Video Article

Promoter Capture Hi-C: High-resolution, Genome-wide Profiling of Promoter Interactions

Stefan Schoenfelder*[1], Biola-Maria Javierre*[1,2], Mayra Furlan-Magaril[1,3], Steven W. Wingett[1,4], Peter Fraser[1,5]

[1]Nuclear Dynamics Programme, The Babraham Institute, Babraham Research Campus

[2]IJC Building, Campus ICO-Germans Trias i Pujol, Josep Carreras Leukemia Research Institute

[3]Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México

[4]Bioinformatics Group, The Babraham Institute, Babraham Research Campus

[5]Department of Biological Science, Florida State University

*These authors contributed equally

Correspondence to: Stefan Schoenfelder at stefan.schoenfelder@babraham.ac.uk, Biola-Maria Javierre at bmjavierre@carrerasresearch.org, Peter Fraser at peter.fraser@babraham.ac.uk

URL: https://www.jove.com/video/57320
DOI: doi:10.3791/57320

Keywords: Genetics, Issue 136, 3D genome organization, promoters, enhancers, long-range interactions, Hi-C, *cis*-regulatory elements, genome-wide association study (GWAS), non-coding disease variants, chromatin loops, CHiCAGO, HiCUP

Date Published: 6/28/2018

Citation: Schoenfelder, S., Javierre, B.M., Furlan-Magaril, M., Wingett, S.W., Fraser, P. Promoter Capture Hi-C: High-resolution, Genome-wide Profiling of Promoter Interactions. *J. Vis. Exp.* (136), e57320, doi:10.3791/57320 (2018).

Abstract

The three-dimensional organization of the genome is linked to its function. For example, regulatory elements such as transcriptional enhancers control the spatio-temporal expression of their target genes through physical contact, often bridging considerable (in some cases hundreds of kilobases) genomic distances and bypassing nearby genes. The human genome harbors an estimated one million enhancers, the vast majority of which have unknown gene targets. Assigning distal regulatory regions to their target genes is thus crucial to understand gene expression control. We developed Promoter Capture Hi-C (PCHi-C) to enable the genome-wide detection of distal promoter-interacting regions (PIRs), for all promoters in a single experiment. In PCHi-C, highly complex Hi-C libraries are specifically enriched for promoter sequences through in-solution hybrid selection with thousands of biotinylated RNA baits complementary to the ends of all promoter-containing restriction fragments. The aim is to then pull-down promoter sequences and their frequent interaction partners such as enhancers and other potential regulatory elements. After high-throughput paired-end sequencing, a statistical test is applied to each promoter-ligated restriction fragment to identify significant PIRs at the restriction fragment level. We have used PCHi-C to generate an atlas of long-range promoter interactions in dozens of human and mouse cell types. These promoter interactome maps have contributed to a greater understanding of mammalian gene expression control by assigning putative regulatory regions to their target genes and revealing preferential spatial promoter-promoter interaction networks. This information also has high relevance to understanding human genetic disease and the identification of potential disease genes, by linking non-coding disease-associated sequence variants in or near control sequences to their target genes.

Video Link

The video component of this article can be found at https://www.jove.com/video/57320/

Introduction

Accumulating evidence suggests that the three-dimensional organization of the genome plays an important functional role in a range of nuclear processes, including gene activation[1,2,3], repression[4,5,6,7,8], recombination[9,10], DNA repair[11], DNA replication[12,13], and cellular senescence[14]. Distant enhancers are found in close spatial proximity to the promoters they regulate[15,16,17], which is essential for proper spatio-temporal gene expression control. Enhancer deletions show that distal enhancers are essential for target gene transcription[18,19,20,21,22], and 'forced chromatin looping' demonstrates that engineered tethering between an enhancer and its target promoter in the *Hbb* locus is sufficient to drive transcriptional activation[23]. Further, genome rearrangements that bring genes under the control of ectopic enhancers can result in inappropriate gene activation and disease[24,25,26]. Together, these examples illustrate that promoter-enhancer interactions are essential for gene control and require tight regulation to ensure appropriate gene expression. The human and mouse genomes are each estimated to harbor around one million enhancers. For the vast majority of these enhancers, target genes are unknown, and the 'rules of engagement' between promoters and enhancers are poorly understood. Assigning transcriptional enhancers to their target genes thus remains a major challenge in deciphering mammalian gene expression control.

Our understanding of three-dimensional genome architecture has been revolutionized by the introduction of 3C[27] (chromosome conformation capture) and its variants[28,29,30,31]. The most powerful of these techniques, Hi-C (high throughput chromosome conformation capture) is designed to identify the entire ensemble of chromosomal interactions within a cell population. Hi-C libraries, typically generated from millions of cells, are highly complex with an estimated 10^{11} independent ligation products between ~4 kb fragments in the human genome[32]. As a consequence, reliable and reproducible identification of interactions between individual restriction fragments (such as those containing a promoter or enhancer)

Copyright © 2018 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

from Hi-C data is not feasible unless Hi-C libraries are subjected to ultra-deep sequencing, which is not an economically viable solution for laboratories preparing Hi-C libraries routinely. To circumvent this shortcoming, we developed Promoter Capture Hi-C to specifically enrich promoter-containing ligation products from Hi-C libraries. We focused on promoters for two reasons. First, promoter-enhancer contacts have been shown to be crucial for proper gene expression levels in numerous studies (see references above), and second, as promoters are largely invariant between cell types, the same capture bait system can be used to interrogate the regulatory circuitry across multiple cell types and conditions. Our approach relies on in-solution hybridization of Hi-C libraries with tens of thousands of biotinylated RNA 120mers complementary to promoter-containing Hi-C ligation products and subsequent capture on streptavidin-coated magnetic beads. This results in PCHi-C libraries with much reduced complexity compared to the original Hi-C library, focusing only on the identification of fragments that are ligated to promoters at significantly high frequencies.

We have used PCHi-C in a number of human and mouse cell types to contribute to a better understanding of gene expression control by uncovering long-range distal promoter interacting regions with putative regulatory function, as well as non-random promoter-promoter contacts in the three-dimensional space of the nucleus. The studies have mapped hundreds of thousands of promoter-enhancer contacts across numerous cell types[33,34,35,36,37,38,39], identified Polycomb Repressive Complex-mediated spatial genome organization in mouse embryonic stem cells[7], demonstrated large-scale rewiring of promoter interactomes during cellular differentiation[37,38,39], and linked non-coding disease-associated sequence variants to gene promoters[35].

PCHi-C is an ideally suited method to map the genome-wide ensemble of DNA sequences interacting with promoters. Related approaches, such as Capture Hi-C of continuous genomic regions (see **Discussion**) are the method of choice to obtain high-resolution interaction profiles for selected genomic regions. PCHi-C and Capture Hi-C are extremely similar from an experimental point of view (the only difference is the choice of capture system), so that the advice and guidelines we provide are applicable to both approaches. Here, we present a detailed description of PCHi-C. We outline the rationale and design of a PCHi-C experiment, provide a step-by-step PCHi-C library generation protocol, and illustrate how the quality of PCHi-C libraries can be monitored at various steps in the protocol to yield high-quality data.

Protocol

1. Formaldehyde Fixation

1. **Cell preparation: Start with a minimum of 2 x 10^7 cells per experiment.**
    1. For cells grown in culture, resuspend the cells in culture medium. For *ex vivo* cells, resuspend in 1x Dulbecco's Modified Eagle Medium (DMEM), supplemented with 10% (vol/vol) fetal bovine serum (FBS).
    2. For adherent cells, remove culture medium and add 30.625 mL of fresh medium with 10% (vol/vol) FBS at room temperature (RT; 20–25 °C).
    3. For suspension cells, collect and centrifuge cells at 400 x g and 20 °C for 3 min. Remove supernatant and re-suspend cell pellet in 30.625 mL of medium with 10% (vol/vol) FBS at RT.
    4. For solid tissues, use trypsin (0.05% to 2.5% final concentration, depending on cell type) or dounce homogenizing to obtain a single cell suspension. After this additional step, treat cells like suspension cells.

2. **Add 4.375 mL of 16% methanol-free paraformaldehyde (open ampoule just prior to use) to a final concentration of 2% (vol/vol). Fix for 10 min at RT with gentle mixing on a rocker.**
   CAUTION: Paraformaldehyde is a hazardous chemical. Follow the appropriate health and safety regulations.
3. Quench reaction by adding 5 mL of freshly-prepared 1 M ice-cold glycine. Mix for 5 min with gentle rocking at RT, and then incubate on ice for 15 min with occasional inverting.
4. **Wash and collect fixed cells.**
    1. For adherent cells, remove supernatant, add 10 mL of ice-cold 1x PBS pH 7.4 on the plate wall and remove it. Add 1 mL of ice-cold 1x PBS pH 7.4, collect cells using a cell scraper and transfer into a 50 mL tube. Repeat twice to collect as many cells as possible. Add ice-cold PBS up to 50 mL final volume.
    2. For suspension cells, centrifuge cells at 760 x g and 4 °C for 5 min, remove supernatant, and re-suspend cell pellet in 50 mL of ice-cold PBS pH 7.4.

5. Centrifuge cells at 400 x *g* and 4 °C for 10 min and carefully remove supernatant. The cell pellet can be snap frozen in liquid nitrogen and subsequently stored at -80 °C for several months.

2. Cell Lysis

1. Re-suspend cell pellet in 50 mL of freshly-prepared ice-cold lysis buffer (10 mM Tris-HCl pH 8, 0.2% (vol/vol) Igepal CA-630, 10 mM NaCl, and one tablet protease inhibitor cocktail) and mix. Incubate on ice for 30 min, mix occasionally by inverting. Centrifuge the nuclei at 760 x *g* and 4 °C for 5 min and remove supernatant.

3. *Hin*dIII Digestion

1. Wash cell nuclei with 1.25x restriction buffer 2. Re-suspend cell pellet in 1 mL of ice-cold 1.25x restriction buffer 2 and transfer into a 1.5 mL tube. Spin the nuclei at 760 x g and 4 °C for 5 min and remove supernatant.
2. Re-suspend cell pellet in 1790 µL of 1.25x restriction buffer 2. Make 5 aliquots, each containing 5–10 million cells in 358 µL of 1.25x restriction buffer 2.
3. Add 11 µL of 10% (wt/vol) SDS per aliquot and shake at 950 revolutions per min (rpm) for 30 min at 37 °C in a thermomixer. If cell clumps appear, dissociate by pipetting, avoiding bubbles.

Copyright © 2018 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

4. Add 75 µL of 10% Triton X-100 (vol/vol) per aliquot and shake at 950 rpm and 37 °C for 15 min in a thermomixer. If cell clumps appear, dissociate by pipetting, avoiding bubbles.
5. **Add 12 µL of 100 U/µL *Hin*dIII 100 (1,200 units in total) per aliquot and incubate at 37 °C overnight (O/N) while shaking at 950 rpm in a thermomixer.**
    1. For the digestion control, transfer 25 µL of sample (5 µL from each aliquot) in a new tube before adding the enzyme (undigested control) and repeat the same procedure after adding the enzyme (digested control). Incubate both tubes in the same manner as the Hi-C library.
6. On the following morning, add 5 µL of 100 U/µL *Hin*dIII (500 units in total) per aliquot and incubate at 37 °C for 2 h while shaking at 950 rpm in a thermomixer.
7. **Digestion control: for the digested and undigested controls (see 3.5.1), perform crosslink reversal (step 6), Phenol:Chloroform extraction, and DNA precipitation (step 7).**
    1. Design a pair of primers that span a *Hin*dIII site. In the same region, design another pair of primers that don't span a *Hin*dIII site. Design primers for quantitative PCR (Q-PCR) using Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/) and the following parameters:
    Primer size: Optimal 20 (Min.: 18, Max.: 27); Primer Tm: Optimal 60 (Min.: 57, Max.: 63);Primer CG% content: Min.: 20, Max.: 80;
    Amplicon size: RT-PCR ~100 bp (for conventional PCR ~300 bp); Mispriming library: human (human primers) or rodent and simple (mouse primers).
    2. Perform Q-PCR to obtain 4 mean Cts (threshold cycle): Ct[D;H], obtained from the digested sample [D] with the pair of primers that span a *Hin*dIII site [H]; Ct[D;-], obtained from the digested sample [D] with the pair of primers that don't span a *Hin*dIII site [-]; Ct[U;H], obtained from the undigested sample [U] with the pair of primers that span a *Hin*dIII site; Ct[U;-], obtained from the undigested sample [U] with the pair of primers that don't span a *Hin*dIII site [-]. Calculate the percentage of digestion as: % digestion = $100-100/2^{(Ct[D,H]-Ct[D,-]) - (Ct[U,H]-Ct[U,-])}$.

4. Biotinylation of Restriction Fragment Overhangs

1. Prepare biotinylation master mix: 30.6 µL of 10x restriction buffer 2, 10.2 µL of H_2O (molecular biology grade), 7.65 µL of 10 mM dCTP, 7.65 µL of 10 mM dGTP, 7.65 µL of 10 mM dTTP, 191.25 µL of 0.4 mM biotin-14-dATP, and 51 µL of 5,000 U/mL DNA polymerase I large (Klenow) fragment.
2. Add 60 µL of biotinylation master mix per aliquot, mix, and incubate at 37 °C for 1 h shaking at 700 rpm (thermomixer) for 5 s, every 30 s. After 1 h, place aliquots on ice.

5. In-nucleus Ligation

1. Prepare ligation master mix: 510 µL of 10x T4 DNA ligase buffer, 51 µL of 10 mg/mL Bovine Serum Albumin (100x BSA), 1754.4 µL of water (molecular biology grade), and 127.5 µL of 1 U/µL T4 DNA ligase (see **Table of Materials**).
2. Add 479 µL of ligation master mix per aliquot mix and incubate at 16 °C for 4 h shaking at 700 rpm for 5 s every 2 min in a thermomixer.
3. Incubate 30 min at RT.

6. Crosslink Reversal

1. Combine all aliquots in a 50 mL centrifuge tube (suitable for high-speed centrifugation).
2. Add 62.5 µL of 10 mg/mL RNase A, mix, and incubate for 30 min at 37 °C.
3. Add 300 µL of 10 mg/mL Proteinase K, mix, and incubate for 30 min at 37 °C.
4. Incubate reaction O/N (or at least 4 h) at 65 °C. On the following morning, add 300 µL of 10 mg/mL Proteinase K, mix, and incubate for 1 h at 65 °C.

7. DNA Purification

1. Add 4337.5 µL of TLE buffer (10 mM Tris-HCl pH 8.0; 0.1 mM EDTA pH 8.0) and mix.
2. Add 1 volume (10 mL) phenol pH 8.0, vortex for 10 s, and centrifuge at RT and 20,000 x *g* for 3 min. Transfer 9 mL of the upper (aqueous) phase to a new 50 mL tube.
    CAUTION: Phenol is a hazardous chemical . Follow the appropriate health and safety regulations.
3. Add 2 mL of TLE buffer to the remaining aqueous phase, vortex for 10 s and centrifuge at RT and 20,000 x *g* for 3 min. Transfer 2.5 mL of the aqueous phase into the new tube from step 7.2, making the final volume 11.5 mL. Discard tube containing the lower (organic) phase.
4. Add 1 volume (11.5 mL) of phenol:chloroform:isoamyl alcohol (25:24:1), vortex for 10 s, and centrifuge at RT and 20,000 x *g* for 3 min. Transfer 11 mL of the upper (aqueous) phase to a new 50 mL tube. Repeat step 7.3. The total sample volume will now be 13.5 mL.
5. Add 1.35 mL of 3 M sodium acetate pH 5.2 and 33.75 mL of ice cold 100% ethanol, mix, and incubate at -80 °C for 45 min, or alternatively overnight at -20 °C.
6. Centrifuge at 4 °C and 20,000 x *g* for 10 min, remove supernatant, re-suspend pellet in 1 mL of freshly-prepared 70% (vol/vol) ethanol, and transfer to a new tube.
7. Centrifuge at 4 °C and at full speed for 3 min in a benchtop centrifuge, then remove supernatant.
8. Re-suspend pellet in 1 mL of ice cold 70% (vol/vol) ethanol and repeat step 7.7. Dry the pellet at 37 °C for 10 min and re-suspend in 650 µL of TLE buffer. Determine the DNA yield by using a fluorescence-based assay to quantify double-stranded DNA.
    NOTE: The protocol can be paused here by snap freezing and storing the sample at -80 °C for several months or at -20 °C for a shorter period of time.

Copyright © 2018 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

8. Quality Controls

1. Monitor library integrity and ligation by DNA electrophoresis. Run 200 ng of library on a 0.8% agarose/1x TBE gel. The DNA should run as a band over 10 kb.
2. Detect known cell-type invariant short- and long-range interactions by conventional PCR. Use 100 ng of template DNA per PCR reaction. Design the PCR primers close and towards the restriction sites following the instructions above (see 3.7.1). Primer sequences for quality control of mouse and human Hi-C libraries are listed in **Table 1**.
3. **Fill-in and ligation control: Cut out the gel bands containing the amplicons from control 8.2, gel-extract DNA, and use the DNA as template for 4 individual PCR reactions with identical primer combinations.**
    1. Purify amplicons using a PCR purification kit and quantify the DNA concentration.
    2. Prepare four digestion reactions (*Hin*dIII [a], NheI [b], *Hin*dIII + NheI [c] and no enzyme [d]) for each amplicon in a final volume of 15 µL: 500 ng of amplicon, 1.5 µL of 10x restriction buffer 2.1, 0.15 µL of 10 mg/mL Bovine Serum Albumin (100x BSA), and 0.1 µL (10 units) of enzyme (*Hin*dIII [a], NheI [b], *Hin*dIII + NheI [c] or water [d]).
    3. Digest for 1 h at 37 °C, then run digestion reactions on a 1.5% (wt/vol) agarose/1x TBE gel.

9. DNA Fragmentation

1. Transfer 50.5 µg of sample in a new tube and add TLE buffer to a final volume of 655 µL. Split sample into 5 sonication vials (see **Table of Materials**) by adding 130 µL of library (10 µg) to each vial. Shear to a size of ~400 bp in an ultra-sonicator (see **Table of Materials**) using the following parameters: duty factor: 10%; peak incident power (w): 140; cycles per burst: 200; time: 55 s.
2. Collect sonicated sample in a fresh 2 mL tube.

10. Double-sided SPRI-bead Size Selection

1. Mix SPRI (Solid Phase Reversible Immobilization) bead solution well by inverting, transfer 1.85 mL of bead solution to a new tube and bring to RT for 15 min.
2. Add 350 µL of water (molecular biology grade) to the sample (final volume 1 mL).
3. Add 600 µL of SPRI bead solution to the sample (total volume 1.6 mL; ratio of SPRI solution to DNA: 0.6 to 1), incubate for 5 min at RT, and spin sample in a benchtop centrifuge for 2–3 s to collect sample.
4. Open the lid, place the sample on the magnetic separation stand for 5 min, transfer clear supernatant into a new tube and discard beads.
5. Concentrate SPRI beads for the second size selection step: Transfer 930 µL of SPRI beads into a new tube, place on the magnetic separation stand for 5 min and discard clear supernatant. Re-suspend the beads in 310 µL of SPRI bead solution.
6. Add 300 µL of concentrated SPRI beads (step 10.5) to the sample (total volume 1.9 mL; ratio SPRI solution to DNA is now 0.9 to 1), incubate at RT for 5 min, and spin sample in a benchtop centrifuge for 2–3 s. Carefully open the lid, place the tube on the magnetic separation stand for 5 min, and discard supernatant.
7. Add 1 mL of freshly prepared 70% ethanol (vol/vol) to the sample tube on the magnetic separation stand, incubate for 30 s, and discard supernatant. Repeat twice.
8. Dry beads at 37 °C in a thermomixer (tube lid open) for no more than 5 min. Add 300 µL of TLE buffer to the sample, mix, and incubate for 10 min at room temperature.
9. Spin sample in a benchtop centrifuge for 2–3 s, open the lid and place the tube on the magnetic separation stand for 5 min. Transfer clear supernatant into a new tube and discard beads.

11. Biotin/Streptavidin Pull-down of Ligation Products

1. Prepare buffers: 1x TB buffer (5mM Tris-HCl pH 8.0; 0.5mM EDTA; 1 M NaCl; 0.05% Tween 20); 2x NTB buffer (10 mM Tris-HCl pH 8.0; 1 mM EDTA; 2 M NaCl); 1x NTB buffer (5 mM Tris-HCl pH 8.0; 0.5 mM EDTA; 1 M NaCl).
2. Add 200 µL of magnetic streptavidin-coupled beads (see **Table of Materials**) into a new tube, place it on the magnetic separation stand for 1 min and remove supernatant.
3. **Wash beads twice with 500 µL of 1x TB buffer.**
    1. For each wash step during the biotin pull-down, end repair and removal of biotin at non-ligated DNA ends, dATP tailing, and adapter ligation steps, re-suspend the beads in the corresponding buffer, rotate at RT and 15 rpm for 3 min, spin the tube in a benchtop centrifuge for 2–3 s, place the tube on the magnetic separation stand for 3 min and remove supernatant.
4. Re-suspend beads in 300 µL of 2x NTB buffer. Mix beads and sample (600 µL total volume) and incubate at RT for 15 min on a rotating wheel at 3 rpm.
5. Reclaim beads on the magnetic separation stand for 3 min and remove the clear supernatant. Wash beads twice in 500 µL of 1x NTB buffer first and then in 200 µL of 1x ligation buffer. Re-suspend the beads in 50 µL of 10x ligation buffer.

12. End Repair and Removal of Biotin at Non-ligated DNA Ends

1. Combine the sample (50 µL in total) with 50 µL of 2.5 mM dNTP mix (12.5 µL of 10 mM of each dNTP), 18.1 µL of 3,000 U/mL T4 DNA Polymerase, 18.1 µL of 10,000 U/mL T4 PNK, 3.7 µL of 5,000 U/mL DNA polymerase I large (Klenow) fragment, and 360.1 µL of H_2O.
2. Mix and incubate at 20 °C for 1 h, shaking 5 s at 700 rpm every 2 min in a thermomixer.
3. Reclaim beads on the magnetic separation stand, remove the clear supernatant, and wash beads twice in 500 µL of 1x TB buffer.
4. Wash beads in 500 µL of 1x NTB buffer, followed by one wash in 500 µL of 1x TLE .

Copyright © 2018 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

5. Reclaim beads on the magnetic separation stand, remove the clear supernatant, and re-suspend beads in 415 µL of 1x TLE buffer.

13. dATP Tailing

1. Combine sample (415 µL) with 50 µL of 10x restriction buffer 2, 5 µL of 10 mM dATP, and 30 µL of 5 U/µL Klenow exo-minus.
2. Mix and incubate at 37 °C for 30 min, shaking 5 s at 700 rpm every 2 min in a thermomixer.
3. Reclaim beads on the magnetic separation stand, remove the clear supernatant, and wash beads twice in 500 µL of 1x TB buffer.
4. Wash beads in 500 µL of 1x NTB buffer.

14. Adapter Ligation

1. Wash beads in 200 µL of 1x ligation reaction buffer (see **Table of Materials**).
2. Re-suspend beads in 200 µL of 1x ligation reaction buffer. Add 4 µL of DNA ligase (see **Table of Materials**) and 16 µL of 15 µM pre-annealed PE adapters (pre-anneal the PE adapters by mixing equal volumes of PE adapter 1 and PE adapter 2 (both at 30 µM) and incubating for a few minutes at RT). Incubate at RT for 15 min.
3. Reclaim beads on the magnetic separation stand, remove the clear supernatant, and wash beads twice in 500 µL of 1x TB buffer.
4. Wash beads in 500 µL of 1x NTB buffer. Then, wash beads in 100 µL of 1x restriction buffer 2, re-suspend beads in 50 µL of 1x restriction buffer 2, and transfer into a new tube.

15. Hi-C Library Amplification

1. Prepare PCR master mix: 100 µL of 5x Phusion buffer; 6 µL of 25 µM PE PCR primer 1.0; 6 µL of 25 µM PE PCR primer 2.0; 14 µL of dNTP mix (10 mM each); 6 µL of Phusion polymerase; 318 µL of H_2O.
2. Mix PCR master mix with the beads (500 µL in total), divide in 10 aliquots of 50 µL, and amplify by PCR using the following conditions:
   30s at 98 °C
   7 cycles of: 10 s at 98 °C; 30 s at 65 °C; 30 s at 72 °C
   7 min at 72 °C
3. Collect PCR reactions into a new tube, reclaim beads on the magnetic separation stand, and transfer supernatant (500 µL) into a new tube.
4. **Purify the library DNA using SPRI beads.**
   1. Mix SPRI beads, transfer 460 µL of beads in a new tube, and bring to RT for 15 min. Add 450 µL of SPRI beads to the PCR reactions (final volume 950 µL), incubate for 5 min at RT, and spin sample in a benchtop centrifuge for 2–3 s to collect sample.
   2. Open the lid, place the sample on the magnetic separation stand for 5 min, and remove supernatant.
   3. Keeping the beads on the magnetic separation stand, add 1 mL of 70% ethanol (vol/vol) to sample tube over an area clear of beads, leave for 30 s, and discard supernatant.
   4. Repeat step 15.4.3 twice more.
   5. Dry beads at 37 °C in a thermomixer (tube lid open) for no more than 5 min.
   6. Add 51 µL of TLE buffer to the sample, mix, and incubate for 10 min at 37 °C, shaking at 950 rpm in a thermomixer.
   7. Spin sample in a benchtop centrifuge for 2–3 s, open the lid and place the tube on the magnetic separation stand for 5 min. Transfer clear supernatant into a new tube and discard beads.
   8. Quantify the concentration of the Hi-C library. After 7 rounds of PCR amplification, we routinely obtain 500–1,500 ng of Hi-C library.

16. Hybrid In-solution Capture

NOTE: Blocker and buffer (SHS1-4) solutions below are from the SureSelect kit (see **Table of Materials**).

1. Transfer 500 ng to 1 µg of Hi-C library into a new tube and evaporate sample on a vacuum Concentrator (see **Table of Materials**; 45 °C; vacuum pressure: level 30.0, ramp 5) until dry.
2. Re-suspend evaporated Hi-C library by adding 3.6 µL of H_2O (molecular biology grade), 2.5 µL of blocker 1, 2.5 µL of blocker 2, and 0.6 µL of custom blocker.
3. Transfer sample into a well of a new PCR tube strip, close with a PCR cap strip and place on ice. Label as "D" (for Hi-C DNA).
4. Prepare the hybridization buffer: 12.5 µL of SHS1 buffer; 0.5 µL of SHS2 buffer; 5 µL of SHS3 buffer; 6.5 µL of SHS4 buffer.
5. Incubate at 65 °C for 5 min in a thermomixer. Transfer into a well of a new PCR tube strip, close with a PCR cap strip and keep at RT. Label as "H" (for hybridization buffer).
6. Into a well of a new PCR tube strip, mix 5 µL of 100 ng/µL biotinylated RNA probes (store at -80 °C and thaw on ice just before use); 0.5 µL of SRNase B (RNase inhibitor) and 1.5 µL of H_2O (molecular biology grade).
7. Close the PCR tube strip with a PCR cap strip and place on ice. Label as "R" (for RNA).
8. Set up PCR machine using the following parameters:
   5 min at 95 °C; 25 h at 65 °C; lid heated; 29 µL PCR reaction volume.
   NOTE: Proceed as quickly as possible during all procedures while the PCR machine is running in order to avoid sample evaporation.
9. Place the "D" PCR tube strip in the PCR machine, close the PCR machine lid, and start the PCR reaction. When the PCR program reaches 65 °C, open the PCR machine lid and place the "H" PCR tube strip in the PCR machine. Close the PCR machine lid and incubate for 3 min. Open the PCR machine lid, place the "R" PCR tube strip on the PCR machine, and close the PCR machine.
10. After 2 min, open the PCR machine lid and all PCR tube strips. Transfer 13 µL of well "H" into well "R", then all volume of well "D" into well "R". Pipet up and down 3 times to mix the reaction, close the PCR tube strip, remove the "H" and "D" PCR tube strips, and close PCR machine lid. Incubate the reaction at 65 °C for 24 h.

Copyright © 2018 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

17. Isolation of Promoter Fragment-containing Ligation Products

NOTE: The following steps are recommended to be done with SureSelect adapter kit and library (see **Table of Materals**).

1. Pre-warm 1.5 mL of wash buffer 2 per sample at 65 °C in advance.
2. Add 60 µL of streptavidin-coupled magnetic beads (see **Table of Materials**) into a new tube, place on the magnetic separation stand for 1 min and remove supernatant.
3. Wash beads three times with 200 µL of 1x binding buffer.
   NOTE: For each wash step during the post-capture isolation of promoter-containing ligation products, re-suspend beads in the corresponding buffer, rotate for 3 min at RT and 15 rpm on a rotating wheel, softly spin the tube in a benchtop centrifuge for 2–3 s to collect sample, place the tube on the magnetic separation stand for 3 min, and remove supernatant.
4. Re-suspend beads in 200 µL of 1x binding buffer. Open the PCR machine and the PCR tube strip (while the PCR program is still running) and transfer the hybridization reaction into the tube with the magnetic beads. Incubate at RT for 30 min on a rotating wheel at 3 rpm.
5. Reclaim beads on the magnetic separation stand and remove the clear supernatant. Re-suspend beads into 500 µL of wash buffer 1, mix, and incubate for 15 min at 20 °C while shaking at 950 rpm in a thermomixer.
6. Reclaim beads on the magnetic separation stand for 3 min and remove the clear supernatant. Re-suspend beads into 500 µL of wash buffer 2, mix and incubate 10 min at 65 °C while shaking at 950 rpm in a thermomixer. Repeat step 17.5 twice more.
7. Reclaim beads on the magnetic separation stand, remove the clear supernatant and re-suspend beads in 200 µL of 1x restriction buffer 2. Reclaim beads on the magnetic separation stand, remove supernatant and re-suspend beads into 30 µL of 1x restriction buffer 2.

18. PCHi-C Library Amplification

1. Prepare PCR master mix: 60 µL of 5x PCR buffer (Phusion buffer), 3.6 µL of 25 µM PE PCR primer 1.0, 3.6 µL of 25 µM PE PCR primer 2.0, 8.4 µL of dNTP mix (10 mM each), 3.6 µL of Phusion polymerase, and 190.8 µL of H_2O.
2. Mix PCR master mix with the beads (300 µL in total), divide in 6 aliquots of 50 µL, and PCR-amplify using the following conditions:
   30 s at 98 °C
   4 cycles of: 10 s at 98 °C, 30 s at 65 °C, 30 s at 72 °C
   7 min at 72 °C
3. Collect all PCR reactions in a new tube, reclaim the beads on the magnet, and transfer supernatant (300 µL; contains PCHi-C library) in to a new tube.
4. Purify the PCHi-C library using SPRI beads, following the steps described above under 15.4.
5. Quantify the concentration of the PCHi-C library.

Representative Results

Promoter Capture Hi-C has been used to enrich mouse[7,34,36,39] and human[33,35,37,38] Hi-C libraries for promoter interactions. A similar protocol (named HiCap) has been described by the Sandberg group[40]. **Figure 1A** shows the schematic workflow for Promoter Capture Hi-C. In the protocol described here, Hi-C libraries are generated using in-nucleus ligation[41], which results in a significantly reduced number of spurious ligation products[42]. For PCHi-C, highly complex mouse or human Hi-C libraries are subjected to in-solution hybridization and capture using 39,021 biotinylated RNAs complementary to 22,225 mouse promoter-containing HindIII restriction fragments, or 37,608 biotinylated RNAs targeting 22,076 human promoter-containing HindIII restriction fragments, respectively. Promoter containing restriction fragments can be targeted at either or both ends by individual biotinylated RNAs (**Figure 1B**). We found that capture of both ends improved coverage of individual promoters (**Figure 1C**; raw sequence reads) nearly two-fold, as expected. Thus, whenever possible (*i.e.,* in non-repetitive regions), we advise to use biotinylated RNAs complementary to both ends of a restriction fragment to be captured.

To assess PCHi-C library quality at an early stage during library preparation, we perform two controls after DNA ligation and purification, as previously described[31]. The first is to use specific primer pairs to amplify ligation products as in 3C[27]. We use primer pairs (**Table 1**) to amplify cell-type invariant long-range ligation products, such as between the *Myc* gene and its known enhancers located approximately 2 Mb away (**Figure 2A**) or between genes of the *Hist1* locus (separated by 1.5 Mb), and between two regions located in close linear proximity ('short-range control').

The second quality control is carried out to determine the efficiency of biotin incorporation during Klenow-mediated fill-in of restriction site overhangs with biotin-dATP. Successful Klenow fill-in and subsequent blunt-end ligation results in the disappearance of the original restriction site between the DNA molecules of a ligation product, and in the case of HindIII in the formation of a new NheI recognition site (**Figure 2B**). The ratio of the HindIII to NheI digested ligation product is a direct readout of biotin incorporation efficiency. A poor quality Hi-C library will show a high level of HindIII digestion, whereas high-quality libraries have near-complete NheI digestion of ligation products (**Figure 2B**).

After Hi-C library preparation (*i.e.,* after biotin-streptavidin pull down of size-selected Hi-C ligation products, adapter ligation and pre-capture PCR), the integrity and size distribution of the Hi-C library is assessed by Bioanalyzer (**Figure 2C**). The same control is carried out at the end of PCHi-C library preparation (*i.e.,* after hybridization capture of promoter-containing ligation products and post-capture PCR). Comparison of the Hi-C and PCHi-C Bioanalyzer profiles shows that as expected, Hi-C libraries are much more concentrated than the corresponding PCHi-C libraries, but the size distribution of the libraries is highly similar, indicating that the capture step in PCHi-C does not introduce a size bias (**Figure 2C**, **D**).

Copyright © 2018 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

After paired-end sequencing, the PCHi-C reads are mapped, quality controlled and filtered using the HiCUP pipeline[43]. High-quality PCHi-C libraries contain between 70-90% 'valid pairs' (*i.e.,* paired-end sequence reads between two restriction fragments that are not neighboring on the linear genomic map; **Figure 3A**, **B**). Using the in-nucleus ligation protocol[41,42], the percentage of *trans* read pairs (*i.e.,* paired-end sequence reads between two restriction fragments that are located on different chromosomes) are usually low, between 5 and 25%, reflecting the existence of chromosome territories, and indicating good library quality. Direct comparison of the percentage of 'valid pairs' between Hi-C libraries and their corresponding PCHi-C libraries[35], shows that in all cases the percentage of valid pairs is higher in the PCHi-C libraries (**Figure 3B**). This is accompanied by a reduction in the percentage of non-valid 'same fragment internal' reads in PCHi-C (**Figure 3C**). This is expected, as the capture step not only enriches for promoter-containing ligation products, but also for restriction fragment ends, due to the position of the capture oligos on the restriction fragments (see **Figure 1B**).

After HiCUP filtering, we determine the capture efficiency. PCHi-C libraries contain three types of valid sequence reads after HiCUP filtering:
1.) Promoter: genome reads (*i.e.,* reads between a captured promoter fragment and a non-promoter HindIII restriction fragment anywhere in the genome)
2.) Promoter: promoter reads (reads between two captured promoter fragments)
3.) Genome: genome reads (background Hi-C ligation products where neither of the ligation product partners maps to a captured promoter). These are discarded prior to downstream analyses.

High-quality PCHi-C libraries have capture efficiencies (sum of categories 1 and 2 above) between 65–90% (**Figure 3D**). A direct comparison to Hi-C libraries shows that PCHi-C results in a ~15-fold enrichment for promoter-containing ligation products (**Figure 3D**), in some cases 17-fold. This is close to the hypothetical maximum (19.6-fold) enrichment for PCHi-C, which is dependent on the percentage of the genome restriction fragments covered by the capture system. Greater enrichment can be achieved by designing capture systems targeting fewer restriction fragments[44,45,46].

Analysis of promoter interactomes demonstrates cell type and lineage-specificity[33,34,35], with pronounced changes during cellular differentiation[37,38,39]. **Figures 4** and **5** show examples of lineage specificity and differentiation dynamics at specific promoters. For example, *ALAD* is constitutively expressed in all cells but its expression is upregulated in erythroblasts[47]. The *ALAD* promoter contacts several distal fragments in all hematopoietic cells and engages in additional interactions specifically in erythroblasts (**Figure 4**). *IL-8* shows no statistically significant interactions in B cells, very few interactions in T cells, but dozens of interactions in cells of the myeloid lineage, including cell-type specific interactions in monocytes, neutrophils and megakaryocytes (**Figure 5**). These examples demonstrate how PCHi-C can be used to unravel cell-type specific interactomes and identify promoter-interacting regions with regulatory potential.

Copyright © 2018 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

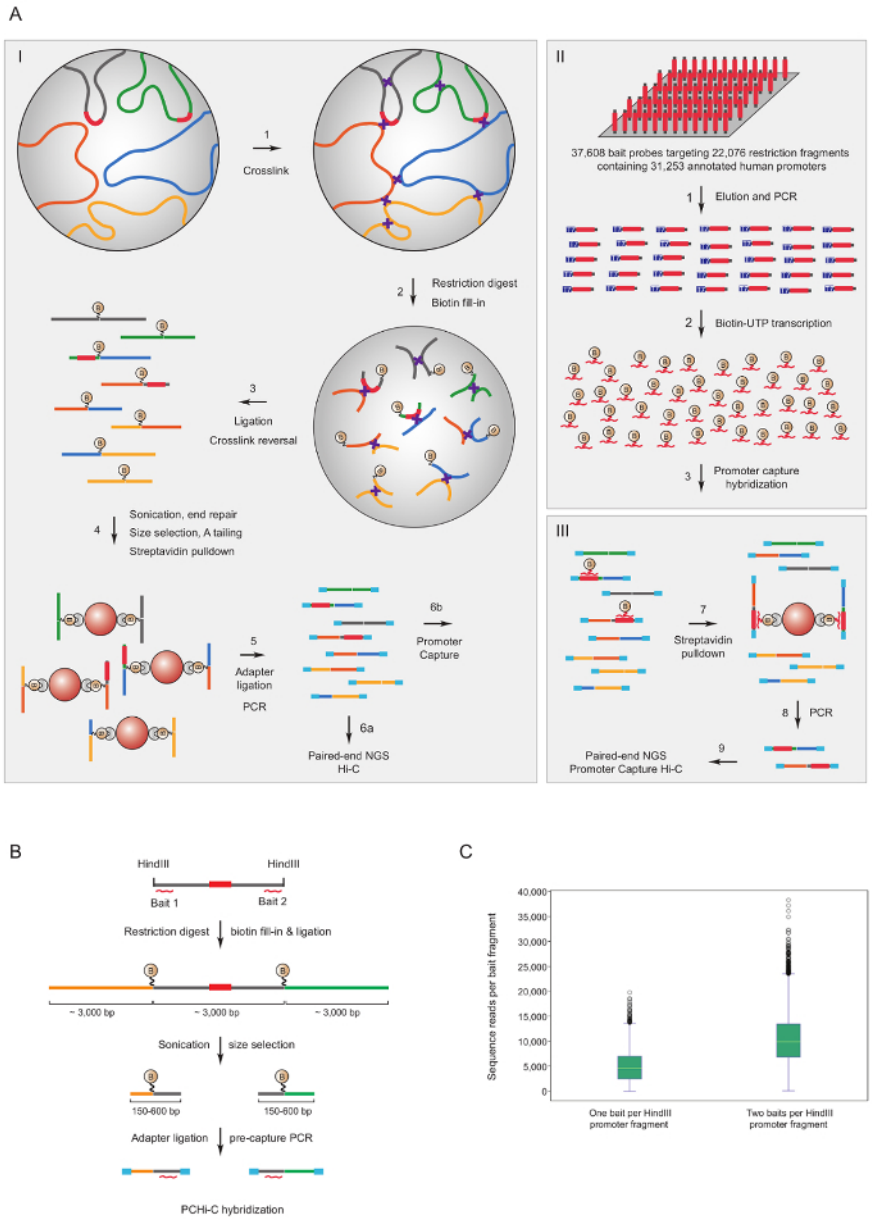

Figure 1: Promoter Capture Hi-C rationale and capture bait design. (A) Schematic workflow of PCHi-C. In-nucleus ligation Hi-C[41,42] (I) is followed by in-solution hybridization with biotinylated RNA baits (II) targeting the restriction fragments of all human (depicted here) or mouse gene promoters (III). **(B)** Bait design for PCHi-C. Biotinylated RNA capture baits (red curved lines) are designed against the ends of promoter-containing restriction fragments (grey; note that the promoter sequences themselves (red) are only targeted by the RNA capture baits if they are located at restriction fragment ends). Ligation products consisting of promoter-containing restriction fragments (grey) and their interacting restriction fragments (yellow and green) are isolated through sequence-complementarity hybridization between RNA bait and DNA target, and subsequent biotin-streptavidin pulldown, as shown in A. **(C)** Comparison of PCHi-C capture efficiency for promoter-containing restriction fragments targeted by one RNA bait capture probe vs two RNA bait capture probes (see schematic in B). Please click here to view a larger version of this figure.

Copyright © 2018 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

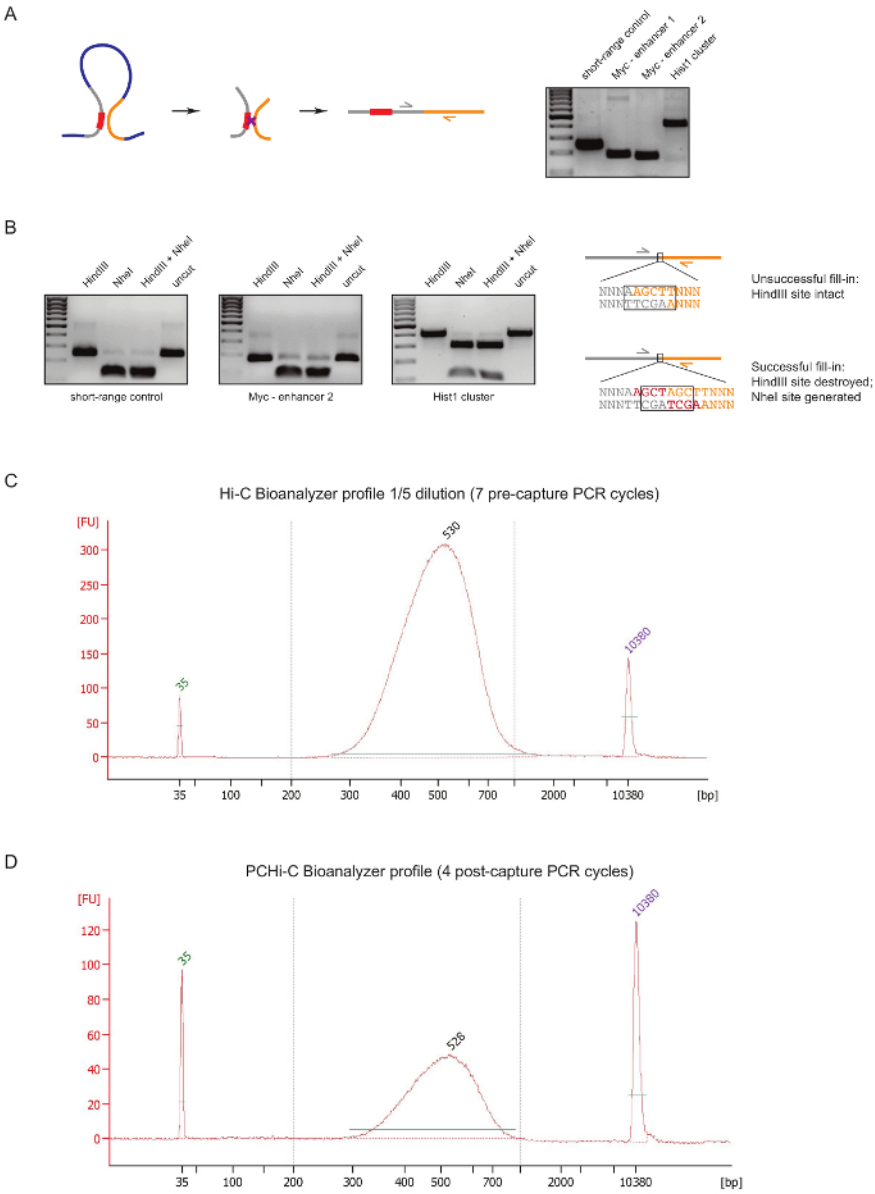

Figure 2: PCHi-C pre-sequencing quality controls. (A) Left, schematic of spatial juxtaposition between promoter and PIR, resulting in a Hi-C ligation product consisting of a promoter-containing restriction fragment (grey; promoter sequence in red) and a PIR restriction fragment (yellow). Right, DNA gel electrophoresis showing examples of Hi-C ligation products amplified using specific primer pairs (as depicted in schematic on the left). **(B)** Left, representative examples of HindIII, NheI and HindIII/NheI restriction digests of Hi-C ligation products (PCR products shown in A). Right, schematic of DNA sequence after Hi-C ligation following unsuccessful (top) or successful (bottom) dNTP Klenow fill-in of restriction junctions and subsequent ligation. **(C)** Representative Hi-C library bioanalyzer profile (1/5 dilution). **(D)** Representative PCHi-C library bioanalyzer profile (no dilution). Please click here to view a larger version of this figure.

Copyright © 2018 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

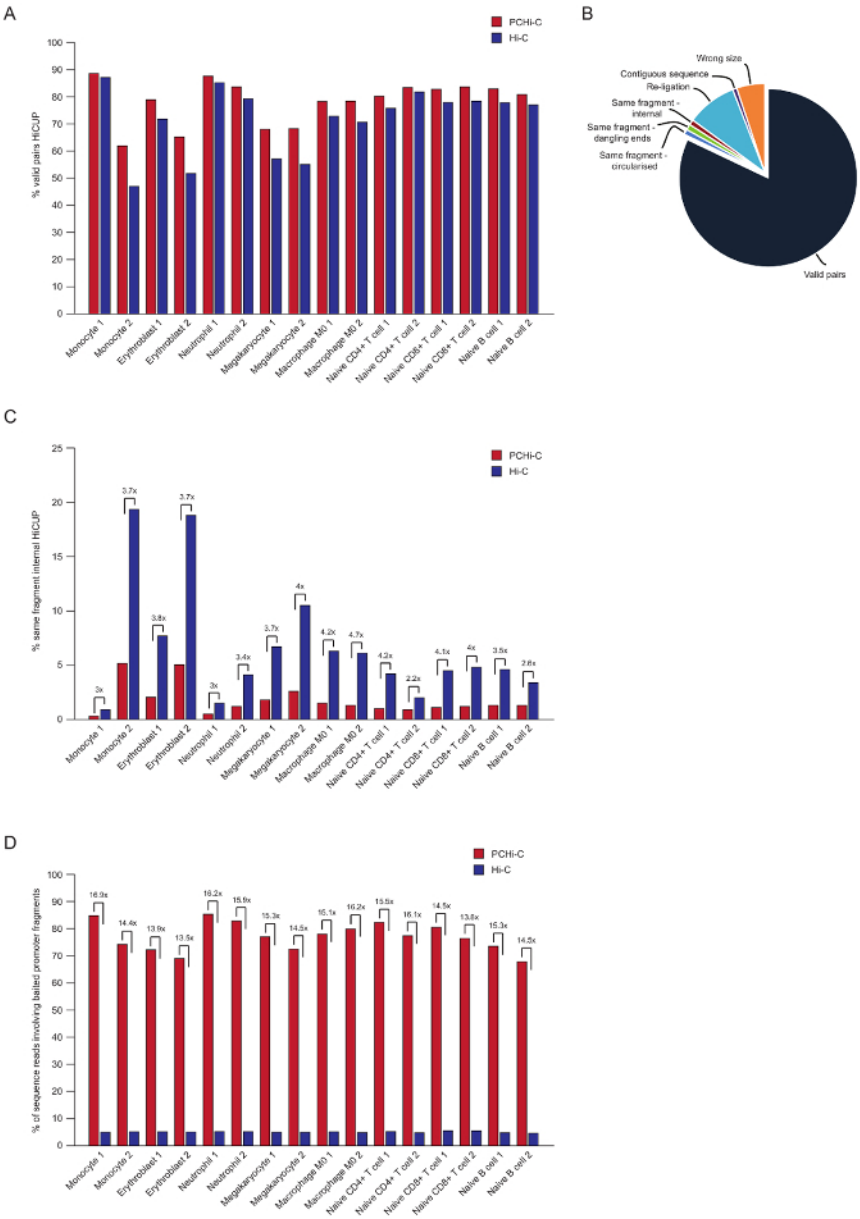

Figure 3: PCHi-C post-sequencing quality controls. (A) Comparison of percentage valid sequence read pairs after HiCUP[43] processing in PCHi-C vs corresponding Hi-C libraries (data from Javierre *et al.*, 2016[35]). **(B)** Representative HiCUP PCHi-C result showing valid read pairs, and other sequence categories that are discarded prior to downstream analyses (data from Javierre *et al.*, 2016[35]). **(C)** Comparison of percentage 'same fragment internal' reads after HiCUP processing in PCHi-C vs corresponding Hi-C libraries (data from Javierre *et al.*, 2016[35]). **(D)** Comparison of percentage sequence reads involving baited promoter fragments (capture efficiency) in PCHi-C vs corresponding Hi-C libraries (data from Javierre *et al.*, 2016[35]). Please click here to view a larger version of this figure.

Copyright © 2018 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

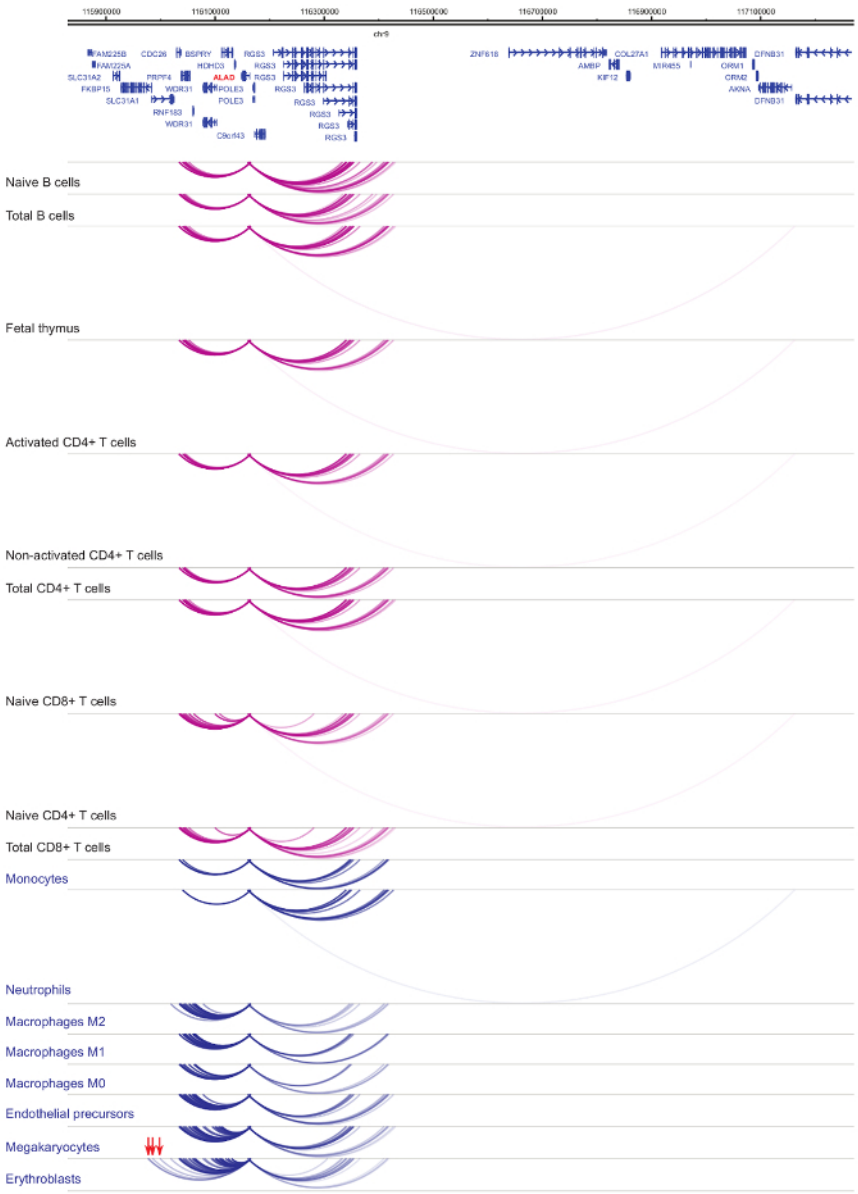

Figure 4: ***ALAD* PCHi-C profile in human hematopoietic cells.** Promoter interactions of myeloid cell types are shown as blue arches, and promoter interactions of lymphoid cell types are shown as purple arches. Erythroblast-specific interactions are indicated by red arrows (data from Javierre *et al.*, 2016[35]). Please click here to view a larger version of this figure.

Copyright © 2018 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

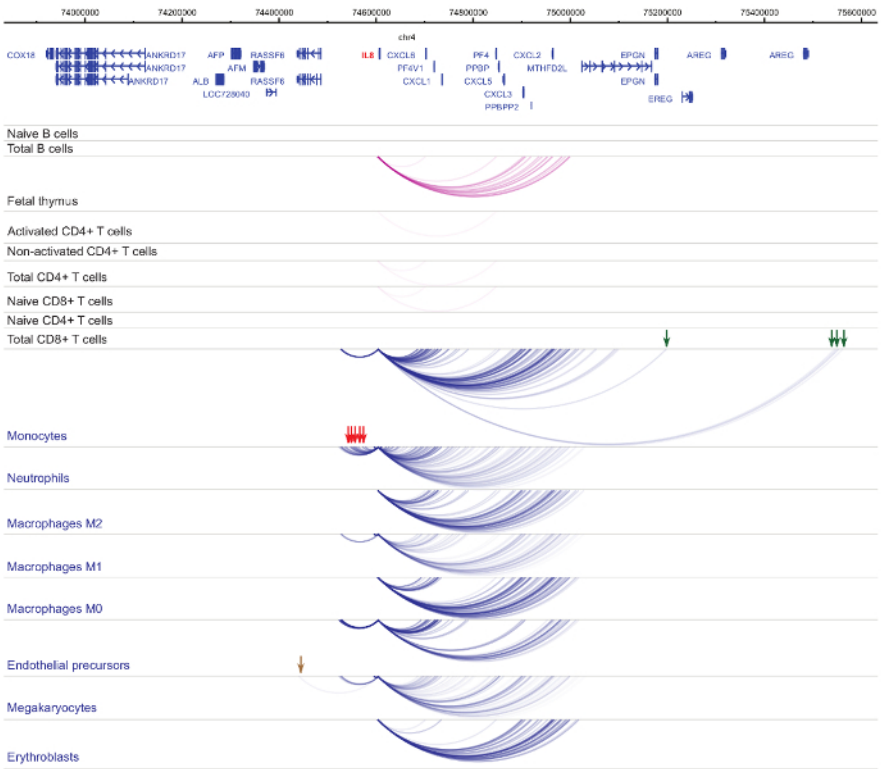

Figure 5: *IL8* PCHi-C profile in human hematopoietic cells. Promoter interactions of myeloid cell types are shown as blue arches, and promoter interactions of lymphoid cell types are shown as purple arches. Monocyte-specific interactions are indicated by green arrows, neutrophil-specific interactions are indicated by red arrows, and a megakaryocyte-specific interaction is indicated by a brown arrow (data from Javierre *et al.*, 2016[35]). Please click here to view a larger version of this figure.

Copyright © 2018 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

Human						
Name	Sequence	Chromosome	Strand	Start GRCh38/ hg38	End GRCh38/ hg38	Primer combinations to test 3C interactions and biotin incorporation
hs AHF64 Dekker	GCATGCATTAGCCTCTGCTGTTCTCTGAAATC	11	+	116803960	116803991	use in combination with hs AHF66 Dekker
hs AHF66 Dekker	CTGTCCAAGTACATTCCTGTTCACAAACCC	11	+	116810219	116810248	use in combination with hs AHF64 Dekker
hs MYC locus	GGAGAACCGGTAATGGCAAA	8	-	127733814	127733833	use in combination with hs MYC +1820 or hs MYC -538
hs MYC +1820	AAAATGCCCATTTCCTTCTCC	8	+	129554527	129554547	use in combination with hs MYC locus
hs MYC -538	TGCCTGATGGATAGTGCTTTC	8	-	127195696	127195716	use in combination with hs MYC locus
hs HIST1 F	AAGCAGGAAAAGGCATAGCA	6	+	26207174	26207193	use in combination with hs HIST1 R
hs HIST1 R	TCTTGGGTTGTGGGACTTTC	6	+	27771575	27771594	use in combination with hs HIST1 F
Mouse						
	Sequence	Chromosome	Strand	Start GRCm38/ mm10	End GRCm38/ mm10	Primer combinations to test 3C interactions and biotin incorporation
	TCATGAGTTCCCCACATCTTTG	8	+	84841090	84841111	use in combination with mm Calr2
	CTGTGGGCACCAGATGTGTAAAT	8	+	84848519	84848541	use in combination with mm Calr1
	TATCAAGGGTGCCCGTCACCTTCAGC	6	+	125163098	125163123	use in combination with Gapdh4 Dekker
	GGGCTTTTATAGCACGGTTATAAAGT	6	+	125163774	125163799	use in combination with Gapdh3 Dekker
	GGAGGAGGGAAAAGGAGTGATT	6	+	52212829	52212850	use in combination with mm Hoxa13

Copyright © 2018 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

	CAGGCATTATTTGCTGAGAACG	6	-	52253490	52253511	use in combination with mm Hoxa7
	GGGTAATGGTGTCACTAACTGG	13	+	23571284	23571305	use in combination with mm Hist1h3e or mm Hist1h4i
	GGGTTTGATGAGTTGGTGAAG	13	+	23566541	23566561	use in combination with mm Hist1h2ae
	TTGGGCCAAAGCCTATATGA	13	+	22043085	22043104	use in combination with mm Hist1h2ae

Table 1: Primer sequences for quality control of human and mouse Hi-C libraries.

Discussion

Modular design of Promoter Capture Hi-C

Promoter Capture Hi-C is designed to specifically enrich Hi-C libraries for interactions involving promoters. These interactions comprise only a subset of ligation products present in a Hi-C library.

Capture Hi-C can easily be modified to enrich Hi-C libraries for any genomic region or regions of interest by changing the capture system. Capture regions can be continuous genomic segments[44,45,46,48], enhancers that have been identified in PCHi-C ('Reverse Capture Hi-C'[35]), or DNase I hypersensitive sites[49]. The size of the capture system can be adjusted depending on the experimental scope. For example, Dryden *et al*. target 519 bait fragments in three gene deserts associated with breast cancer[44]. The capture system by Martin *et al*. targets both continuous genomic segments ('Region Capture': 211 genomic regions in total; 2,131 restriction fragments) and selected promoters (3,857 gene promoters)[45].

SureSelect libraries are available in different size ranges: 1 kb to 499 kb (5,190–4,806), 500 kb to 2.9 Mb (5,190–4,816), and 3 Mb to 5.9 Mb (5,190–4,831). As each individual capture biotin-RNA is 120 nucleotides long, these capture systems accommodate a maximum of 4,158, 24,166 and 49,166 individual capture probes, respectively. This corresponds to 2,079, 12,083, and 24,583 targeted restriction fragments, respectively (note that the numbers for restriction fragments are lower bounds based on the assumption that two individual capture probes can be designed for every restriction fragment — in reality due to repetitive sequences this will not be the case for every restriction fragment (see also **Figure 1B**, **C**), resulting in a higher number of targetable restriction fragments for a constant number of available capture probes).

The protocol described here is based on the use of a restriction enzyme with a 6 bp recognition site to uncover long-range interactions. Using a restriction enzyme with a 4 bp recognition site for greater resolution of more proximal interactions is also possible[40,49].

Limitations of PCHi-C

One inherent limitation of all chromosome conformation capture assays is that their resolution is determined by the restriction enzyme used for the library generation. Interactions that occur between DNA elements located on the same restriction fragment are invisible to 'C-type' assays. Further, in PCHi-C, in some cases more than one transcription start site can be located on the same promoter-containing restriction fragment, and PIRs in some cases harbor both active and repressive histone marks, making it difficult to pinpoint which regulatory elements mediate the interactions, and to predict the regulatory output of promoter interactions. Using restriction enzymes with 4 bp recognition sites mitigates this issue but comes at the expense of vastly increased Hi-C library complexity (Hi-C libraries generated with 4 bp recognition site restriction enzymes are at least 100 times more complex than Hi-C libraries generated with 6 bp recognition site restriction enzymes), and the associated costs for next generation sequencing.

Another limitation is that the current PCHi-C protocol requires millions of cells as starting material, precluding the analysis of promoter interactions in rare cell types. A modified version of PCHi-C to enable the interrogation of promoter contacts in cell populations with 10,000 to 100,000 cells (for example cells during early embryonic development or hematopoietic stem cells) would therefore be a valuable addition to the Capture Hi-C toolbox.

Finally, like all methods that rely on formaldehyde fixation, PCHi-C only records interactions that are 'frozen' at the time point of fixation. Thus, to study the kinetics and dynamics of promoter interactions, methods such as super-resolution live cell microscopy are required alongside PCHi-C.

Methods to dissect spatial chromosome organization at high resolution

The vast complexity of chromosomal interaction libraries prohibits the reliable identification of interaction products between two specific restriction fragments with statistical significance. To circumvent this problem, sequence capture has been used to enrich either Hi-C[33,34,40,44] or 3C[50,51] libraries for specific interactions. The major advantage of using Hi-C libraries over 3C libraries for the enrichment step is that Hi-C, unlike 3C, includes an enrichment step for genuine ligation products. As a consequence, the percentage of valid reads in PCHi-C libraries is

Copyright © 2018 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

approximately 10-fold higher than in Capture-C libraries[50], which contained around 5–8% valid reads after HiCUP filtering. Sahlen *et al.* have directly compared Capture-C to HiCap, which like PCHi-C uses Hi-C libraries for capture enrichment, in contrast to Capture-C which uses 3C libraries. Consistent with our findings, they found that Capture-C libraries are mainly composed of un-ligated fragments[40]. In addition, HiCap libraries had a higher complexity than Capture-C libraries[40].

A variant of Capture-C, called next-generation Capture-C[52] (NG Capture-C) uses one oligo per restriction fragment end, as previously established in PCHi-C[33,34], instead of overlapping probes used in the original Capture-C protocol[50]. This increases the percentage of valid reads compared to Capture-C modestly, but NG Capture-C employs two sequential rounds of capture enrichment, and a relatively high number of PCR cycles (20 to 24 cycles in total, compared to 11 cycles typically for PCHi-C), which inevitably results in higher numbers of sequence duplicates and lower library complexity. In trial experiments during the optimization of PCHi-C, we found that the percentage of unique (*i.e.,* not duplicated) read pairs was only around 15% when we used 19 PCR cycles (13 cycles pre-capture + 6 cycles post-capture; data not shown), however optimization to a lower number of PCR cycles, typically yields 75–90% unique read pairs. Thus, reducing the number of PCR cycles substantially increases the amount of informative sequence data.

A recent method combines ChIP with Hi-C to focus on chromosomal interactions mediated by a specific protein of interest (HiChIP[53]). Compared to ChIA-PET[54], which is based on a similar rationale, HiChIP data contains a higher number of informative sequence reads, allowing for higher-confidence interaction calling[53]. It will be very interesting to directly compare the corresponding HiChIP and Capture Hi-C data sets once they become available (for example HiChIP using an antibody against the cohesin unit Smc1a[53] with Capture Hi-C for all Smc1a bound restriction fragments) side by side. One inherent difference between these two approaches is that Capture Hi-C does not rely on chromatin immunoprecipitation, and therefore is capable of interrogating chromosomal interactions irrespective of protein occupancy. This enables comparison of 3D genome organization in the presence or absence of specific factor binding, as has been used to identify PRC1 as a key regulator of mouse ESC spatial genome architecture[7].

PCHi-C and GWAS

Genome-wide association studies (GWAS) have revealed that greater than 95% of disease-associated sequence variants are located in non-coding regions of the genome, often at great distances to protein-coding genes[55]. GWAS variants are often found in close proximity to DNase I hypersensitive sites, which is a hallmark of sequences with potential regulatory activity. PCHi-C and Capture Hi-C have been used extensively to link promoters to GWAS risk loci implicated in breast cancer[44], colorectal cancer[48], and autoimmune disease[35,45,46]. A PCHi-C study on 17 different human hematopoietic cell types found SNPs associated with autoimmune disease were enriched in PIRs in lymphoid cells, whereas sequence variants associated with platelet and red blood cell specific traits were predominantly found in the macrophages and erythroblasts, respectively[35,56]. Thus, tissue-type specific promoter interactomes uncovered by PCHi-C may help to understand the function of non-coding disease-associated sequence variants and identify new potential disease genes for therapeutic intervention.

Characteristics of promoter-interacting regions

Several lines of evidence link promoter interactomes to gene expression control. First, several PCHi-C studies have demonstrated that genomic regions interacting with promoters of (highly) expressed genes are enriched in marks associated with enhancer activity, such as H3K27 acetylation and p300 binding[33,34,37]. We found a positive correlation between gene expression level and the number of interacting enhancers, suggesting that additive effects of enhancers result in increased gene expression levels[34,35]. Second, naturally occurring expression quantitative trait loci (eQTLs) are enriched in PIRs that are connected to the same genes whose expression is affected by the eQTLs[35]. Third, by integrating TRIP[57] and PCHi-C data, Cairns *et al.* found that TRIP reporter genes mapping to PIRs in mouse ESCs show stronger reporter gene expression than reporter genes at integration sites in non-promoter-interacting regions[58], indicating that PIRs possess transcriptional regulatory activity. Together, these findings suggest that promoter interactomes uncovered by PCHi-C in various mouse and human cell types include key regulatory modules for gene expression control.

It is worth noting that enhancers represent only a small fraction (~20%) of all PIRs uncovered by PCHi-C[33,34]. Other PIRs could have structural or topological roles rather than direct transcriptional regulatory functions. However, there is also evidence that PCHi-C may uncover DNA elements with regulatory function that do not harbor classical enhancer marks. In a human lymphoid cell line, the *BRD7* promoter was found to interact with a region devoid of enhancer marks that was shown to possess enhancer activity in reporter gene assays[33]. Regulatory elements with similar characteristics may be more abundant than currently appreciated. For example, a CRISPR-based screen for regulatory DNA elements identified unmarked regulatory elements (UREs) that control gene expression but are devoid of enhancer marks[59].

In other cases, PIRs have been shown to harbor chromatin marks associated with transcriptional repression. PIRs and interacting promoters bound by PRC1 in mouse ESCs were engaged in an extensive spatial network of repressed genes bearing the repressive mark H3K27me3[7]. In human lymphoblastoid cells, a distant element interacting with the *BCL6* promoter repressed transgene reporter gene expression[33], suggesting that it may function to repress *BCL6* transcription in its native context.

PIRs enriched for occupancy of the chromatin insulator protein CTCF in human ESCs and NECs[37] may represent yet another class of PIRs. Collectively, these results suggest that PIRs harbor a collection of gene regulatory activities yet to be functionally characterized.

Disclosures

Authors have nothing to disclose.

Acknowledgements

We thank Valeriya Malysheva for critical reading of the manuscript and expert help with Figure 1. This work was supported by the Medical Research Council, UK (MR/L007150/1) and the UK Biotechnology and Biological Sciences Research Council, UK (BB/J004480/1).

Copyright © 2018 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

References

1. Osborne, C. S. *et al.* Active genes dynamically colocalize to shared sites of ongoing transcription. *Nature Genetics* . **36**, 1065-1071 (2004).
2. Schoenfelder, S. *et al.* Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nature Genetics* . **42**, 53-61 (2010).
3. de Wit, E. *et al.* The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature.* **501**, 227-231 (2013).
4. Bantignies, F. *et al.* Polycomb-dependent regulatory contacts between distant *Hox* loci in *Drosophila. Cell.* **144**, 214-226 (2011).
5. Engreitz, J. M. *et al.* The *Xist* lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science.* **341**, 1237973 (2013).
6. Denholtz, M. *et al.* Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell.* **13**, 602-616 (2013).
7. Schoenfelder, S. *et al.* Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nature Genetics* . **47**, 1179-1186 (2015).
8. Kundu, S. *et al.* Polycomb Repressive Complex 1 generates discrete compacted domains that change during differentiation. *Molecular Cell* . **65**, 432-446 e435 (2017).
9. Skok, J.A., Gisler, R., Novatchkova, M., Farmer, D., de Laat, W., Busslinger, M. Reversible contraction by looping of the *Tcra* and *Tcrb* loci in rearranging thymocytes. *Nature Immunology* . **8**, 378-387 (2007).
10. Zhang, Y. *et al.* Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell.* **148**, 908-921 (2012).
11. Aymard, F. et al. Genome-wide mapping of long-range contacts unveils clustering of DNA double-strand breaks at damaged active genes. *Nature Structural & Molecular Biology* . **24**, 353-361 (2017).
12. Ryba, T. *et al.* Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Research.* **20**, 761-770 (2010).
13. Pope, B. D. *et al.* Topologically associating domains are stable units of replication-timing regulation. *Nature.* **515**, 402-405 (2014).
14. Chandra, T. *et al.* Global reorganization of the nuclear landscape in senescent cells. *Cell Reports* . **10**, 471-483 (2015).
15. Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F., Fraser, P. Long-range chromatin regulatory interactions *in vivo. Nature Genetics* . **32**, 623-626 (2002).
16. Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., de Laat, W. Looping and interaction between hypersensitive sites in the active *beta-globin* locus. *Molecular Cell* . **10**, 1453-1465 (2002).
17. Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., Shiroishi, T. Chromosomal dynamics at the *Shh* locus: limb bud-specific differential regulation of competence and active transcription. *Developmental Cell.* **16**, 47-57 (2009).
18. Zuniga, A. *et al.* Mouse limb deformity mutations disrupt a global control region within the large regulatory landscape required for *Gremlin* expression. *Genes & Development* . **18**, 1553-1564 (2004).
19. Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M., Shiroishi, T. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific *Shh* expression and truncation of the mouse limb. *Development.* **132**, 797-803 (2005).
20. D'Haene, B. *et al.* Disease-causing 7.4 kb *cis*-regulatory deletion disrupting conserved non-coding sequences and their interaction with the *FOXL2* promotor: implications for mutation screening. *PLoS Genet.* **5**, e1000522 (2009).
21. Sur, I. K. *et al.* Mice lacking a *Myc* enhancer that includes human SNP *rs6983267* are resistant to intestinal tumors. *Science.* **338**, 1360-1363 (2012).
22. Herranz, D. *et al.* A NOTCH1-driven *MYC* enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nature Medicine* . **20**, 1130-1137 (2014).
23. Deng, W. *et al.* Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell.* **149**, 1233-1244 (2012).
24. Groschel, S. *et al.* A single oncogenic enhancer rearrangement causes concomitant *EVI1* and *GATA2* deregulation in leukemia. *Cell.* **157**, 369-381 (2014).
25. Lupianez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell.* **161**, 1012-1025 (2015).
26. Franke, M. *et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* . **538**, 265-269 (2016).
27. Dekker, J., Rippe, K., Dekker, M., Kleckner, N. Capturing chromosome conformation. *Science.* **295**, 1306-1311 (2002).
28. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genetics* . **38**, 1348-1354 (2006).
29. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genetics* . **38**, 1341-1347 (2006).
30. Dostie, J. *et al.* Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Research* . **16**, 1299-1309 (2006).
31. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* **326**, 289-293 (2009).
32. Belton, J.M., McCord, R.P., Gibcus, J.H., Naumova, N., Zhan, Y., Dekker, J. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods.* **58**, 268-276 (2012).
33. Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature Genetics.* **47**, 598-606 (2015).
34. Schoenfelder, S. *et al.* The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* **25**, 582-597 (2015).
35. Javierre, B. M. *et al.* Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell.* **167**, 1369-1384 e1319 (2016).
36. Wilson, N. K. *et al.* Integrated genome-scale analysis of the transcriptional regulatory landscape in a blood stem/progenitor cell model. *Blood.* **127**, e12-23 (2016).

Copyright © 2018 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

37. Freire-Pritchett, P. *et al.* Global reorganisation of *cis*-regulatory units upon lineage commitment of human embryonic stem cells. *Elife.* **6** (2017).
38. Rubin, A. J. *et al.* Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nature Genetics .* **49**, 1522-1528 (2017).
39. Siersbaek, R. *et al.* Dynamic rewiring of promoter-anchored chromatin loops during adipocyte differentiation. *Molecular Cell .* **66**, 420-435 e425 (2017).
40. Sahlen, P. *et al.* Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. *Genome Biology.* **16**, 156 (2015).
41. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature.* **502**, 59-64 (2013).
42. Nagano, T., Varnai, C., Schoenfelder, S., Javierre, B.M., Wingett, S.W., Fraser, P. Comparison of Hi-C results using in-solution versus in-nucleus ligation. *Genome Biology.* **16**, 175 (2015).
43. Wingett, S. *et al.* HiCUP: pipeline for mapping and processing Hi-C data. *F1000 Res.* **4**, 1310 (2015).
44. Dryden, N. H. *et al.* Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. *Genome Research .* **24**, 1854-1868 (2014).
45. Martin, P. *et al.* Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. *Nature Communications.* **6**, 10069 (2015).
46. McGovern, A. *et al.* Capture Hi-C identifies a novel causal gene, *IL20RA*, in the pan-autoimmune genetic susceptibility region *6q23. Genome Biol.*ogy **17**, 212 (2016).
47. Hodge, D. *et al.* A global role for EKLF in definitive and primitive erythropoiesis. *Blood.* **107**, 3359-3370 (2006).
48. Jager, R. *et al.* Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nature Communications.* **6**, 6178 (2015).
49. Joshi, O. *et al.* Dynamic reorganization of extremely long-range promoter-promoter Interactions between two states of pluripotency. *Cell Stem Cell .* **17**, 748-757 (2015).
50. Hughes, J. R. *et al.* Analysis of hundreds of *cis*-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nature Genetics .* **46**, 205-212 (2014).
51. Kolovos, P. *et al.* Targeted Chromatin Capture (T2C): A novel high-resolution high-throughput method to detect genomic interactions and regulatory elements. *Epigenetics Chromatin.* **7**, 10 (2014).
52. Davies, J. O. *et al.* Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nature Methods .* **13**, 74-80 (2016).
53. Mumbach, M. R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nature Methods .* **13**, 919-922 (2016).
54. Fullwood, M. J. *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature.* **462**, 58-64 (2009).
55. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science.* **337**, 1190-1195 (2012).
56. Petersen, R. *et al.* Platelet function is modified by common sequence variation in megakaryocyte super enhancers. *Nat. Commun.* **8**, 16058 (2017).
57. Akhtar, W. *et al.* Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell.* **154**, 914-927 (2013).
58. Cairns, J. *et al.* CHiCAGO: Robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biology.* **17**, 127 (2016).
59. Rajagopal, N. *et al.* High-throughput mapping of regulatory DNA. *Nature Biotechnology.* **34**, 167-174 (2016).

Copyright © 2018 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License