gel before and after excision. Purify by the gel-crush protocol \ (Appendix V). Resuspend the pellet in 40  $\mu$ l TE buffer. 7. Quantitate DNA concentration using Picogreen \ (Appendix II). Perform a quality control check using an Agilent DNA 7500 LabChip according to the manufacturer's guidelines. 8. Prepare and sequence the sample with Roche 454 Titanium sequencing according to the manufacturer's guidelines.

## Timing

A: 1-2 days B: 3-4 days C: 2-3 days Total timing: 6-9 days

## Troubleshooting

Possible Problem 1: Quality Control run of annealed linkers yield two or more discrete bands. The reason might be that unequal molar amounts of oligonucleotides resulting in incomplete annealing. The solution is to test different ratios of oligos to determine the optimal ratio for stoichiometric annealing; if necessary, run an Agilent 1000 DNA chip to double check. Possible Problem 2: Quality Control run of PCR products shows a very weak smear/no smear. Possible reasons could include sub-optimal PCR cycle conditions, in which case solutions may include increasing the number of cycles used \ (do not use more than 25 cycles), decreasing annealing temperature or increasing elongation time, and another possible reason might be insufficient template DNA, in which case the solution is to increase the starting amount of template DNA in the PCR reaction. Possible Problem 3: Very bright smear/ doughnut-shaped band/ no band observed after PCR scale-up. Possible reasons may be that too much DNA was loaded onto the PAGE gel. Solutions include decreasing the number of PCR reactions, or splitting samples up into more wells. Possible Problem 4: Library has many repeated reads. Possible reasons may be that the library has low complexity. To minimize this problem, use high amounts of starting and template DNA to maximize the amount of DNA used in the PCR; reduce the number of PCR cycles. We have found that despite trying these options, many libraries still have many repeated reads, suggesting that chromatin interactions are inherently rare events. Possible Problem 5: Poor results observed upon sequencing and data analysis \ (few/ low quality chromatin interactions). One reason may be that there were problems with upstream chromatin preparation procedures; eg. poor cross-linking. Troubleshoot chromatin preparation procedures – ensure that formaldehyde used is fresh and functional, and ensure that sonication worked by running a quality control agarose gel. Another reason may be that the region of interest is repetitive. To troubleshoot, check to ensure that a region which is not highly repetitive \ (and hence difficult to analyze by sequencing followed by unique mapping) is not used. Possible Problem 6: Mapping errors are

observed (mapping is wrong upon manual double-checking of a few examples using UCSC BLAT and other mapping methods). This could be because the mapping was incorrectly done. To troubleshoot, ensure that only unique mappings are used to identify chromatin interactions.

## Anticipated Results

A successful library preparation would show the following successful quality controls: (1) Well-sonicated chromatin (Figure 2a); (2) A smear of about 200 to more than 500 bp in the quality control gel run after PCR (Figure 2b) (3) chromatin interactions following library sequencing (Figure 2c-e). In an experiment performed on a keratin gene cluster region, 454 GSFLX (a prior version to Titanium) sequencing generated approximately half a million sequences. All the sequences were mapped to reference genome (hg18 human genome assembly) to identify the target regions in relation to the bait region, and 0.44 million sequences (95%) showed at least one hit to the genome. Sequences that did not show at least two mapping regions were filtered away, as these could be incompletely sequenced ligation products or DNA sequences that did not ligate. 95,050 (21.6%) reads had at least 2 hits, with the first hit mapping correctly to the primer site. In further experiments, the longer sequencing read lengths offered by the new 454 Titanium system, as well as even longer ligation times, could address this issue. Sequences were filtered to remove redundant clonal amplifications (repeated sequences). Of 95,050 reads, 3660 (3.9%) sequences were found to be unique. Many redundant clonal amplification sequences were observed, indicating that clonal amplifications within 4C libraries are indeed an issue. Previous use of restriction enzymes to fragment the chromatin would not have been able to distinguish clonal amplifications from *bona fide* enriched chromatin interaction signals. In future experiments, further increasing the amount of starting template, and amount used as PCR template, could help to reduce the redundancy. Also, while reducing the number of PCR cycles could reduce amount of selection performed on the 4C library and hence increase non-specific 4C ligation noise, this modification might also reduce the amount of redundancy seen in the library. The majority of the non-redundant sequences either mapped randomly along the genome as potential non-specific 4C products, or mapped to the “bait” region within 1 kb, suggesting self-ligation products in the 4C experiment. 3,429 (93.7%) sequences were intrachromosomal ligation products, and they comprised 3388 (98.8%) self-ligation products, 30 (0.9%) inter-ligation products representing expected interactions, and 11 (0.3%) other long-range inter-ligations typical of random noise. Manual inspection of the 30 sequences that mapped to expected interaction loci revealed 7 unique ligation events, which were then collapsed into 4 distinct interactions. 2 interaction clusters (2 or more overlapping unique PETs) were found, namely chr12:50828808-50880947 (genomic span=52 kb; 2 ligation events) and chr12:50828801-50883887 (genomic span=54 kb; 3 ligation events). The remaining 2 interactions – chr12:50828832-51024898, with a genomic span of 196 kb, and chr12:50828839-51575465, with a genomic span of 746.6 kb – were represented by single unique ligation events (Figure 2c-e). We noted that the 4C data showed a very clean background (Figure 2d, e). From the bait region to the first and the second interaction sites (about 200 kb distance), there are no background sequences in the intervals. Given that we prepared the chromatin material used for our 4C analysis by sonication, which is different from the standard 3C and 4C protocols, this result suggests that

the sonication method could be very efficient in “shaking off” non-specific chromatin fragments randomly attached to the specific chromatin interaction complexes. We expect to obtain discrete interaction peaks formed by clusters of inter-ligation sequences from sonicated material, because we expect real interactions to be captured by proximity ligation process. While the detached non-specific chromatin fragments would still be present in the DNA pool, they would not be amplified by the 4C PCR detection method. We specifically looked at the KRT gene cluster site (previously identified by ChIA-PET [9]), where the “bait” region lay, from which we designed our inverse 4C PCR primers. Moving right from the “bait” region, we identified overlapping sequence clusters that correlated very well with the locations of the interaction sites identified by ChIA-PET data, cross-validating both the ChIA-PET data as well as the 4C protocol (Figure 2c, e). Interestingly, analysis of chromatin interactions by 4C and ChIA-PET in the keratin region suggests that chromatin interactions are correlated with gene expression coordination. Both ChIA-PET and 4C data shows that KRT7, KRT8, and KRT18 are all pulled into the “hub” of the same interaction complex. KRT7, 8, and 18 are known to be expressed in breast carcinomas. In particular, KRT8 and KRT18 are tightly coexpressed genes, and the gene products bind tightly to each other, pairing up by the formation of a heterodimer between KRT8 (a “type II” keratin) and KRT18 (a “type I” keratin). Without the formation of a heterodimer, type I and type II keratins are rapidly degraded [20]. These two genes are connected by many inter-ligations. By contrast, KRT5, 6, 1, 2, and other keratins involved in other aspects such as in hair development for example KRT72 and KRT75, are not expressed, and they are present in the “loop” of the interaction complex. Hence, chromatin interactions in the keratin region may bring together relevant genes into transcriptional foci, and loop out irrelevant genes, in order to achieve tightly coordinated gene expression regulation. In conclusion, our novel sonication-based 4C protocol has enabled the identification of bona fide chromatin interactions for ChIA-PET validation of chromatin interactions in the keratin gene cluster, demonstrating that chromatin interactions in the keratin cluster may function to coordinate gene transcription. With sonication, non-specific noise could be “shook off”, thus reducing the very high non-specific noise seen in the original 3C and 4C protocols. Moreover, with sonication as opposed to restriction enzyme digestion, a previously unrecognized problem with regards to high amounts of sequenced clonal amplifications may be reduced. With further optimizations, sonication-based 4C could become a robust method for use in conjunction with next-generation sequencing to identify and study chromatin interactions. Moreover, sonication-based 4C could be coupled with faster and cheaper third-generation sequencing methods such as Pacific Biosciences [21] as they become available. With the improved throughput, reduced sample requirements of the new third-generation sequencing methods, it may become possible to omit part C of the sonication-based 4C protocol presented here, and simply sequence the entire library of proximity-ligated sonicated chromatin, which would allow for ultra-high-throughput analysis of global chromatin interactions at high resolution.

## References

1. de Laat W, Klous P, Kooren J, Noordermeer D, Palstra RJ, et al. (2008) Three-dimensional organization of gene expression in erythroid cells. *Curr Top Dev Biol* 82: 117-139.
2. Simonis M, de Laat W (2008) FISH-eyed and genome-wide views on the spatial organisation of gene expression. *Biochim Biophys Acta*

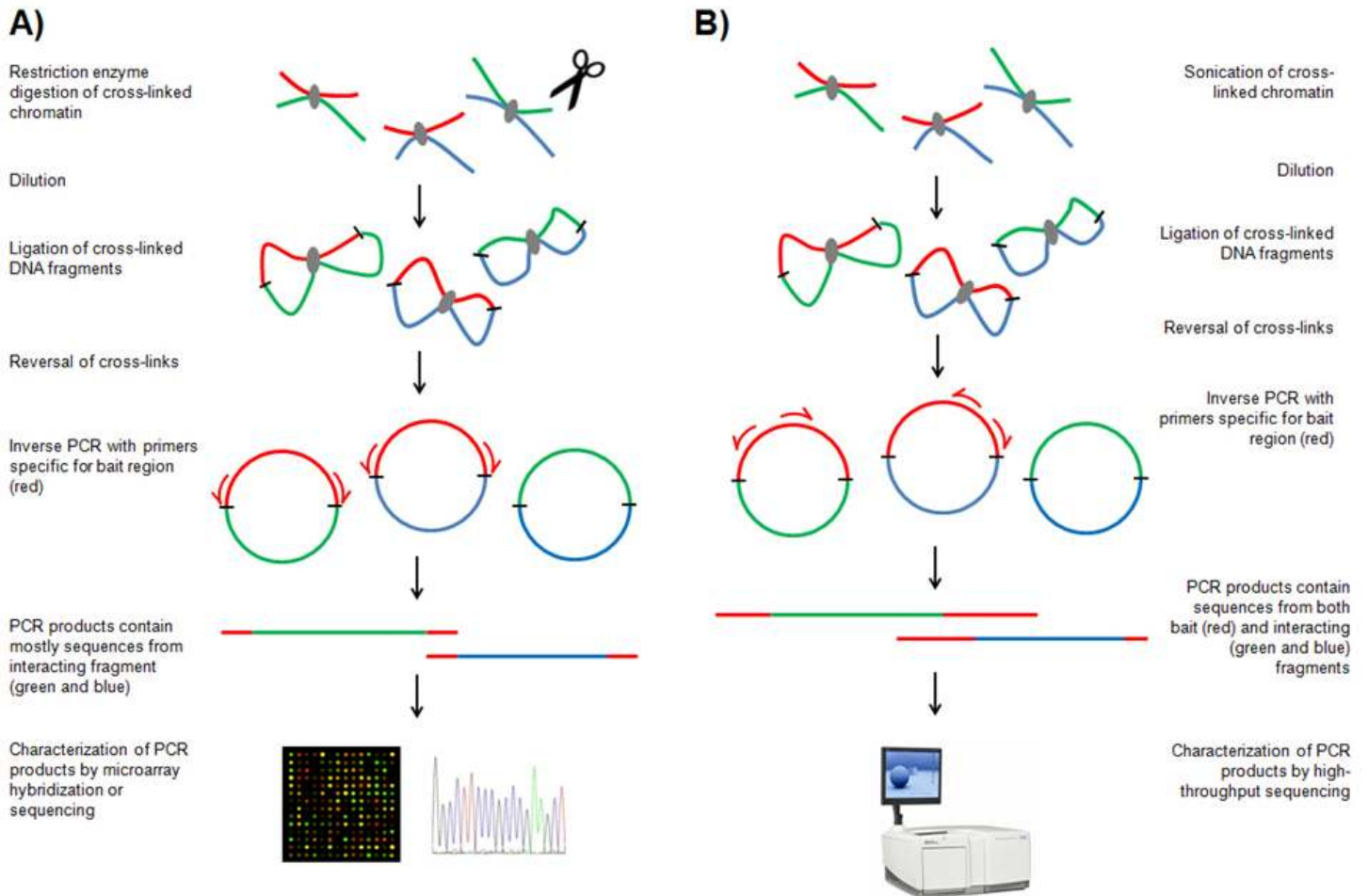
1783: 2052-2060. 3. Dekker J, Rippe K, Dekker M, Kleckner N \ (2002) Capturing chromosome conformation. *Science* 295: 1306-1311. 4. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, et al. \ (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip \ (4C). *Nat Genet* 38: 1348-1354. 5. Zhao Z, Tavoosidana G, Sjolinder M, Gondor A, Mariano P, et al. \ (2006) Circular chromosome conformation capture \ (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet* 38: 1341-1347. 6. Wurtele H, Chartrand P \ (2006) Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. *Chromosome Res* 14: 477-495. 7. Lomvardas S, Barnea G, Pisapia DJ, Mendelsohn M, Kirkland J, et al. \ (2006) Interchromosomal interactions and olfactory receptor choice. *Cell* 126: 403-413. 8. Fullwood MJ, Han Y, Wei CL, Ruan X, Ruan Y \ (2010) Chromatin interaction analysis using paired-end tag sequencing. *Curr Protoc Mol Biol* Chapter 21: Unit 21 15 21-25. 9. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, et al. \ (2009) An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* 462: 58-64. 10. Lieberman-Aiden E, van Berkum NL, Williams L, Imapkaev M, Ragooczy T, et al. \ (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326: 289-293. 11. Dekker J \ (2006) The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nat Methods* 3: 17-21. 12. Simonis M, Kooren J, de Laat W \ (2007) An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods* 4: 895-901. 13. Splinter E, Grosveld F, de Laat W \ (2004) 3C technology: analyzing the spatial organization of genomic loci in vivo. *Methods Enzymol* 375: 493-507. 14. Gondor A, Rougier C, Ohlsson R \ (2008) High-resolution circular chromosome conformation capture assay. *Nat Protoc* 3: 303-313. 15. Ohlsson R, Gondor A \ (2007) The 4C technique: the 'Rosetta stone' for genome biology in 3D? *Curr Opin Cell Biol* 19: 321-325. 16. Palstra RJ, Simonis M, Klous P, Brassat E, Eijkelkamp B, et al. \ (2008) Maintenance of long-range DNA interactions after inhibition of ongoing RNA polymerase II transcription. *PLoS One* 3: e1661. 17. Rozen S, Skaletsky H \ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Totowa, NJ: Humana Press. pp. 365-386. 18. Smit AFA, Hubley R, Green P \ (1996-2004) RepeatMasker Open-3.0. 19. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, et al. \ (2003) The UCSC Genome Browser Database. *Nucleic Acids Res* 31: 51-54. 20. Lu X, Lane EB \ (1990) Retrovirus-mediated transgenic keratin expression in cultured fibroblasts: specific domain functions in keratin stabilization and filament formation. *Cell* 62: 681-696. 21. Eid J, Fehr A, Gray J, Luong K, Lyle J, et al. \ (2009) Real-time DNA sequencing from single polymerase molecules. *Science* 323: 133-138.

## Acknowledgements

The authors acknowledge the Genome Technology and Biology Group and the Cancer Biology and Pharmacology group at the Genome Institute of Singapore for technical support in developing the protocol. The authors acknowledge the bioinformatics group supervised by Dr Ken Sung, as well as Mr. Atif Shahab, Mr. Chan Chee Seng, and Mr. Fabianus H. Mulawadi for computing support; and Drs Shujun Luo and Gary Schroth for Illumina sequencing support. M.J.F., P.Y.H.H., and Y.H. are supported by

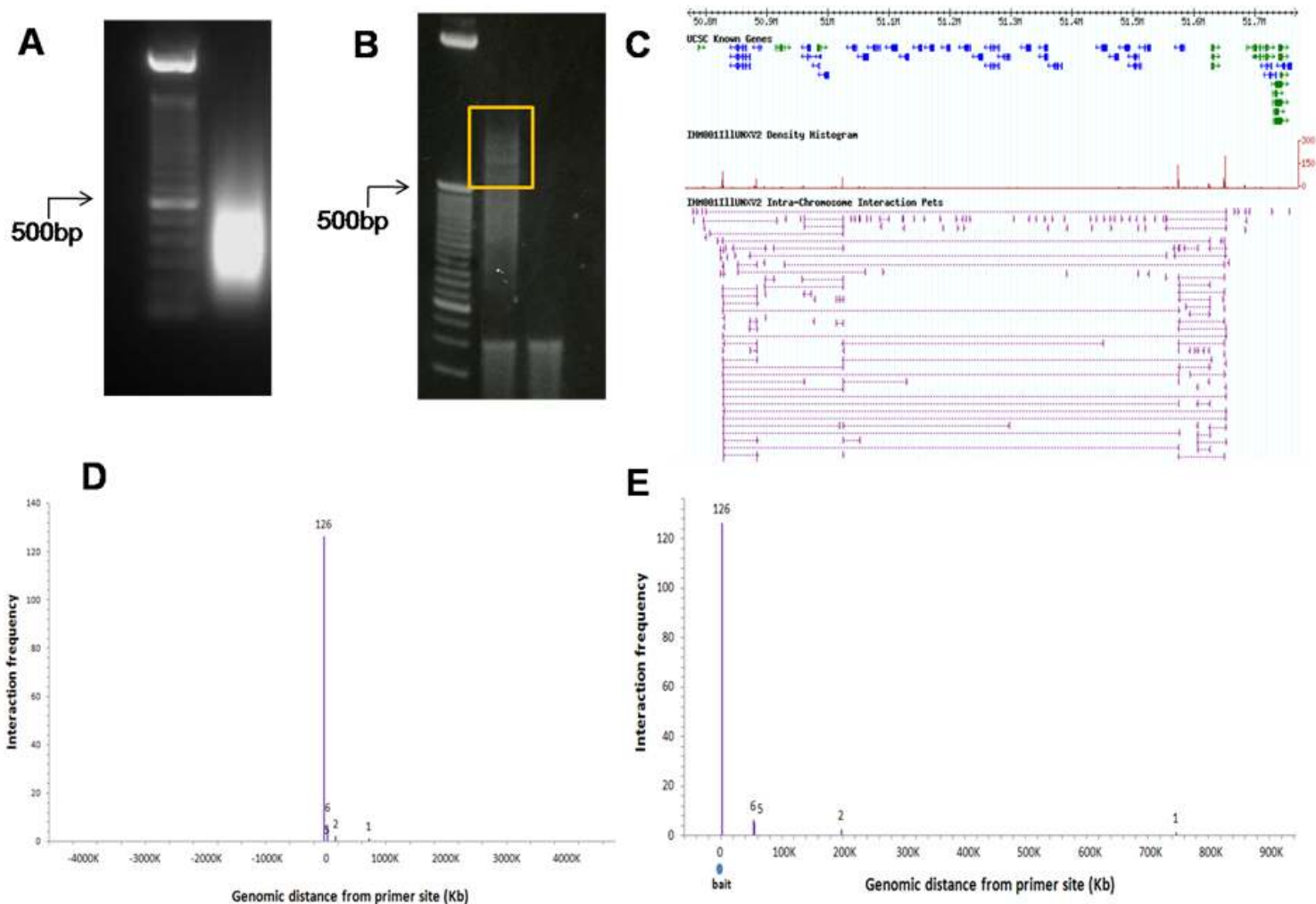
A\*\*STAR Scholarships. M.J.F. is supported by a 2009 L'Oreal For Women In Science National Fellowship and a 2010 Lee Kuan Yew Post-Doctoral Fellowship. Y.R. and C.L.W. are supported by A\*\*STAR of Singapore and NIH ENCODE grants (R01 HG004456-01, R01HG003521-01, and part of 1U54HG004557-01).

## Figures



**Figure 1**

Schematic comparison of 4C procedures. a. Outline of the original 4C method. b. Outline of the sonication-based 4C method.



**Figure 2**

4C validations a. Sonication quality control gel showing that the chromatin has been successfully fragmented into sizes of about 200 – 2000 bp. b. The 4C PCR products using the "bait" primer pair based at the KRT chromatin interaction region. The boxed range of DNA amplicon was gel-excised for sequencing analysis. c. Chromatin interactions at the KRT gene cluster identified by ERα ChIA-PET analysis. d. An enlarged view (10 Mb) of 4C sequence mapping centered at the KRT gene cluster shows that the 4C data is very clean. e. The 4C sequences mapped at the KRT gene cluster locus, aligned with the view in c. The highest 4C sequence mapping peak is at the 4C "bait" site (indicated by a blue dot). The interaction anchors of this interaction complex were mapped 4C sequences.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement0.pdf](#)