CAPTURE Hi-C LIBRARY GENERATION AND ANALYSIS TO DETECT CHROMATIN INTERACTIONS

Running Title: Capture Hi-C

Authors: Giulia Orlando[1*], Ben Kinnersley[1], Richard S Houlston[1]

Contact Information:

Giulia Orlando – email: giulia.orlando@icr.ac.uk, telephone: +442087224735
Ben Kinnersley – email: ben.kinnersley@icr.ac.uk, telephone: +4402087224424
Richard S Houlston – email: richard.houlston@icr.ac.uk, telephone: +442087224175

1. Division of Genetics and Epidemiology, The Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey, SM2 5NG, United Kingdom.

[*] To whom correspondence should be addressed

Significance Statement

In recent years, the structural and regulatory role of the genome has been reported, emphasizing the importance of identifying functionally active genomic regions. The development of chromosome conformation capture (3C) technologies allows researchers to investigate the 3D structure of chromatin, gaining knowledge about DNA packaging and mapping *cis*-regulatory elements to their target genes. Here we describe capture Hi-C (CHi-C), a technology that provides a platform to investigate genome-wide chromatin interactions with selected regions of interest. The protocols we provide include both laboratory and bioinformatics methods and can be customized to the researcher's needs, allowing the study of the regulatory role of the genome in any cell type.

ABSTRACT

Chromosome conformation capture (3C), coupled with next-generation sequencing (Hi-C), provides a means for deciphering not only the principles underlying genome folding and architecture, but more broadly, the role 3D chromatin structure plays in gene regulation and the replication and repair of DNA. The recent implemented modification, *in situ* Hi-C, maintains nuclear integrity during digestion and ligation steps, reducing random ligation of Hi-C fragments. Although Hi-C allows for genome-wide characterization of chromatin contacts, it requires high-depth sequencing to discover significant contacts. To address this, Capture Hi-C (CHi-C) enriches standard Hi-C libraries for regions of biological interest, for example by specifically targeting gene promoters, aiding identification of biologically significant chromatin interactions compared to conventional Hi-C, for an equivalent number of sequence reads. Illustrating the application of CHi-C applied to genome-wide analysis of chromatin interactions with promoters, we detail the protocols for *in situ* Hi-C and CHi-C library generation for sequencing, as well as the bioinformatics tools for data analysis.

Keywords: Capture Hi-C (CHi-C), di-tags, gene expression regulation, promoter-enhancer contacts

INTRODUCTION

Chromosome conformation capture (3C) technologies (Dekker, 2008; Dekker *et al.*, 2002; Rao *et al.*, 2014), have increased our understanding of how genome folding influences gene regulation (Dekker and Heard, 2015; Montefiori *et al.*, 2016; Servant *et al.*, 2015).

The high-throughput version of 3C termed Hi-C, has become a well-established technique for studying the 3D structure of chromatin providing fundamental insights about DNA packaging and its role in mediating promoter-enhancer interactions (Lieberman-Aiden *et al.*, 2009). The recent implementation of *in situ* Hi-C, maintaining an intact nucleus during digestion and ligation steps of library generation, improves resolution and sensitivity by reducing background noise (**Figure 1**) (Rao *et al.*, 2014). The Hi-C di-tags, defined as two different DNA fragments separated by a ligation junction, are generated following DNA de-crosslinking and purification, biotin pull-down and PCR amplification (**Figure 1**).

While Hi-C allows genome-wide characterization of chromatin contacts, increasing its effective resolution by a factor of n requires a quadratic (n^2) increase in the number of reads. To overcome this problem, several capture strategies have been developed to increase the proportion of relevant interactions (Hughes *et al.*, 2014; Jager *et al.*, 2015; Mifsud *et al.*, 2015; Sahlen *et al.*, 2015). The most widely used is promoter Capture Hi-C (PCHi-C) which enriches standard Hi-C libraries for regions of biological interest using custom-based biotinylated RNA baits specific for ~ 20,000 gene promoters (**Figure 1**) (Javierre *et al.*, 2016; Mifsud *et al.*, 2015).

Here we describe CHi-C applied to genome-wide chromatin interactions with promoters. We provide protocols detailing the preparation of *in situ* Hi-C (Rao *et al.*, 2014) and CHi-C libraries for sequencing (**Figure 1**). To illustrate CHi-C data analysis we make use of three replicates of an exemplar naïve B-cell promoter capture Hi-C dataset (Javierre *et al.*, 2016) , publicly available from the European genome-phenome archive (EGAD00001002268).

STRATEGIC PLANNING

Consideration of experimental design

For each sample, we advise creating CHi-C libraries for three biological replicates (*i.e.* cells fixed on different days) of $2.5x10^7$ cells in order to robustly demonstrate interactions. Before proceeding to the capture stage, we advise checking the quality of each Hi-C library by low coverage sequencing using an appropriate technology, such as MiSeq (Illumina Inc.). The low coverage Hi-C sequence that is generated can be used to assess successful enrichment of targeted regions in the subsequent capture step, by comparing the percentage of reads mapping to regions captured by RNA baits between Hi-C and CHi-C libraries. Additionally, low coverage data can be used to map topologically associating domains (TADs) (Dixon *et al.*, 2012).

BASIC PROTOCOL 1

DESIGNING OF THE CAPTURE RNA-BAITS

Depending on the biological questions, the biotinylated RNA-baits can be designed to target and enrich specific genomic regions for the relevant contacts. For example, to study germline susceptibility and identify candidate genes, disease-specific genome-wide associated loci can be targeted (Dryden *et al.*, 2014; Jager *et al.*, 2015). While capture technology can be tailored to answer differing biological questions in different organisms, we focus on the application of promoter CHi-C enriching for contacts involving human promoters of both protein-coding and non-coding transcripts. Below we provide a guideline for custom-designing the capture kit.

Materials

Ensembl genome browser (https://www.ensembl.org/index.html)
Bedtools (protocol version: 2.2.5)
SureDesign Custom Design Tool (https://earray.chem.agilent.com/suredesign)

1. Select the restriction enzyme to use for the Hi-C protocol (we use HindIII for promoter CHi-C).

   The choice of digestion enzyme is specific for the capture kit employed, as the RNA baits are designed around the restriction fragment sites. HindIII is a 6-base restriction enzyme, generating fragments on average 3kb long. Other enzymes can be used to perform digestion, such as the 4-base cutter MboI. However, compatibility with the capture kit has to be ensured prior to starting.

2. Generate the *in silico* cut genome of the desired organism using HiCUP.

   *Please refer to **BASIC PROTOCOL 7** for more details.*

3. Using the Ensembl browser, select the regions you are interested in capturing (in our case, the promoters).

   i. Open the browser at: https://www.ensembl.org/biomart/
   ii. In the Dataset window select “Ensembl Regulation 91” and “Human Regulatory features”
   iii. In the Filters window select all the autosomal and sexual chromosomes in the “Chromosome” window to restrict the search to the mapped genome.
   iv. In the same Filters window select “Promoters” in the “Feature type” window.
   v. In the Attributes window select “Chromosome/scaffold name”, “Start (bp)”, “End (bp)” and “Feature type”.
   vi. Download your list of promoters using the “Results” icon.

4. Using bedtools, intersect the selected regions with the cut genome and output a file containing the intersection.

   The aim of this step is to identify the corresponding HindIII fragment encompassing each promoter. Please refer to CPBI Unit 11.12 for the use of Bedtools.

5. Select the restriction fragments that overlap the relevant regions to design the RNA-baits with the following constraints:

   i. Design one 120bp long RNA-bait for each end of the restriction fragments
   ii. Keep the GC content of target sequence between 25-65%
   iii. Verify the genomic sequences complementary to the RNA-bait does not contain more than 2 successive Ns
   iv. Ensure that the RNA-baits are within 330bp of the restriction fragment cut-site.

   For those interested in capturing human promoters, we recommend using the existing RNA bait design (Agilent Design ID: 0681871)(Mifsud et al., 2015).

BASIC PROTOCOL 2

***IN SITU* Hi-C LIBRARY PREPARATION AND PURIFICATION**

Preparing *in situ* Hi-C libraries requires maintaining an intact nucleus during restriction enzyme digestion, biotinylation of DNA overhanging ends and ligation. Initially cells are treated with formaldehyde to cross-link DNA and protein, followed by a cell lysis step which disrupts the cytoplasm, while preserving the nucleus. The nuclear membrane is permeabilized using mild-detergents, such as Triton X-100, to allow restriction enzymes, DNA polymerases (required for the biotin-filling step) and the T4 DNA ligase to enter the nucleus. Finally, the DNA is de-crosslinked and the ligated DNA (*i.e.* Hi-C di-tags) is recovered, generating the *in situ* Hi-C library.

Materials

16% Formaldehyde (ThermoScientific)
1M glycine
Phosphate-buffered saline solution (PBS)
1M Tris-HCl, pH8
5M NaCl
1% Igepal CA-630 (Sigma)
Protease inhibitors (Sigma)
10% SDS (Sigma)
10% Triton X-100 (Sigma)
10X NEB buffer 2 (NEB)
100U/µl HindIII (NEB)
0.4mM biotin-14-dATP (ThermoScientific)
100mM dCTP, dGTP, dTTP

5U/µl DNA Polymerase I, Large (Klenow) Fragment (NEB)
10X T4 DNA ligase buffer with 10 mM ATP (NEB)
10mg/ml Bovine Serum Albumin (NEB)
400 U/µl T4 DNA Ligase (NEB)
Distilled water (Gibco)
3M sodium acetate, pH 5.2
100% Ethanol
70% Ethanol
Qubit BR DNA Kit (Invitrogen)
20mg/ml proteinase K (ThermoScientific)
1.5ml and 2ml DNA low-binding tubes (Eppendorf)

Cell crosslinking

1. Grow cells under recommended culture conditions to around 80% confluence.
2. Harvest $2.5x10^7$ cells, either collect suspension cells or detach adherent cells with trypsin, count and collect by centrifugation at 300xg for 5 minutes.
3. Re-suspend cells in fresh media at $1x10^6$ cells per 1ml media. Add formaldehyde solution to a final concentration of 1% v/v. Incubate at room temperature for 10 minutes on a tube roller shaker.
4. Add glycine solution to a final concentration of 0.2M to quench the formaldehyde. Incubate at room temperature for 5 minutes on a tube roller shaker.
5. Centrifuge for 5 minutes at 300xg at 4°C.
6. Re-suspend cells in 1ml of cold PBS and spin for 5 minutes at 300xg at 4°C. Discard supernatant and either flash-freeze cell pellets in dry ice and store at -80°C or proceed with the protocol.

Cell lysis and in-situ digestion

7. For each cell pellet to lyse prepare 1.8 ml of ice-cold Hi-C lysis buffer (10mM Tris-HCl pH8.0, 10mM NaCl, 0.2% Igepal CA630) and combine it with 200µl of protease inhibitors. Add 1 ml of lysis buffer to one crosslinked pellet of cells.
8. Incubate cell suspension on ice for 15 minutes. To help the lysis mix the suspension by inverting the tube every 5 min.

    Lysis buffer composition and duration of lysis might need to be optimised for different cell types.
9. Centrifuge the mix at 2,500xg at 4°C for 5 minutes. Remove the supernatant.
10. Wash pelleted nuclei with 500µl of the remaining ice-cold Hi-C lysis buffer. Centrifuge at 2,500xg at 4°C for 5 minutes.
11. Gently re-suspend pellet in 250µl of 0.5% SDS and incubate at 62°C for 10 minutes.
12. Add 725µl of water and 125µl of 10% Triton X-100 to quench the SDS. Mix slowly to avoid excessive foaming. Incubate at 37°C for 15 minutes.
13. Add 125µl of 10X NEB buffer2 and mix slowly the solution.

*Quality control 1 (**QC1**). Transfer 50µl into a 1.5 ml tube as undigested chromatin control (CTR1). Keep it at 4°C.*

14. Add 15µl of HindIII restriction enzyme, transfer the mix into a 2ml tube and digest chromatin overnight at 37°C on a tube rotator.

In-situ biotin marking of digested DNA ends and proximity ligation

15. Heat-inactivate HindIII at 65°C for 20 minutes.

    ***QC1**. Transfer 50µl into a 1.5ml tube as digested chromatin control (CTR2). Keep it at 4°C.*

16. Divide the rest of the sample into 4 aliquots (~250-300µl) and perform the following step for each aliquot.
17. To fill-in the digested fragment overhangs and mark the DNA ends with biotin, to each aliquot add:

    37.5µl of 0.4mM biotin-14-dATP
    1.5µl of 10mM dCTP
    1.5µl of 10mM dGTP
    1.5µl of 10mM dTTP
    8µl of DNA Polymerase I, Large (Klenow) Fragment

18. Mix by pipetting and incubate at 37°C for 90 minutes on a tube rotator.
19. Prepare 900µl of ligation master mix for each aliquot:

    658µl of distilled water
    120µl of 10X NEB T4 DNA ligase buffer
    100µl of 10% Triton X-100
    12µl of 10mg/ml Bovine Serum Albumin
    10µl of T4 DNA Ligase

20. Mix by inverting and incubate at 16°C overnight at 900rpm on a thermomixer.

DNA de-crosslinking and purification

21. To each aliquot add 50µl of 20mg/ml proteinase K and 120µl of 10% SDS. Incubate at 55°C for 30 minutes at 900rpm on a thermomixer. For CTR1 and CTR2 samples, add 2.5µl of proteinase K and 6µl of 10% SDS.
22. To de-crosslink the DNA, add 130µl of 5M sodium chloride and incubate at 68°C for four hours at 900rpm on a thermomixer. Add 6.5µl for CTR1 and CTR2 samples.
23. Cool tubes at room temperature and split each of the four 1.5ml aliquots into two 750µl aliquots in 2ml tubes (total of eight tubes per samples).
24. One by one, add 1.6X volumes of cold 100% ethanol and 0.1X volumes of 3M sodium acetate, pH 5.2, quickly mix by inverting and keep the tubes on ice. For CTR1 and CTR2 add 128µl of 100% ethanol and 8µl of sodium acetate.
25. Incubate all the aliquots at -80°C for 15 minutes.
26. Centrifuge at 18,000xg at 4°C for 15 minutes. Keep the tubes on ice after spinning and slowly remove the supernatant by pipetting.
27. Using 0.8-1.5ml of cold 70% ethanol, re-suspend the DNA pellets combining all the aliquots in a 1.5 ml tube. Centrifuge at 18,000xg at 4°C for 10 minutes.

The DNA pellets do not dissolve in ethanol, thus apply extra care not to lose the DNA while recovering it.

28. Discard all the supernatants and wash the pellet with 800µl of cold 70% ethanol. For CTR1 and CTR2 wash it in 200µl of ethanol. Centrifuge at 18,000xg at 4°C for 5 minutes.
29. Dissolve the pellet in 200-300µl of 10mM Tris-HCl, pH 8, and incubate at 37°C for 15 minutes.

    If the DNA pellet does not dissolve increase the incubation time and slowly mix by pipetting.
30. Quantify the amount of DNA in CTR1, CTR2 and in the ligated library with Qubit dsDNA BR assay and store it at 4°C (short-term storage) or at -20°C (long-term storage).

    DNA amounts depend on cell types (DNA content per cell), lysis and ethanol purification efficiency.
31. **QC1**: Load 200 ng of each sample (CTR1, CTR2 and ligated library) on 0.7% agarose gel, run it for 30 minutes at 100 V to check digestion and ligation efficiency (**Figure 2**).

BASIC PROTOCOL 3

DNA SONICATION AND SIZE SELECTION

The parameters for DNA sonication differ between instruments and should be appropriately optimised for the technology used. We provide an example on how to optimise sonication conditions using a focused-ultrasonicator LE220 (Covaris) (**SUPPORT PROTOCOL 1**). Sonication is required to shear Hi-C ligated DNA into fragments that are both compatible for the capture kit and high-throughput sequencing instruments. After sonication a double-sided size-selection is performed removing shorter and longer fragments, thus enriching for the desired size range of Hi-C di-tags.

Materials

AMPure XP beads (Beckman Coulter)
1M Tris-HCl, pH8
80% ethanol
1.5ml DNA low-binding tubes (Eppendorf)
RPT tips (TipOne®)
96-microTUBE plates (Covaris)
Focused-ultrasonicator LE220 (Covaris)
Magnetic rack (ThermoScientific)
2100 Bioanalyzer (Agilent)
DNA 1000 Bioanalyzer kit (Agilent)

DNA sonication

1. Take 40µg of the ligated library and proceed with it for the further steps. Freeze the rest at -20°C to store as a backup.

    *For each sample, we suggest processing 40µg for the non-indexed library, used for the capture step, and 30µg for the indexed library, used in the **QC3**.*

2. To obtain an efficient sonication each microTUBE (each well of the Covaris plate) should contain a maximum of 10µg of DNA aliquot in a final volume of 130µl of 10mM Tris-HCl, pH 8. Dilute the DNA accordingly and aliquot 130µl sample in four wells.
3. Shear to a size of 200-650bp using the following parameters:

   Target bp size: 400bp
   Instrument: Covaris LE220
   Volume of Library: 130µl in a Covaris microTUBE
   Fill Level: 6
   Duty Cycle: 10%
   PIP (Peak Incident Power): 175
   Cycles/Burst: 200
   Time: 60 seconds

4. Recover the entire volume from each well and transfer it in four 1.5ml tubes.
   QC2*: Save 5µl of any of the aliquots as a post sonication control.*
5. Wash each of the Covaris wells with 70µl of 10mM Tris-HCl, pH 8, and add to the four samples in the tubes, bringing the total volume to 200µl.

Double-sided size selection

6. Mix Ampure XP beads well by vortexing, keep at room temperature for at least 30 minutes.
7. For each of the four sonicated samples perform the following step separately.
8. Add 110 µl (0.55X) of Ampure XP bead suspension into the 200µl of sonicated DNA (Tube A).
9. Mix thoroughly, incubate for 20 minutes at room temperature.
10. Place the tube on a magnetic rack and recover the unbound supernatant containing the DNA in the desired size range (<700-1000bp) into a fresh tube (Tube B).
    QC2*: (Tube A) Keep the beads from any of the four aliquots as a control (>700-1000bp). Recover the bound fragments by washing twice with 80% ethanol and elute the fragments in 50 µl Tris-HCl, pH 8.*
11. For the second step of size selection protocol, prepare 240µl of Ampure XP beads. Concentrate the Ampure XP beads by placing the tube (Tube C) containing 240µl of beads on the magnetic rack, and remove all but 60µl of the supernatant (*i.e.* discard around 180µl of the supernatant). Take tube containing the beads off the magnetic and re-suspend the beads in the remaining 60µl volume (Tube C).
12. Add 60µl of concentrated beads (see above) from tube (Tube C) into the sample tube (Tube B) (0.9X). Mix well, incubate at room temperature for 10 minutes, place on magnetic, and discard supernatant containing fragments <200bp. The beads are now bound by DNA fragments between 200-650bp in size.
13. (Tube B) Wash twice with 80% ethanol. Leave the beads on the magnetic for 2 minutes to allow remaining ethanol to evaporate after the last wash.
    It is important that the 80% ethanol is freshly prepared.
14. Combine the four aliquots (Tube B) by re-suspending all the beads in a total of 200µl 10mM Tris-HCl, pH 8, and incubate at room temperature for 5 minutes, place on

magnetic and transfer supernatant, now containing your size-selected DNA into a fresh tube (Tube D). Discard tube (Tube B) containing the beads.

QC2*: Transfer 5 µl of the sheared DNA to a fresh 1.5ml tube.*

15. (Tube D) Add 100µl of 10mM Tris-HCl, pH8, to reach 300µl and store the size-selected DNA fragments at 4°C (short-term storage) or at -20/-80°C (long-term storage).
16. **QC2**: To check if both sonication and size-selection have worked, load the post-sonication control, the >700-1000bp and 200-650bp samples onto a DNA 1000 chip and run it on a Bioanalyzer instrument according to the manufactures' recommendations (**Figure 3**).

SUPPORT PROTOCOL 1

DNA SONICATION CALIBRATION

DNA sonication calibration has to be performed to insure optimal conditions to achieve the right target size and range of Hi-C DNA fragments.

Materials

1M Tris-HCl, pH8
1.5ml DNA low-binding tubes (Eppendorf)
RPT tips (TipOne®)
96-microTUBE plates (Covaris)
Focused-ultrasonicator LE220 (Covaris)
2100 Bioanalyzer (Agilent)
DNA 1000 Bioanalyzer kit (Agilent)

DNA sonication calibration

1. For each sonication condition to test, prepare 10µg of the ligated library in a final volume of 130µl 10mM Tris-HCl, pH8.
2. To obtain an efficient sonication transfer the samples to a microTUBE.
3. Test different sonication conditions by modulating either the PIP or the duration of sonication.
4. An example of the different conditions to use is reported in Table 1.
5. After sonication, recover the entire volume from each well and transfer it in 1.5ml tubes.
6. Load each sample onto a DNA 1000 chip and run it on a Bioanalyzer instrument according to the manufactures' recommendations (**Figure 4**).
7. Select the condition where the fragment size distribution peak most closely overlaps the desired target fragment size (*i.e*. for a target size of 400bp, we select COND8).

BASIC PROTOCOL 4

BIOTIN PULL-DOWN, END REPAIR AND ADAPTER LIGATIONS

The biotin-pull down step enriches for Hi-C di-tags that have been digested, biotin-filled and ligated to maximize the number of "true" Hi-C contacts. The pull-down is followed by an end-repair step to create blunt end DNA fragments damaged by the sonication, prior to the

attachment of 3' dATP for the subsequent ligation of adapters. These steps occur while the Hi-C di-tags are bound to the streptavidin beads.

Materials

Dynabeads MyOne Streptavidin T1 beads (Invitrogen)
1M Tris-HCl, pH7.5
1M Tris-HCl, pH8
0.5M EDTA
5M NaCl
Tween-20 (Sigma)
10X T4 DNA ligase buffer with 10 mM ATP (NEB)
10U/µl T4 PNK (NEB)
3U/µl T4 DNA polymerase I (NEB)
5U/µl DNA polymerase I, Large (Klenow) Fragment (NEB)
100mM dATP, dCTP, dGTP, dTTP
10X NEB buffer 2 (NEB)
5U/µl Klenow exo minus (NEB)
15µM pair-end (PE) annealed adapters (IDT)
Magneticic rack (ThermoScientific)
1.5ml DNA low-binding tubes (Eppendorf)
RPT tips (TipOne®)

Biotin pull-down of Hi-C-biotinylated fragments, end repair and adapter ligations

1. Re-suspend the Dynabeads MyOne Streptavidin T1 magnetic beads on a vortex mixer.
2. Prepare the beads for the biotin pull-down by washing 150µl of 10mg/ml Dynabeads MyOne Streptavidin T1 beads with 400µl of 1X Tween Washing Buffer (1X TWB: 5mM Tris-HCl, pH7.5, 0.5mM, EDTA; 1M NaCl; 0.05% Tween 20).
   1X TWB can be stored at room temperature.
3. Separate on a magnetic rack and discard the supernatant.
4. Re-suspend the beads in 300µl of 2X Binding Buffer (2X BB: 10mM Tris-HCl, pH7.5, 1mM EDTA; 2M NaCl) and add it to the whole 300µl size-selected library.
   2X BB can be stored at room temperature.
5. Incubate at room temperature for 15 minutes on a tube rotator to bind Hi-C biotinylated fragments to the streptavidin beads.
6. Separate on a magnetic rack and discard the supernatant.
7. To remove any non-specific binding, wash the beads by adding 600µl of 1X TWB. Transfer the mixture to a new tube. Heat the tubes at 55°C for 2 minutes with mixing at 900rpm. Reclaim the beads using a magnetic rack and discard the supernatant.
8. Repeat the wash as described in step 7.
9. Re-suspend beads in 100µl 1X T4 DNA ligase buffer and transfer to a new tube. Reclaim beads and discard the buffer.
10. To repair ends of sheared DNA, re-suspend beads in 100µl of freshly prepared master mix:

    88µl of 1X T4 DNA ligase buffer with 10mM ATP
    2µl of 25mM dNTP mix

5µl of 10U/µl T4 PNK
4µl of 3U/µl T4 DNA polymerase I
1µl of 5U/µl DNA polymerase I, Large (Klenow) Fragment

11. Incubate at room temperature for 30 minutes.
12. Once the reaction is completed, separate on a magnetic rack and remove the solution.
13. Repeat the wash step 7 twice.
14. Re-suspend beads in 100µl 1X Buffer 2 and transfer to a new tube.
15. Reclaim beads and discard the supernatant.
16. Re-suspend beads in freshly prepared 100µl of dATP attachment master mix:

    90µl of 1X NEB Buffer 2
    5µl of 10mM dATP
    5µl of 5U/µl Klenow exo minus

17. Incubate at 37°C for 30 minutes.
18. Once the reaction is completed, separate on a magnetic rack and remove the solution.
19. Repeat wash step 7 twice.
20. Re-suspend beads in 100µl 1X NEB T4 DNA ligase buffer and transfer to a new tube.
21. Reclaim beads and discard the solution.
22. Re-suspend in 50µl of 1X T4 DNA ligase buffer. Add 1200U of T4 DNA ligase and 3µl of 15µM annealed pair-end (PE) Illumina adapters. Mix well.

    Add appropriate PE adapters depending on whether you want to use indexed or non-indexed ones. Note that the promoter capture protocol we describe requires non-indexed adapters.
23. Incubate at 20°C for 2 hours with no mixing. Gently mix every 30 minutes.
24. Separate on a magnetic rack and discard the solution.
25. Repeat wash step 7 twice.
26. Re-suspend beads in 100µl 10mM Tris-HCl, pH8, and transfer to a new tube.
27. Reclaim beads and discard the solution.
28. Re-suspend in 50µl 10mM Tris-HCl, pH8.

    The Hi-C library is now bound to the beads and the DNA fragments have been ligated to the Illumina adaptors.
29. Keep the sample at 4°C (short-term storage) or at -20°C (long-term storage).

SUPPORT PROTOCOL 2

ADAPTERS ANNEALING

Before proceeding to PCR amplification, a specific pair of adapters compatible with the high-throughput sequencing (HTS) Illumina platforms has to be annealed and subsequently ligated to the library as described in **BASIC PROTOCOL 4**. The annealing step requires ramping PCR conditions to ensure the correct annealing of oligos.

Materials

10X T4 polynucleotide kinase buffer (NEB)

1.5ml DNA low-binding tubes (Eppendorf)
1M Tris-HCl, pH8
200µM Illumina PE adapter oligos (IDT)

Adapters annealing

1. Reconstitute adapter oligos (**Table 2**) to 200µM in 1X T4 polynucleotide kinase buffer.
   Use only HPLC-purified oligos.
2. Combine PE adapter 1 and PE adapter 2 at a 1:1 ratio (10µl:10µl).
3. Anneal the oligos using the following conditions:

95°C	5min	(ramp up at 0.5°C/s)
20°C	5s	(ramp down at 0.1°C/5s)
4°C	Hold	

4. Add 133.3µl 10mM Tris-HCl, pH8, to 20µl of adapter mix to obtain 15µM working concentration.
5. Store the annealed adapters at -20°C in small aliquots to minimise repeated freeze-thawing.

BASIC PROTOCOL 5

LIBRARY PCR AMPLIFICATION

Next it is necessary to amplify the Hi-C di-tags to ensure sufficient material for target capture. To do so, a PCR step is introduced, using primers compatible with the sequencing adapters. Firstly, an optimization step is performed to estimate the number of PCR cycles required to obtain the appropriate amount of Hi-C library. It is important not to perform an excessive number of PCR cycles in order to preserve the library complexity and avoid generating PCR duplicates.

Materials

10µM PCR Primer 1 (IDT)
10µM PCR Primer 2 (IDT)
Herculase II Fusion DNA Polymerase kit (Agilent)
1.5ml DNA low-binding tubes (Eppendorf)
RPT tips (TipOne®)
AMPure XP beads
80% ethanol
1M Tris-HCl, pH8
2100 Bioanalyzer (Agilent)
High sensitivity (HS) Bioanalyzer kit (Agilent)
Qubit high sensitivity (HS) DNA kit (Invitrogen)
Nuclease-free water (Gibco)
MiSeq Reagent Kit v3 (Illumina Inc.)

PCR optimization

1. The Hi-C library is amplified directly off the streptavidin beads with 4, 6 and 8 cycles of PCR, using the PCR primers
   Use only HPLC-purified PCR primers.
2. Perform PCR reactions using Herculase II Fusion DNA Polymerase kit.
   To avoid mis-incorporation of erroneous bases we do not recommend using a simple Taq Polymerase.
3. Prepare the following master mix reaction for PCR amplification:

   10µl of 5X Herculase Reaction buffer
   0.5µl of dNTP mix (25 mM each)
   1.25µl of PCR Primer 1 (10 µM)
   1.25µl of PCR primer 2 (10 µM)
   35.5µl of Water
   0.5µl of Herculase II fusion DNA polymerase

4. Test 4, 6 and 8 cycles in duplicates (6 reactions in total, using up to 6µl of the 50µl suspension beads) with the following PCR program

1 cycle:	2 min	98°C	(initial denaturation)
4, 6, or 8 cycles:	20 sec	98°C	(denaturation)
	30 sec	65°C	(annealing)
	30 sec	72°C	(extension)
1 cycle:	5 min	72°C	(final extension)
1 cycle:	hold	4°C	

5. After the PCR is completed, combine the two aliquots amplified with the same number of cycles.
6. For each of the samples, proceed with the following steps.
7. Separate on a magnetic rack, discard the supernatant and transfer to a fresh tube.
8. Mix Ampure XP beads well by vortexing, keep at room temperature for at least 30 minutes.
9. Add 100µl of beads to the PCR reaction (1X). Mix by pipetting and incubate at room temperature for 10 minutes.
10. Separate on a magnetic and remove the supernatant.
11. Keeping the beads on the magnetic rack, wash twice with 800µl of 80% ethanol without mixing.
    It is important that the 80% ethanol is freshly prepared.
12. Keep the beads on the magnetic rack at room temperature for 5 minutes to allow the ethanol to evaporate after the last wash.
13. Add 100µl of 10mM Tris-HCl pH8, to elute the amplified library. Mix and incubate at room temperature for 5 minutes.
14. Separate on a magnetic rack, and transfer the solution to a new tube.
15. Repeat step 9 to 12.

This further step of cleaning-up has been introduced to remove PCR primer dimers detected in the Bioanalyzer profile.

16. Add 25µl of 10mM Tris-HCl, pH8, to elute the amplified library. Mix and incubate at room temperature for 5 minutes.
17. Separate on a magnetic rack, and transfer the solution to a new tube.
18. Assess the concentration of each sample by quantifying the library using HS Qubit according to the manufacturer's recommendations.
19. For each cycle conditions, estimate the final amount of library by: (i) calculating how much DNA is obtained in a single PCR reaction;(ii) by multiplying this amount by the number of the remaining reactions to run (typically 44-50 reactions).
20. Choose the lower amount of cycles necessary to obtain ~ 1.5µg of DNA.

    At least 750ng DNA is required for the capture, thus the target amount is enough for at least two rounds of capture. Usually if all the steps are performed correctly 6 PCR cycles are optimal to get the required amount.
21. Finally, to assess the fragment distribution after the PCR run the samples onto a DNA HS chip and run it on a Bioanalyzer instrument according to the manufacturer's recommendations (**Figure 5**).

Final library amplification

22. Prepare the PCR reaction mix for all the remaining beads (~50µl). Aliquot the reaction (49µl) in a PCR plate and add 1µl of beads in each well on a PCR 96-well plate. Mix well and perform the PCR step with the chosen amount of cycles.
23. Once the PCR is over, recover all the reactions in 1.5ml tubes and separate the beads from the supernatant using a magnetic rack. The final amplified Hi-C library is in the supernatant.
24. Mix Ampure beads well by vortexing, maintain at room temperature for at least 30 minutes.
25. Aliquot the supernatant containing the amplified library in 600µl aliquots and add 1X of AMPure XP beads (600µl) to recover the library.
26. Mix by pipetting and incubate at room temperature for 10 minutes.
27. Separate on a magnetic rack and remove the supernatant. The Hi-C library is now bound to the beads.
28. Repeat wash steps 11 and 12.
29. Add 200µl of 10mM Tris-HCl, pH8, to elute DNA. Mix by pipetting and incubate at room temperature for 5 minutes.
30. Separate on a magnetic rack, and transfer the solution to a new tube.
31. Add 200µl (1X) of AMPure XP beads to recover the library. Mix by pipetting and incubate at room temperature for 10 minutes.
32. Separate on a magnetic rack and remove the clear supernatant.
33. Repeat wash steps 11 and 12.
34. Add 210µl of 10mM Tris-HCl, pH8, or in nuclease-free water to elute DNA. Mix by pipetting and incubate at room temperature for 5 minutes.
35. Separate on a magnetic rack, and transfer the solution to a new tube.
36. The result is a final *in situ* Hi-C library ready to be quantified and sequenced using an Illumina sequencing platform.

37. Quantify the library by HS Qubit and assess the distribution of the fragments by HS Bioanalyzer kit (Figure 6).
38. Keep the Hi-C library at 4°C (short-term storage) or at -20°C (long-term storage).
39. **QC3**: To assess library quality, run the sample on a MiSeq platform (Illumina Inc.) according to the manufacturer's recommendations. For each library, you need around 3-5 million reads. Refer to **BASIC PROTOCOL 7** for further information about how to process the data.

BASIC PROTOCOL 6

PROMOTER CAPTURE Hi-C (PCHi-C)

Promoter interactions are captured using an RNA-bait system (Agilent). Following hybridization, captured di-tags are recovered using biotin pull down. Finally, a PCR step is performed to amplify DNA necessary for HTS.

Materials

SSEL TE Reagent Kit ILM, PE Full Adapter (Agilent)
Custom promoter capture RNA-baits - Design ID: 0681871 (Agilent)
1.5ml DNA low-binding tubes (Eppendorf)
Parafilm
Dynabeads MyOne Streptavidin T1 beads (Invitrogen)
96-well plate mixer
RPT tips (TipOne®)
10µM PCR Primer 1 (IDT)
10µM PCR Primer 2 (IDT)
Herculase II Fusion DNA Polymerase kit (Agilent)
1.5ml DNA low-binding tubes (Eppendorf)
RPT tips (TipOne®)
AMPure XP beads
80% ethanol
1M Tris-HCl, pH8
2100 Bioanalyzer (Agilent)
High sensitivity (HS) Bioanalyzer kit (Agilent)
Qubit high sensitivity (HS) DNA kit (Invitrogen)
Nuclease-free water (Gibco)

Hybridization of Hi-C library to capture promoter interactions

1. For each Hi-C library prepared, carry out one hybridization and capture.
2. The hybridization reaction requires 750ng of prepared DNA in a volume of 3.4µl (concentration of 221ng/µl).
3. Transfer the volume equivalent to 1µg of library to an Eppendorf tube. Break off the cap, cover with parafilm, and poke several holes in the parafilm with a narrow gauge needle.
4. Dehydrate using a vacuum concentrator on low heat (≤45°C).
5. Reconstitute with 4.5µl of nuclease-free water to a final concentration of 221ng/µl. Pipette up and down along the sides of the tube for optimal recovery.

6. Transfer each 3.4µl library sample (750ng) to a separate well of a 96-well plate. Seal the wells and keep on ice.
7. Thaw all the frozen reagents on ice except the SureSelect Hyb 3 which is to be thawed at room temperature.
8. Slowly thaw the RNA-baits on ice. For each Hi-C library 5µl of RNA-baits are required.
   Aliquot the capture RNA-baits (5µl/aliquot) and store at -80°C until needed.
9. Prepare the required mixes in excess.
10. For each capture, prepare 13µl of hybridization buffer:

    6.63µl of SureSelect Hyb 1
    0.27µl of SureSelect Hyb 2
    2.65µl of SureSelect Hyb 3
    3.45µl of SureSelect Hyb 4

11. Keep the prepared Hybridization Buffer at room temperature until it is used.
    If a precipitate forms, warm the Hybridization Buffer at 65°C for 5 minutes.
12. Prepare the SureSelect Block mix and keep the mixture on ice until it is used:

    2.5µl of SureSelect Block 1
    2.5µl of SureSelect Block 2
    0.6µl of SureSelect Custom Block

13. For each capture prepare 2µl of the SureSelect RNase Block mix using nuclease-free water (1 part of RNase Block, 3 parts water). Prepare the dilution in excess and keep it on ice until required.
14. To each Hi-C library sample well prepared, add 5.6µl of the SureSelect Block mix prepared in step 11. Pipette up and down to mix.
15. Seal the wells and transfer the sealed plate to the thermal cycler.
16. Use a heated lid, set at 105°C, to hold the temperature at 65°C and run the following program:

1 cycle:	5 min	95°C
1 cycle:	Hold (at least 5 min)	65°C

    This step allows the denaturation of the DNA and the blocking of the adapters with the blocking oligos contained in the SureSelect Block mix. Make sure that the library DNA + Block Mix samples are held at 65°C for at least 5 minutes before adding the hybridization buffer.
17. To avoid changes in the temperature, warm-up the hybridization buffer at 65°C for 5 minutes.
18. Mix the 5µl of RNA-baits with 2µl of SureSelect RNase Block mix. Keep the solution on ice until needed.
19. Without removing the plate from the thermocycler, remove the seal and add the RNA-baits/RNase Block mix to a well in close proximity to the sample. Close the seal and incubate at 65°C for an additional 2 min.

20. Without removing the plate from the thermocycler, remove the seal and quickly add both 13µl of the pre-warmed hybridization buffer and 7µl of RNA-baits/RNase Block mix to the blocked Hi-C sample.
21. Mix well by pipetting up and down 8 to 10 times. The hybridization reaction wells now contain approximately 27-29µl, depending on the degree of evaporation during the thermal cycler incubation.
22. Incubate the hybridization mixture for 18-24 hours at 65°C with a heated lid at 105°C.

    Wells must be adequately sealed to minimize evaporation, or your results can be negatively impacted. You must avoid evaporation from the small volumes of the capture during the incubation.

Streptavidin pull-down to recover captured Hi-C library

23. Warm-up SureSelect Wash Buffer 2 at 65°C.
24. Re-suspend the Dynabeads MyOne Streptavidin T1 magnetic beads on a vortex mixer.
25. For each hybridization sample, prepare 50µl of the re-suspended beads into a new tube.
26. Wash the beads by adding 200µl of SureSelect Binding Buffer.
27. Mix by pipetting up and down until beads are fully re-suspended.
28. Separate the beads on a magnetic rack, wait until the solution is clear, then remove and discard the supernatant.
29. Repeat steps 26-28 two more times for a total of three washes.
30. Re-suspend the beads in 200µl of SureSelect Binding Buffer and transfer to a new PCR plate.
31. Maintain the hybridization plate at 65°C while transferring the entire volume (25-29µl) of each hybridization mixture to the wells tube containing 200µl of washed streptavidin beads. Mix well by slowly pipetting up and down until the solution is well mixed.
32. Seal the wells, then incubate the capture plate on a 96-well plate mixer, mixing vigorously (1400–1800 rpm) for 30 minutes at room temperature.
33. Transfer the mixture to a new tube, place the sample on a magnetic rack and discard the supernatant. The promoter capture Hi-C library is now attached on the beads.
34. Re-suspend the beads in 200µl of SureSelect Wash Buffer 1. Mix by pipetting until beads are fully re-suspended.
35. Incubate the samples at room temperature for 15 minutes.
36. Briefly spin in a centrifuge and transfer the tube on a magnetic rack. Wait for the solution to clear, then remove and discard the supernatant.
37. Wash the beads with SureSelect Wash Buffer 2. Re-suspend the beads in 200µl of 65°C pre-warmed Wash Buffer 2. Pipette up and down until beads are fully re-suspended.
38. Incubate the sample tube at 65°C for 10 minutes
39. Briefly spin in a centrifuge and transfer the tube on a magnetic rack. Wait for the solution to clear, then remove and discard the supernatant.
40. Repeat from step 37 to 39 for a total of 3 washes.
41. Make sure to remove all of the wash buffer during the final wash.

42. Re-suspend the beads in 30µl of nuclease-free water. The captured library is retained on the streptavidin beads.
43. Keep the PCHi-C library at 4°C (short-term storage) or at -20°C (long-term storage).

Final amplification of captured library

44. Prepare the PCR reaction mix and run the PCR program (**BASIC PROTOCOL 5**, ***PCR optimisation***) for all the beads (~30µl).
45. Prepare the following master mix reaction for PCR amplification:

    10µl of 5X Herculase Reaction buffer
    0.5µl of dNTP mix (25 mM each)
    1.25µl of PCR Primer 1 (10 µM)
    1.25µl of PCR Primer 2 (10 µM)
    34µl of Water
    0.5 µl of Herculase II fusion DNA polymerase

46. Aliquot the reaction (47.5µl) in a PCR plate and add 2.5µl of beads in each well on a PCR 96-well plate. Mix well and perform the PCR step with 5 cycles.
    The number of cycles can be optimised thus to obtain enough DNA for sequencing with the lower amount of PCR cycles.
47. Mix Ampure beads well by vortexing, keep at room temperature for at least 30 minutes.
48. Once the PCR is over, recover all the reactions in 1.5ml tubes and separate the beads from the supernatant using a magnetic rack. The final amplified promoter capture Hi-C library is in the supernatant.
    Keep the recovered beads as back-up for a second round of PCR amplification, if needed.
49. Aliquot the supernatant containing the amplified library in 600µl aliquots and add 1X of AMPure XP beads (600µl) to recover the library.
50. Repeat from step 25 to 33 of **BASIC PROTOCOL 5, *Final library amplification***.
51. Separate on a magnetic rack and remove the supernatant. The promoter capture Hi-C library is now bound to the beads.
52. Add 20µl of nuclease-free water to elute DNA. Mix by pipetting and incubate at room temperature for 5 minutes.
53. Separate on a magnetic rack, and transfer the solution to a new tube.
54. The result is a final *in situ* promoter capture Hi-C library ready to be quantified and sequenced using an Illumina sequencing platform.
55. Keep the promoter capture Hi-C library at 4°C (short-term storage) or at -20°C (long-term storage).
56. Quantify the library by HS Qubit and assess the distribution of the fragments by HS Bioanalyzer kit (**Figure 6**).

BASIC PROTOCOL 7

SEQUENCE ALIGNMENT AND QUALITY CONTROL OF PROMOTER CAPTURE HI-C LIBRARIES

This protocol assumes that Hi-C or CHi-C libraries have been prepared as described here and subjected to paired-end Illumina sequencing, producing FASTQ files containing sequence reads from each end of a read pair.

Here we illustrate use of HiCUP software (Wingett *et al.*, 2015) in initial processing of sequenced Hi-C and promoter capture Hi-C libraries. HiCUP handles sequence read alignment, pairing of reads sequenced from both ends into a Hi-C di-tag and removal of duplicate reads (*i.e.* PCR artefacts). Through comparison with an *in-silico* digested genome reference for library preparation, invalid read-pairs (*i.e.* same circularised, same fragment dangling ends, same fragment internal, re-ligation, continuous sequence, wrong size) can be identified and filtered out (**Figure 7**, **Table 3**). HiCUP outputs a bam file containing unique, valid read-pairs used for subsequent analysis. HiCUP additionally generates output reports of proportions of valid and invalid di-tags, useful to evaluate sequenced libraries, as well as initial quality controls (see **BASIC PROTOCOL 5: QC3**). We illustrate use of HiCUP on three replicates of an exemplar naïve B-cell promoter capture Hi-C dataset (Javierre *et al.*, 2016) which is available from the European genome-phenome archive (EGAD00001002268).

Materials

Linux environment (This protocol was tested on Intel Xeon Haswell 2.6 GHz CPU cores using up to 50Gb memory, see Table 7)
HiCUP software (Wingett *et al.*, 2015)
https://www.bioinformatics.babraham.ac.uk/projects/download.html#hicup
(protocol version: 0.5.10)
R (R Development Core Team, 2013) (protocol version: 3.4.0)
Samtools (Li *et al.*, 2009) (protocol version: 1.5)
Bowtie (Langmead, 2010) or Bowtie2 (Langmead and Salzberg, 2012) (Bowtie2 protocol version: 2.3.4)
Perl

Create/obtain input files

1. Reference genome fasta (for this protocol we use the hg19 fasta as reference from ftp://gsapubftp-anonymous@ftp.broadinstitute.org/bundle/hg19/ucsc.hg19.fasta.gz)
2. Digest map file (digesting reference genome into restriction fragments used in Hi-C library preparation – for this protocol we assume HindIII)
    a. Run *hicup_digester* perl script
    ```
    ./hicup_v0.5.10/hicup_digester --genome Human_hg19 --re1 A^AGCTT,HindIII ucsc.hg19.fasta.gz
    ```
    b. Output tab-delimited bed file format (Col = column):
    Col1 - chromsome
    Col2 – fragment start position
    Col3 – fragment end position
    Col4 – per-chromosome fragment number
    Col5 – restriction enzyme 1 fragment number (can give multiple restriction enzymes)
    Col6 – 5’ restriction site
    Col7 – 3’ restriction site

3. Rmap file – created from first three columns of previously generated digest map file
    a. Format – tab-delimited file:
       Col1 - chromosome
       Col2 – fragment start position
       Col3 – fragment end position
       Col4 – unique fragment identifier

4. Baitmap file (mapping bait intervals to restriction fragments) – for this protocol we use a set of 37,608 bait intervals mapping to 22,076 HindIII fragments. These baits were designed to capture ensembl-annotated promoters of protein-coding, noncoding, antisense, snRNA, miRNA and snoRNA transcripts – annotating 18,202 protein-coding and 10,929 non-protein genes as previously described (Javierre *et al.*, 2016; Mifsud *et al.*, 2015). See **BASIC PROTOCOL 1** for details on bait design.
    a. Format – tab-delimited file:
       Col1 – chromosome
       Col2 – baited fragment start position
       Col3 – baited fragment end position
       Col4 – unique fragment identifier (from Rmap file)
       Col5 – feature *e.g.* gene promoter

5. Bowtie indices (for this protocol we used pre-made bowtie2 indices for hg19 available from ftp://ftp.ccb.jhu.edu/pub/data/bowtie2_indexes/hg19.zip)

Running HiCUP

6. Create configuration file (example is provided in `/hicup_v0.5.10/config_files/hicup_example.conf`) detailing:
    a. Output directory
       `Outdir: /path/to/output`
    b. Path to alignment program (Bowtie or Bowtie2 – in this protocol we use Bowtie2)
       `Bowtie2: /apps/bowtie2/2.3.4/bowtie2`
    c. Path to reference genome indices
       `Index: /hg19_bowtie2_indices/hg19`
    d. Path to genome digest file produced by hiccup_digester
       `Digest: Digest_Human_hg19_HindIII_None_14-06-50_03-01-2018.txt`
    e. FASTQ format (can leave blank and HiCUP will try to determine automatically)
       `Format: Sanger`
    f. FASTQ files to be analysed
       `FASTQ_R1.fastq.gz`
       `FASTQ_R2.fastq.gz`
7. Run HiCUP
   `./hiccup_v0.5.10/hicup --config hicup_NB_34.conf`

HiCUP output

8. `Sample_combined_R1_2.hicup.bam`

9. `Sample_combined_R1_2.HiCUP_summary_report.html` (**Figure 7**) detailing summary of reads filtered and total valid read-pairs

Estimating capture efficiency

Hi-C and CHi-C library preparation aims to maximise proportion of valid sequenced fragments and in the case of CHi-C to maximise enrichment for baited regions (*i.e.* those where at least one read in pair overlaps a baited restriction fragment. Calculation of the proportion of baited fragments relative to total valid fragments reflects capture efficiency, which can aid optimisation and troubleshooting prior to HTS.

10. Run HiCUP `get_captured_reads` Perl script
    g. Requires main hicup directory to be in perl path, and samtools
    ```
    export PERL5LIB=/hicup_v0.5.10
    module load samtools
    ```
    h. Script takes as input a baitmap file (as per Step 4) and the hicup output bam file (Step 8)
    i. Running script
    ```
    /hicup_v0.5.10/Misc/get_captured_reads --baits
    baits.baitmap --header 0 sample.hicup.bam
    ```
11. Output file `sample.hicup.capture_summary.txt` contains "Percent_Total_Captured" as measure of efficiency, which is "Ditags_Processed"/"Total_Captured". This is also summarised in `sample.hicup.capture_charts.pdf` (**Figure 8**).
12. By comparing "Percent_Total_Captured" between PCHi-C and Hi-C libraries for the same biological replicate (if available), the enrichment of baited fragments can be estimated.

Summary read counts for naïve B-cell Hi-C and promoter CHi-C sequenced libraries as processed through HiCUP are provided in **Table 3**.

BASIC PROTOCOL 8

ANALYSIS OF CAPTURE HI-C DATA TO IDENTIFY SIGNIFICANTLY INTERACTING FRAGMENTS: USING THE CHiCAGO SOFTWARE PACKAGE

The CHiCAGO analysis package (Cairns *et al.*, 2016) has been developed specifically for CHi-C to address experimental biases (such as differential capture efficiency) and the asymmetry of CHi-C interaction matrices that affect calling of significant interactions. Here we illustrate use of CHiCAGO to identify significant interactions from an input set of unique valid di-tags generated from sequenced CHi-C libraries (as described in **BASIC PROTOCOL 7)**. We describe the generation of input files, running the software and visualisation of CHiCAGO output.

Materials

Linux environment (This protocol was tested on Intel Xeon Haswell 2.6 GHz CPU cores using up to 150Gb memory, see Table 7)
R (R Development Core Team, 2013) (protocol version: 3.4.0)
Bedtools (protocol version: 2.2.5)

Installing CHiCAGO

1. Set R path
   ```
   export R_LIBS_USER=/path/to/R/3.40
   ```
2. Install latest version of CHiCAGO R package from bitbucket
   ```
   install.packages("argparser")
   install.packages("devtools")
   library(devtools)
   install_bitbucket("chicagoTeam/Chicago",
   subdir="Chicago")
   install_bitbucket("chicagoTeam/Chicago",
   subdir="PCHiCdata")
   ```

Creating input files for CHiCAGO analysis

3. Baitmap and Rmap files are as previously generated above. It is recommended that these files are copied to a "DesignDir"
4. Run the following CHiCAGO_tools scripts (from https://bitbucket.org/chicagoTeam/chicago/) within the "DesignDir" (if only one .baitmap and .rmap file is in the DesignDir folder these will be used as input by default).
   ```
   python /CHiCAGO_tools/makeNPerBinFile.py
   python /CHiCAGO_tools/makeNBaitsPerBinFile.py
   python /CHiCAGO_tools/makeProxOEFile.py
   ```
5. Run `bam2chicago.sh`
   ```
   /CHiCAGO_tools/bam2chicago.sh \
   Sample.hicup.bam \
   /DesignDir/sample.baitmap /DesignDir/sample.rmap
   Output_sample_name
   ```
   Output will be "Ouput_sample_name.chinput"

Running CHiCAGO

6. Run "runChicago.R" Rscript
   ```
   Rscript /CHiCAGO_tools/runChicago.R \
   --design-dir /DesignDir/ --cutoff 0 \
   --output-dir /CHiCAGO_output/ \
   Replicate_1.chinput,\
   Replicate_2.chinput,\
   Replicate_3.chinput \
   OUTPUT_NAME
   ```
7. Run parameters:
   a. "--cutoff 0" – specify "CHiCAGO score" threshold to limit output interactions file (by default 5). To output all interactions set cutoff to 0.
   b. CHiCAGO output several plots to allow assessment of the algorithm performance (**Figure 9**).
   c. By default output interactions are in Wash U format (`sample_washU_text.txt`)

Visualisation in WashU Epigenome Browser

8. Go to https://epigenomegateway.wustl.edu/ and select "Wash U EpiGenome Browser"
9. Select reference genome *e.g.* Human hg19
10. Go to "CUSTOM tracks" and select "Upload them from your computer"
11. Upload the *_washU_text.txt file created from CHiCAGO
12. Go to "Setup" then select "Pairwise Interaction" format and "add as TRACK"
13. Right click over newly added track and select "Arc" to visualise interactions as arcs between fragments (**Figure 10**).

ALTERNATE PROTOCOL 1

ANALYSIS OF CAPTURE HI-C DATA TO IDENTIFY SIGNIFICANTLY INTERACTING FRAGMENTS: USING THE GOTHiC SOFTWARE PACKAGE

GOTHiC (Mifsud *et al.*, 2017) is an approach for analysis of Hi-C data that uses a simple binomial probabilistic model to call significantly interacting DNA fragments. Here we illustrate GOTHiC as a complementary approach. GOTHiC has previously been used in analysis of promoter CHi-C data (Mifsud *et al.*, 2015) although there may be software version differences (the version for this protocol is 1.14.0), therefore if in doubt users should contact the authors directly.

Materials

Linux environment (This protocol was tested on Intel Xeon Haswell 2.6 GHz CPU cores using up to 150Gb memory, see Table 7)
R (R Development Core Team, 2013) (for this protocol we use version 3.4.0)
Download and installation of HiCUP (as per **BASIC PROTOCOL 7**)
Perl

Prepare input files for GOTHiC

1. Run hicup2gothic conversion script

Install GOTHiC

2. Run R

```
R
> source("https://bioconductor.org/biocLite.R")
> biocLite("GOTHiC")
> library(GOTHiC)
```

Run GOTHiC

3. Load GOTHiC library and run GOTHiChicup command

```
R
> library(GOTHiC)
> output <- GOTHiChicup("input.hicup.bam.gothic",
sampleName='sample_name', res=1,
"Digest_file", cistrans='all', parallel=FALSE, cores=1)
```

   a. "input.hicup.bam.gothic" – input interactions file produced by hicup2gothic conversion script as above
   b. "sample_name" – name for output pdf *P*-value distribution plot (see below)

    c. “res=1” – set bp resolution (for fragment level *e.g.* similar to CHiCAGO, set to 1)
    d. “Digest_file” – HiCUP reference restriction fragment interval file
4. Results stored in “output” dataframe (interacting fragments/regions 1 and 2):
    a. chr1
    b. locus1
    c. chr2
    d. locus2
    e. relCoverage1
    f. relCoverage2
    g. probability
    h. expected
    i. readCount
    j. pvalue
    k. qvalue
    l. logObservedOverExpected
2. Density distribution plot of *P*-values is provided in “sample_name.pdf” (**Figure 11**).
5. Save output dataframe as file

```
> write.table(output, file=”output.tsv", append = FALSE, quote = FALSE, sep = "\t", eol = "\n", na = "NA", dec = ".", row.names = FALSE, col.names = TRUE, qmethod="escape", fileEncoding="")
```

6. Filter significant interactions (*e.g.* q<0.05 in each replicate)

COMMENTARY

Background Information

Chromosome conformation capture (3C) techniques have been recently developed to study the 3D structure of chromatin identifying spatial proximity between distal genomic regions (Bonev and Cavalli, 2016). All 3C techniques involve a restriction digestion step followed by ligation forming proximity-ligated DNA fragments that were not contiguously located in the genome (Hi-C di-tags). Hi-C couples 3C techniques with HTS and allows building a genomic-wide map of DNA contacts (Dekker *et al.*, 2013). Numerous variations of Hi-C protocol have been implemented to improve resolution; for example, restriction enzyme digestion using the 4-base cutter MboI increases resolution from 10kb to 1kb, compared against the 6-base cutter HindIII. A major improvement was achieved by performing ligation *in situ* (Rao et al., 2014). Ligation was initially performed *in dilution* in conjunction with lysis of the nuclear membrane whereas *in situ* ligation preserves an intact nucleus, reducing the background noise created by randomly ligated fragments in solution. The Hi-C protocol we described here is based on the *in situ* Hi-C protocol implemented by Lieberman-Aiden lab (Rao *et al.*, 2014), adapted to include a capture step to enrich for genomic regions of interest (see below). We increased the required number of cells to ensure a high amount of DNA without compromising the library complexity. Also, we describe a modification to the digestion step by using the 6-base cutter HindIII, and the ligation step by increasing the length of incubation, as well as the sonication and size selection steps, to obtain a 200-650bp library size.

While Hi-C has the advantage of generating whole genome contact matrices over other 3C techniques, such as 3C or 4C, the amount of sequencing necessary to obtain reliable contact maps is often prohibitive due to the high cost of HTS. To circumvent this issue, several groups employed a capture step, with the aim of enriching for region of interest reducing the amount and, therefore, the cost of sequencing (Hughes *et al.*, 2014; Mifsud *et al.*, 2015; Sahlen *et al.*, 2015). Several variants of these techniques have been reported (CaptureC, HiCap and CHi-C). The most widely used is promoter CHi-C which allows the identification of a genome-wide promoter interaction map using custom designed RNA-baits. Each RNA-bait is 120bp long and complementary to either the beginning or the end of a HindIII fragment mapping to a gene promoter. The promoter kit used in our protocol comprises 32,314 RNA-baits mapping to 19,023 promoter fragments. Here we integrated Hi-C with promoter capture.

Critical Parameters

The number of cells and the DNA content per cell are crucial to ensure efficient DNA digestion for a given quantity of restriction enzyme. Too much DNA can decrease digestion efficiency reducing the final number of valid Hi-C di-tags. QC1 is employed to estimate digestion efficiency before proceeding further with the protocol (**Figure 2**).

Sonication has to be carefully optimised to produce fragments with the desired target size and therefore ensuring high capture efficiency (**Figure 3**). Finally, it is important to remove PCR primer dimers after the final step of amplification to avoid the wastage of sequencing reagents. Primer dimers can be identified in the Bioanalyzer profile as a peak approximately at 75bp.

Troubleshooting

See Table 6 for troubleshooting details.

Statistical Analyses

Handling of biological replicates and identifying significant interactions
CHiCAGO combines reads from all biological replicates in calculation of a di-tag's interaction score. The default score threshold for significant interactions is > 5 (Cairns *et al.*, 2016).

In contrast, GOTHiC is run separately on biological replicates. Read pairs with a FDR *q*-value < 0.05 in each replicate have previously been advocated to represent "significant" interactions (Mifsud *et al.*, 2017).

Significant interactions from analysis of the same naïve B-cell PCHi-C dataset using CHiCAGO and GOTHiC are reported in Table 4 and Table 5 respectively.

CHiCAGO background model
Parameters for the CHiCAGO *P*-value weighting procedure were estimated from macrophage PCHi-C libraries. While they may serve reasonably well for human libraries, it is possible to estimate these parameters using the `fitDistCurve.R` script, after running CHiCAGO on individual sample replicates (as detailed in **BASIC PROTOCOL 7**) and producing .rds files:

```
Rscript fitDistCurve.R sample –inputs
sample1.rds,sample2.rds,sample3.rds
```

This produces a `sample.settings` file which can be provided as an option to `runChicago.R` via `–settings-file` (http://regulatorygenomicsgroup.org/wp-content/uploads/Chicago_vignette.html#using-different-weights).

Understanding Results

Evaluating CHi-C library quality
As previously discussed, the higher the proportion of valid to invalid di-tags, the better quality the Hi-C library and the greater the value-for-money in terms of usable reads sequenced. A good quality library has 80-90% valid read pairs of which 80-90% are *cis* di-tags (**Figure 7, Table 3**). Lower proportions suggest sub-optimal performance of steps in CHi-C library generation. High percentage of *cis* versus *trans* di-tags indicate that the nucleus has remained intact and ligation carried out *in situ*. Finally, a low proportion of duplicate reads is a consequence of a low number of PCR cycles. Since only 3-5 million reads are sufficient for evaluation, we suggest multiplexing 4-6 Hi-C libraries on a MiSeq (Illumina Inc.). Subsequent sequencing data is analysed using the HiCUP pipeline as described in **BASIC PROTOCOL 7**.

Evaluating bait capture efficiency
As previously described, efficiency of bait capture can be checked by calculating the proportion of unique valid reads mapping to the baited fragments. For promoter CHi-C, the

capture reads should be 55-70% of the total library compared to 4-5% in the Hi-C library, leading to a 10-15 fold enrichment of capture reads. It is advisable to sequence at least two lanes on a HiSeq sequencer (Illumina Inc.) to attain a total of 1 billion reads for each captured library. This assumes all baits have the same capability of binding to the DNA.

Assessing enrichment for biological features

CHiCAGO analysis allows assessment of whether significantly interacting DNA fragments are enriched for certain biological features (http://regulatorygenomicsgroup.org/wp-content/uploads/Chicago_vignette.html#peak-enrichment-for-features). This can be achieved by providing an argument to `runChicago.R.` of `–en-feat-list EN_FEAT_LIST` where `EN_FEAT_LIST` is a file of format <feature-name> <feature-bed-file-location> in which the feature bed file contains chromosome, start and end coordinates of the features.

Visualisation of CHi-C interactions

As per steps 8-13 of **BASIC PROTOCOL 8**, CHiCAGO output can be uploaded as a track on the Wash U epigenome browser allowing visualisation of interactions.

Time Considerations

For the experimental procedure 7 days are required to complete the CHiC protocol. For the computational protocol rough timings for various aspects of the analyses are presented in Table 7.

FIGURE LEGENDS

Figure 1. General work-flow for CHi-C library generation and analysis. The figure represents the main steps for the generation of CHi-C libraries, illustrating the Hi-C protocol and the capture of selected di-tags using custom RNA baits.

Figure 2. Quality control 1 (QC1) profiles. Undigested, digested and ligated DNA samples were run on an ethidium-bromide stained 0.7% agarose gel to check for digestion and ligation efficiency. The left-panel shows a successful digestion (lane 1 vs lane 2) where a clear smear is present after treatment with HindIII enzyme; the same smear is less prominent after treatment with DNA ligase (lane 2 vs lane 3), demonstrating that the ligation has occurred. The right panel shows low digestion efficiency as exemplified by the lack of a smear following digestion (lane 1 vs lane 2).

Figure 3. Quality control 2 (QC2) profiles. DNA samples were loaded on DNA 1000 chip to show successful sonication (upper panel), elimination of the fragment > 600bp (middle panel) and enrichment of the 200-650bp fragments after size-selection (lower panel).

Figure 4. DNA sonication calibration. DNA samples were loaded on DNA 1000 chip to determine the optimal parameters of sonication for a 400bp target size (COND 8). The conditions used are reported in Table 1.

Figure 5. Hi-C library profiles after PCR optimisation. DNA samples were loaded on DNA HS chip to show successful PCR amplification following adapter ligation. The profiles show an exponential increase in DNA concentration corresponding to a higher number of PCR cycles performed; upper panel 4 PCR cycles, middle panel 6 PCR cycles, lower panel 8 PCR cycles. It is important to note that after ligating the adapter, there is a shift in the size range distribution of 75bp.

Figure 6. Final Hi-C and CHi-C library profiles. DNA samples were loaded on DNA HS chip to assess that the final PCR amplification was performed correctly. As shown, the fragment distribution of a final Hi-C library (upper panel) and of a final CHi-C library (lower panel) should have the same profile.

Figure 7. HiCUP summary report. The figure shows an example of HiCUP summary report including details of (a) truncation and mapping of the reads, (b) filtering of valid and invalid di-tags, (c) distribution of di-tag sizes selected for further processing (by default 150bp and 800bp respectively) and finally (d) read-pair de-duplication. Total unique di-tags are all included in the output HiCUP bam file (*i.e.* truncated and paired reads, valid and unique di-tags).

Figure 8. Example pie-chart showing percentages of captured reads.

Figure 9. Example CHiCAGO diagnostic output plots. The figure shows an example of the (a) distance function estimate, of the (b) Brownian other-end factors estimate and of the (c) technical noise estimates for both baits and other ends.

Figure 10. Example WashU Epigenome Browser session with significant (*i.e.* CHiCAGO score > 5) PCHi-C interactions added as a track.

Figure 11. Example *P*-value distribution plot output from GOTHiC.

TABLES

Table 1. DNA sonication parameters

	PIP	Time (s)
COND1	140	110
COND2	-	100
COND3	-	85
COND4	-	70
COND5	175	100
COND6	-	90
COND7	-	75
COND8	-	60
COND9	-	55

Table 2. Sequences of the Illumina PE adapter oligos and PCR primers

Name	Sequence
PE adapter 1	ACACTCTTTCCCTACACGACGCTCTTCCGATC*T
PE adapter 2	/5Phos/GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
PCR primer 1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T
PCR primer 2	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T

Table 3. Summary of HiCUP read filtering. Reported are summaries of sequenced promoter CHi-C libraries from three naïve B cell replicates (NB 34, 35, 36) and Hi-C libraries from two of these replicates (NB 35, 36) as described in Javierre *et al.* (Javierre *et al.*, 2016). Analysis was carried out as described in **BASIC PROTOCOL 7**.

		NB 34 PCHi-C	
		Read 1	**Read 2**
Truncation	**Total Reads**	696,454,802	696,454,802
	Not Truncated	635,861,025	640,082,064
	Truncated	60,593,777	56,372,738
	Too short to map	9,608,682	8,671,553
	Average length of truncated sequence	30.10	30.29
Mapping	**Unique alignments**	569,629,384	562,416,547
	Multiple Alignments	107,240,512	107,900,726
	Failed to Align	9,976,224	17,465,976
	Paired	476,536,107	476,536,107
		Di-tag count	
Filtering	**Valid Pairs**	415,053,501	
	Invalid Pairs	61,482,606	
	Same Circularised	4,564,226	
	Same Fragment Dangling Ends	1,312,225	
	Same Fragment Internal	3,508,348	
	Re-ligation	42,604,443	
	Continuous Sequence	2,644,736	
	Wrong Size	6,848,628	
	Total Pairs	476,536,107	
		All Di-Tags	**Unique Di-Tags**
De-duplication	**Read Pairs**	415,053,501	289,580,871
	Cis-close (< 10Kbp)	59,908,805	41,606,630
	Cis-far (> 10Kbp)	259,136,374	181,024,659
	Trans	96,008,322	66,949,582
		Unique Di-Tags	**Efficiency**
Promoter captured	**Unique**	179,105,418	61.85%

		NB 35 Hi-C	
		Read 1	**Read 2**
Truncation	**Total Reads**	273,579,404	273,579,404
	Not Truncated	243,991,683	245,635,115
	Truncated	29,587,721	27,944,289
	Too short to map	4,236,438	3,835,561
	Average length of truncated sequence	30.46	30.65
Mapping	**Unique alignments**	202,247,247	200,832,554
	Multiple Alignments	57,925,388	57,285,168
	Failed to Align	9,170,331	11,626,121
	Paired	161,662,095	161,662,095
		Di-tag count	
Filtering	**Valid Pairs**	131,468,107	
	Invalid Pairs	30,193,988	
	Same Circularised	1,041,466	
	Same Fragment Dangling Ends	2,119,343	
	Same Fragment Internal	5,642,738	
	Re-ligation	13,709,277	
	Continuous Sequence	652,160	
	Wrong Size	7,029,004	
	Total Pairs	161,662,095	
		All Di-Tags	**Unique Di-Tags**
De-duplication	**Read Pairs**	131,468,107	129,285,387
	Cis-close (< 10Kbp)	19,082,128	18,679,090
	Cis-far (> 10Kbp)	83,776,188	82,444,493
	Trans	28,609,791	28,161,804
		Unique Di-Tags	**Efficiency**
Promoter captured	**Unique**	6,230,495	4.82%

		NB 35 PCHi-C	
		Read 1	**Read 2**
Truncation	**Total Reads**	715,157,005	715,157,005
	Not Truncated	647,431,660	650,869,810
	Truncated	67,725,345	64,287,195
	Too short to map	9,761,284	8,876,739
	Average length of truncated sequence	30.43	30.64
Mapping	**Unique alignments**	581,175,416	578,205,011
	Multiple Alignments	108,298,830	109,167,203
	Failed to Align	15,921,475	18,908,052
	Paired	489,876,305	489,876,305
		Di-tag count	
Filtering	**Valid Pairs**	411,601,907	
	Invalid Pairs	78,274,398	
	Same Circularised	3,652,356	
	Same Fragment Dangling Ends	5,780,731	
	Same Fragment Internal	6,590,808	
	Re-ligation	45,940,047	
	Continuous Sequence	2,179,920	
	Wrong Size	14,130,536	
	Total Pairs	489,876,305	
		All Di-Tags	**Unique Di-Tags**
De-duplication	**Read Pairs**	411,601,907	301,482,079
	Cis-close (< 10Kbp)	57,532,280	41,350,809
	Cis-far (> 10Kbp)	258,877,626	190,229,885
	Trans	95,192,001	69,901,385
		Unique Di-Tags	**Efficiency**
Promoter captured	**Unique**	195,920,542	64.99%

		NB 36 Hi-C	
		Read 1	**Read 2**
Truncation	**Total Reads**	270,628,948	270,628,948
	Not Truncated	244,334,728	246,107,174
	Truncated	26,294,220	24,521,774
	Too short to map	3,865,244	3,396,309
	Average length of truncated sequence	30.39	30.65
Mapping	**Unique alignments**	201,213,730	199,832,108
	Multiple Alignments	57,912,976	57,300,335
	Failed to Align	7,636,998	10,100,196
	Paired	159,547,601	159,547,601
		Di-tag count	
Filtering	**Valid Pairs**	126,383,895	
	Invalid Pairs	33,163,706	
	Same Circularised	1,455,817	
	Same Fragment Dangling Ends	2,301,891	
	Same Fragment Internal	7,701,755	
	Re-ligation	9,391,389	
	Continuous Sequence	468,908	
	Wrong Size	11,843,946	
	Total Pairs	159,547,601	
		All Di-Tags	**Unique Di-Tags**
De-duplication	**Read Pairs**	126,383,895	125,303,312
	Cis-close (< 10Kbp)	16,168,108	15,969,982
	Cis-far (> 10Kbp)	81,142,040	80,488,857
	Trans	29,073,747	28,844,473
		Unique Di-Tags	**Efficiency**
Promoter captured	**Unique**	6,209,523	4.96%

		NB 36 PCHi-C	
		Read 1	**Read 2**
Truncation	**Total Reads**	715,650,932	715,650,932
	Not Truncated	653,428,791	657,337,337
	Truncated	62,222,141	58,313,595
	Too short to map	9,111,518	8,062,429
	Average length of truncated sequence	30.42	30.70
Mapping	**Unique alignments**	585,070,946	581,627,101
	Multiple Alignments	105,333,764	106,103,895
	Failed to Align	16,134,704	19,857,507
	Paired	491,761,777	491,761,777
		Di-tag count	
Filtering	**Valid Pairs**	415,019,121	
	Invalid Pairs	76,742,656	
	Same Circularised	5,071,301	
	Same Fragment Dangling Ends	6,108,819	
	Same Fragment Internal	6,531,821	
	Re-ligation	32,769,075	
	Continuous Sequence	1,654,617	
	Wrong Size	24,607,023	
	Total Pairs	491,761,777	
		All Di-Tags	**Unique Di-Tags**
De-duplication	**Read Pairs**	415,019,121	327,661,912
	Cis-close (< 10Kbp)	50,745,960	39,376,525
	Cis-far (> 10Kbp)	263,315,304	208,328,960
	Trans	100,957,857	79,956,427
		Unique Di-Tags	**Efficiency**
Promoter captured	**Unique**	238,729,545	72.86%

Table 4. CHiCAGO interactions summary. Combined interactions are from running CHiCAGO with all three naïve B cell replicates (NB 34, 35, 36) (Javierre *et al.*, 2016). Analysis was carried out as described in **BASIC PROTOCOL 8**.

	NB 34	NB 35	NB 36	Combined
Total interactions	91,447,913	96,835,175	119,187,369	263,023,303
Score > 5	176,980	187,263	219,509	178,525
Cis	21,970,823	22,712,563	25,877,559	41,356,532
Cis score > 5	137,292	142,830	152,107	170,916
Far cis (>10Mb)	29,806,305	31,478,096	38,731,588	86,842,388
Far cis (>10Mb) score > 5	27,039	31,638	55,167	2,400
Trans	39,670,784	42,644,515	54,578,221	134,824,382
Trans score > 5	7,020	6,304	5,866	383

Table 5. GOTHiC interactions summary. Combined interactions are those meeting the q < 0.05 score in each of the three naïve B cell replicates (NB 34, 35, 36) (Javierre *et al.*, 2016). Analysis was carried out as described in ALTERNATIVE PROTOCOL 1.

	NB 34	NB 35	NB 36	Combined
Total interactions	184,793,789	188,778,389	202,092,971	
FDR q < 0.05	7,662,199	7,743,526	7,740,697	4,205,950
Cis	63,954,097	63,849,922	62,573,319	
Cis fdr q < 0.05	7,645,345	7,723,321	7,709,262	4,204,814
Far cis (>10Mb)	54,249,487	55,461,019	60,204,743	
Far cis (>10Mb) fdr q < 0.05	15,872	19,196	30,328	917
Trans	66,590,204	69,467,447	79,314,908	
Trans fdr q < 0.05	982	1,009	1,106	219

Table 6. Troubleshooting.

Problem	Possible Cause	Solution
Low digestion efficiency	Higher amount of cells	Repeat the entire protocol using half of the formaldehyde fixed pellet
Low DNA amount	Loss during ethanol precipitation	Repeat the entire protocol making sure that the ethanol is added before the sodium acetate
No PCR products	Inefficient digestion/biotin filling/ligation or inefficient biotin pull down and/or adaptor ligation	Firstly, repeat the biotin pull down increasing the amount of DNA. If still no product is obtained repeat the entire protocol using half of the formaldehyde fixed pellet
Low valid di-tags percentage	Inefficient digestion/biotin filling/ligation	Repeat the entire protocol using half of the formaldehyde fixed pellet
Low enrichment	Poor experimental	Check reagents, re-do capture

efficiency	capture efficiency Incorrect baitmap file used	 Check baitmap file
Large proportion of reads do not align to reference genome	Experimental contamination from non-reference organism Incorrect bowtie indices used	Prepare new library, taking care to avoid contamination Check and download/create correct indices
Analysis pipeline steps will not complete	Insufficient available memory	Repeat command, making more memory available
HiCUP takes an excessive amount of time to run (>96 hours)	Very large read count will slow-down alignment stage if no parallelisation	If possible, consider parallelising by increasing thread number

Table 7. Timings and memory restrictions for major parts of CHi-C analysis. Unless otherwise stated these were based on analysis of one naïve B cell replicate (NB 34; 696,454,802 total reads (Javierre *et al.*, 2016)).

Process	**Time (hours)**	**Memory (Gb)**	**Comment**
Running HiCUP	~80	~50	For one replicate, single threaded
Estimating capture coverage	~2	~50	For one replicate
Preparing chinput file	~2	~50	For one replicate
Running CHiCAGO	~1	~150	For all biological replicates
Preparing GOTHiC input	~0.5	<10	For one replicate
Running GOTHiC	~20	~150	For one replicate

Acknowledgements
The work of GO, BK and RSH is funded by Cancer Research UK.

LITERATURE CITED

Bonev, B. and Cavalli, G., 2016. Organization and function of the 3D genome. *Nature reviews. Genetics*. 17(12):772.
Cairns, J*., et al.*, 2016. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome biology*. 17(1):127.
Dekker, J., 2008. Mapping in vivo chromatin interactions in yeast suggests an extended chromatin fiber with regional variation in compaction. *The Journal of biological chemistry*. 283(50):34532-34540.
Dekker, J. and Heard, E., 2015. Structural and functional diversity of Topologically Associating Domains. *FEBS Lett*. 589(20 Pt A):2877-2884.
Dekker, J., Marti-Renom, M.A. and Mirny, L.A., 2013. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature reviews. Genetics*. 14(6):390-403.
Dekker, J*., et al.*, 2002. Capturing chromosome conformation. *Science*. 295(5558):1306-1311.
Dixon, J.R*., et al.*, 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 485(7398):376-380.
Dryden, N.H*., et al.*, 2014. Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. *Genome research*. 24(11):1854-1868.
Hughes, J.R*., et al.*, 2014. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet*. 46(2):205-212.
Jager, R*., et al.*, 2015. Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nature communications*. 6:6178.
Javierre, B.M*., et al.*, 2016. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell*. 167(5):1369-1384 e1319.
Langmead, B., 2010. Aligning short sequencing reads with Bowtie. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]*. Chapter 11:Unit 11 17.
Langmead, B. and Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 9(4):357-359.
Li, H*., et al.*, 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25(16):2078-2079.
Lieberman-Aiden, E*., et al.*, 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 326(5950):289-293.
Mifsud, B*., et al.*, 2017. GOTHiC, a probabilistic model to resolve complex biases and to identify real interactions in Hi-C data. *PloS one*. 12(4):e0174744.
Mifsud, B*., et al.*, 2015. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet*. 47:598-606.
Mifsud, B*., et al.*, 2015. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature genetics*. 47(6):598-606.
Montefiori, L*., et al.*, 2016. Extremely Long-Range Chromatin Loops Link Topological Domains to Facilitate a Diverse Antibody Repertoire. *Cell Rep*. 14(4):896-906.
R Development Core Team. 2013. R: A language and environment for statistical computing.
Rao, S.S*., et al.*, 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 159(7):1665-1680.

Sahlen, P*., et al.*, 2015. Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. *Genome biology*. 16:156.
Servant, N*., et al.*, 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome biology*. 16:259.
Wingett, S*., et al.*, 2015. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res*. 4:1310.

KEY REFERENCES

Rao, S.S., et al., 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 159(7):1665-1680. **Describing the *in situ* protocol used in BASIC PROTOCOL 2,3,4 and 5.**

Mifsud, B.*, et al.*, 2015. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet*. 47:598-606. **Describing the promoter CHi-C protocol used in BASIC PROTOCOL 1 and 6.**

Wingett, S.*, et al.*, 2015. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res*. 4:1310. **Describing HiCUP software used in BASIC PROTOCOL 7.**

Cairns, J.*, et al.*, 2016. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome biology*. 17(1):127. **Describing CHiCAGO software used in BASIC PROTOCOL 8.**

Mifsud, B.*, et al.*, 2017. GOTHiC, a probabilistic model to resolve complex biases and to identify real interactions in Hi-C data. *PloS one*. 12(4):e0174744. **Describing GOTHiC software used in ALTERNATIVE PROTOCOL 1.**

INTERNET RESOURCES

https://www.genomics.agilent.com/en/SureSelect-DNA-Library-Preps-Mechanical-Shearing/ - home page for the SureSelect Target Enrichment System containing documentation relative to the capture technologies described in BASIC PROTOCOL 6

https://www.bioinformatics.babraham.ac.uk/projects/hicup/ - home page for HiCUP, containing documentation on installation and running of the software.

http://regulatorygenomicsgroup.org/chicago - home page for CHiCAGO, containing documentation on installation and running of the software.

http://www.bioconductor.org/packages/release/bioc/manuals/GOTHiC/man/GOTHiC.pdf - bioconductor manual for GOTHiC.

http://bowtie-bio.sourceforge.net/bowtie2 - home page for bowtie2, containing documentation on installation and running of the software as well as links to reference genomic indexes.

https://www.r-project.org/ - Website containing documentation and downloads for R.

https://epigenomegateway.wustl.edu/ - WashU epigenome browser – enables plotting of Hi-C and CHi-C contacts against a reference genome and biological annotation features (such as histone markers).

https://emea.illumina.com/ - home page for Illumina Inc. technologies for high-throughput sequencing