nature protocols

PROTOCOL

https://doi.org/10.1038/s41596-021-00633-y

Check for updates

SPRITE: a genome-wide method for mapping higher-order 3D interactions in the nucleus using combinatorial split-and-pool barcoding

Sofia A. Quinodoz[1,3], Prashant Bhat[1,2], Peter Chovanec[1], Joanna W. Jachowicz[1], Noah Ollikainen[1], Elizabeth Detmar[1], Elizabeth Soehalim[1] and Mitchell Guttman[1]✉

A fundamental question in gene regulation is how cell-type-specific gene expression is influenced by the packaging of DNA within the nucleus of each cell. We recently developed Split-Pool Recognition of Interactions by Tag Extension (SPRITE), which enables mapping of higher-order interactions within the nucleus. SPRITE works by cross-linking interacting DNA, RNA and protein molecules and then mapping DNA–DNA spatial arrangements through an iterative split-and-pool barcoding method. All DNA molecules within a cross-linked complex are barcoded by repeatedly splitting complexes across a 96-well plate, ligating molecules with a unique tag sequence, and pooling all complexes into a single well before repeating the tagging. Because all molecules in a cross-linked complex are covalently attached, they will sort together throughout each round of split-and-pool and will obtain the same series of SPRITE tags, which we refer to as a barcode. The DNA fragments and their associated barcodes are sequenced, and all reads sharing identical barcodes are matched to reconstruct interactions. SPRITE accurately maps pairwise DNA interactions within the nucleus and measures higher-order spatial contacts occurring among up to thousands of simultaneously interacting molecules. Here, we provide a detailed protocol for the experimental steps of SPRITE, including a video (https://youtu.be/6SdWkBxQGlg). Furthermore, we provide an automated computational pipeline available on GitHub that allows experimenters to seamlessly generate SPRITE interaction matrices starting with raw fastq files. The protocol takes ~5 d from cell cross-linking to high-throughput sequencing for the experimental steps and 1 d for data processing.

Introduction

Although all cells in an organism share the same genomic DNA sequence packaged within the nucleus of each cell, different genes are expressed depending on the cell type. It has become increasingly clear that the three-dimensional (3D) organization of the genome plays a crucial gene regulatory role in controlling gene expression in various biological contexts, including during embryonic development and cellular differentiation[1–4]. Most methods for genome-wide mapping of 3D genome structure rely on proximity ligation[5–7] (e.g., Hi-C), which works by cross-linking interacting loci through DNA–protein fixation, ligating the ends of DNA regions that are in close spatial proximity in the nucleus, performing affinity purification of ligated fragments (e.g., biotin selection), and sequencing to map pairwise interactions. These techniques have revealed that the genome is organized into structures including chromosome territories, compartments that correspond to transcriptional activity[2,5], topologically associating domains (TADs) corresponding to locally interacting regions of DNA[8], and chromatin loops that connect enhancers with promoters[9].

Although proximity ligation methods have led to important progress in understanding the multiple layers of 3D genome organization, these approaches are limited to identifying DNA interactions that are close enough to directly ligate[10]. Accordingly, these techniques often fail to detect known interactions occurring around larger-scale nuclear structures identified by microscopy, such as nuclear bodies. Because nuclear bodies can range in size from 0.5 to 2 μm, such DNA regions may be too far apart to directly ligate[11]. As a result, these larger-scale interactions may be missed by proximity ligation methods. Additionally, proximity ligation-based methods are primarily limited to measuring pairwise interactions between two genomic loci, and therefore cannot measure how many DNA sites simultaneously organize

[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. [2]David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [3]Present address: Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ, USA. ✉e-mail: mguttman@caltech.edu

within structures in the nucleus. Accordingly, these methods are limited in their ability to generate comprehensive global models of how *multiple* genomic loci are arranged in 3D space. These limitations make it difficult to answer long-standing questions such as which DNA regions simultaneously co-localize around the same nuclear body, how DNA is organized at various genomic scales (i.e., compartments, TADs, loops) and whether some genes exhibit preferential association with distinct nuclear bodies.

Another widely used approach for mapping the 3D genome is microscopy. In situ imaging of specific genomic loci by fluorescence in situ hybridization can reveal the spatial positioning of various genomic loci in 3D space. Recently, multiplexed adaptations of fluorescence in situ hybridization involving iterative hybridization of probes to thousands of DNA molecules have enabled multiplexed visualization in 3D space[12,13]. Although incredibly powerful, microscopy approaches are limited by using probes against a known set of genomic regions in contrast to sequencing-based approaches, which do not require a priori knowledge of interacting genomic regions.

To address these limitations, we developed SPRITE (Split-Pool Recognition of Interactions by Tag Extension), a proximity-ligation independent method that enables genome-wide detection of multiple DNA molecules that co-occur within a range of differently sized cross-linked complexes in the nucleus[14]. We refer to these proximal DNA–DNA contacts as 'interactions' throughout the text. Using SPRITE, we identify genome structures previously discovered by Hi-C, including chromosome territories, compartments, TADs and loop structures. Furthermore, because SPRITE is not constrained to mapping pairwise interactions, we were able to identify many DNA interactions that occur within higher-order (beyond pairwise) structures in the nucleus. In addition to SPRITE, proximity-ligation independent methods (such as genome architecture mapping and ChIA-DROP) can map multiway contacts[15,16], and recently C-based methods (such as C-walks, multi-contact 3C and 4C) have been modified to detect multiway contacts using long-read sequencing[17–19]. Finally, because SPRITE is not constrained to mapping interactions that are close enough to directly ligate, it can map the landscape of long-range intra- as well as inter-chromosomal interactions that organize around larger nuclear structures such as nuclear bodies, including nucleoli and nuclear speckles[14]. In addition to SPRITE, proximity labeling methods (such as TSA-seq and DamID) have also enabled mapping of preferential association of DNA regions relative to various nuclear bodies[20–22].

Overview of SPRITE

SPRITE is a high-throughput sequencing-based method used for mapping genome-wide higher-order interactions between DNA molecules. Briefly, SPRITE involves the following stages (Fig. 1).

Day 1

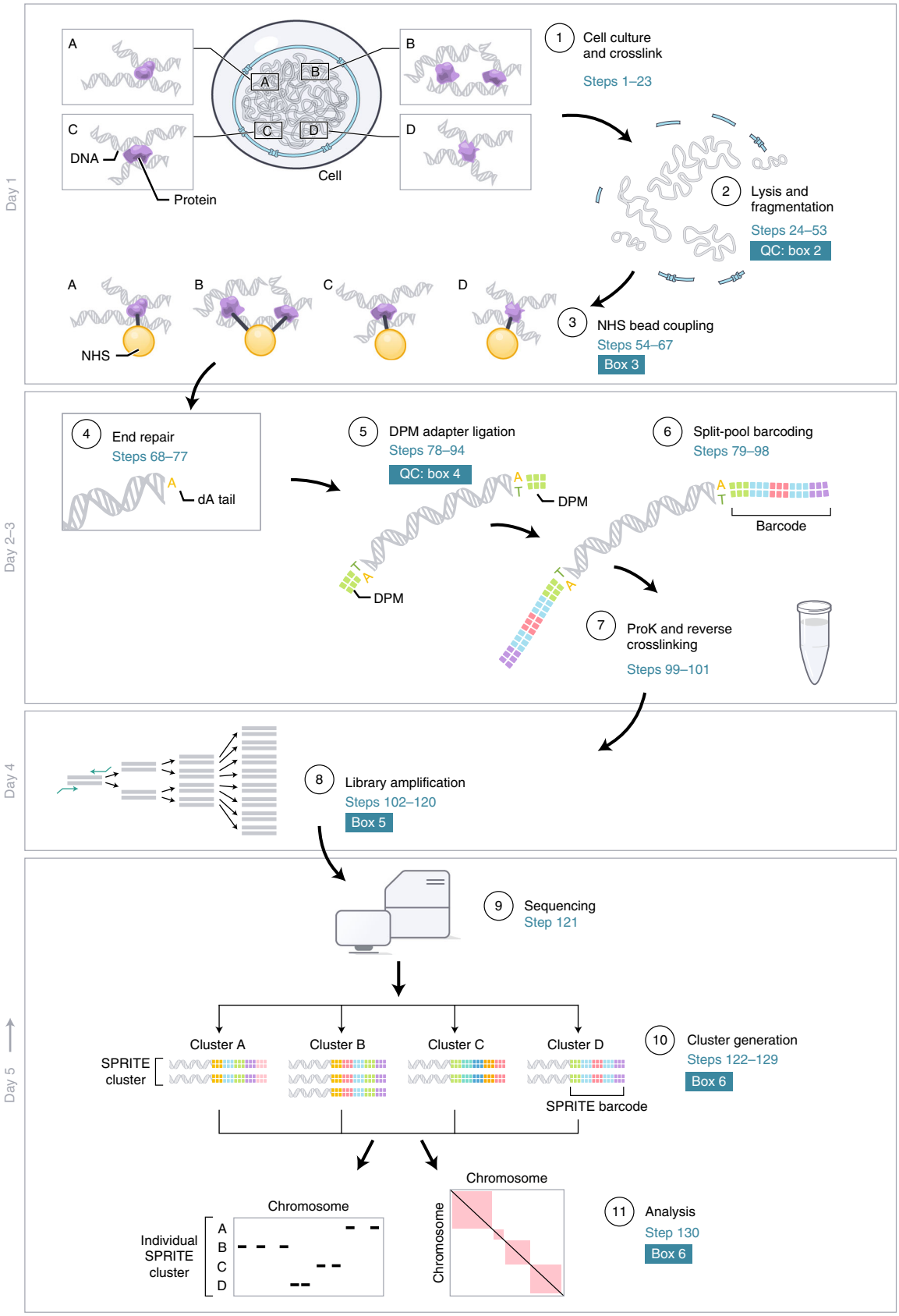
1 Cross-linking. Cells are cross-linked using a combination of formaldehyde and a protein–protein cross-linker called disuccinimidyl glutarate (DSG).
2 Lysis. Cells are lysed and the DNA is fragmented using sonication and DNase digestion.
3 NHS coupling. Cross-linked lysate is covalently coupled to NHS-ester beads overnight.

Days 2–3

4 End repair and dA-tailing. DNA is end-repaired and dA-tailed through treatment with end repair and dA-tailing enzymes.
5 Split-pool barcoding: DNA phosphate modified (DPM) ligation. DPM adaptor, containing a dT overhang, is ligated to the complementary ends of dA-tailed DNA. This serves as the first round of split-pool barcoding prior to performing four additional barcoding rounds. At this step, complexes coupled to beads are split into a plate of 96 unique tag adaptors where T4 DNA ligase adds each of these well-specific tagged DPM adaptors onto the genomic DNA in each complex. The sample is then pooled into a single reservoir and split again into a 96-well plate for another round of tagging.
6 Split-pool barcoding: Odd, Even, Odd and Terminal adaptor ligations. Including DPM ligation, five total rounds of split-and-pool barcoding are performed to achieve >8 billion unique combinations of sequences. The entire split-and-pool barcoding procedure can either be completed in a single day or split across 2 days.
7 Reverse cross-linking. Proteinase K (ProK) is added to digest proteins and samples are incubated overnight to reverse cross-links.

Day 4

8 Library preparation. PCR is performed to add sequencing adaptors and amplify the libraries prior to sequencing.


Day 1
A
B
C
D
DNA
Protein
Cell
1 Cell culture and crosslink
Steps 1–23
2 Lysis and fragmentation
Steps 24–53
QC: box 2
NHS
3 NHS bead coupling
Steps 54–67
Box 3
Day 2–3
4 End repair
Steps 68–77
dA tail
5 DPM adapter ligation
Steps 78–94
QC: box 4
DPM
6 Split-pool barcoding
Steps 79–98
Barcode
7 ProK and reverse crosslinking
Steps 99–101
Day 4
8 Library amplification
Steps 102–120
Box 5
Day 5
9 Sequencing
Step 121
Cluster A
Cluster B
Cluster C
Cluster D
SPRITE cluster
SPRITE barcode
10 Cluster generation
Steps 122–129
Box 6
Chromosome
Individual SPRITE cluster
11 Analysis
Step 130
Box 6

◀ **Fig. 1 | Overview of SPRITE procedure.** Day 1: (1) cells are dual cross-linked with DSG and formaldehyde (Steps 1–23); (2) cells are lysed and chromatin is fragmented using sonication and DNase digestion (Steps 24–53); (3) cross-linked complexes in lysate are coupled to NHS beads overnight (Steps 54–67). Days 2–3: (4) DNA is blunt-ended, phosphorylated and dA-tailed (Steps 68–77) prior to (5) ligation with DPM adaptor (Steps 78–94); (6) four additional rounds of split-and-pool ligations are performed with the Odd, Even, Odd and Terminal tags, which we refer to as a barcode (Steps 79–98); (7) after split-and-pool, samples are split into several aliquots and DNA is reverse cross-linked overnight by addition of ProK enzyme and heat (Steps 99–101). Day 4: (8) final SPRITE libraries are amplified (Steps 102–120). Day 5 onward: (9) DNA is sequenced (Step 121), and (10) all molecules sharing the same barcodes are matched to generate SPRITE clusters (Steps 122–129); (11) DNA interactions occurring in SPRITE clusters can be analyzed as pairwise interactions, visualized using intra- and interchromosomal heatmaps, or as multiway interactions, visualized using individual clusters (Step 130).

Day 5+: sequencing and analysis

9 Sequencing. High-throughput sequencing is performed to sequence SPRITE libraries.

10 Cluster generation. DNA reads sharing the same barcode are matched to generate SPRITE 'clusters' corresponding to molecules from the same cross-linked complex and aligned to the genome.

11 Analysis. Using the SPRITE clusters, pairwise and higher-order maps of DNA interactions in the nucleus are constructed at various resolutions.

Advantages

Higher-order spatial interactions

One of the advantages of SPRITE is that it can measure interactions between multiple molecules interacting in the same 3D space. This enables construction of higher-order spatial interactions between DNA molecules such as those including multiple genes involved in a shared biological process (e.g., histone gene clusters). This is in contrast to proximity ligation methods, such as Hi-C, which primarily capture pairwise (two-way) interactions.

Global spatial maps

Because SPRITE utilizes a proximity-ligation independent method to detect interactions, it is not limited to identifying only those molecules that are close enough to directly ligate. Instead, larger SPRITE clusters can detect long-range interactions including those occurring between regions on separate chromosomes. For example, SPRITE clusters containing >1,000 DNA reads in a cross-linked complex identify long-range interactions occurring between multiple DNA loci that are simultaneously associating around the nucleolus (e.g., mouse chromosomes 10, 12, 18 and 19), while SPRITE clusters containing 10–100 reads can identify interchromosomal interactions between genomic sites associating with nuclear speckles. The ability to detect interactions across a range of distances enables measurement of cross-linked complexes to reconstruct various close-range (e.g., loops) and long-range interactions (e.g., around nuclear bodies) in the nucleus.

Applications

Mapping organization around nuclear landmarks

Recently, SPRITE coupled with immunoprecipitation was applied to bone-marrow-derived dendritic cells to map multiway enhancer–promoter interactions after stimulation with lipopolysaccharide[23]. This presents a unique opportunity to map higher-order DNA organization around various nuclear landmarks (e.g., proteins, chromatin modifications) for which there are validated chromatin immunoprecipitation antibodies.

SPRITE in different cell lines

In addition to many commonly used mammalian cell types (fibroblasts, embryonic stem cells, immune cells and cardiomyocytes)[14,23,24], we and others have successfully applied SPRITE to nonmammalian cell types with minimal modifications, including *Arabidopsis thaliana* and *Dalotia coriaria* in unpublished work. In these cases, we homogenized cellular material prior to cross-linking.

Genome assembly

Constructing an accurate reference genome for a model organism is critical for sequence analysis. Recently, Hi-C has been used to assemble genomes[25,26] owing to the fact that most interactions between DNA molecules occur within, as opposed to between, individual chromosomes. To use

SPRITE for genome assembly, we recommend generating contact maps with SPRITE clusters of size 2–100 or 2–1,000 reads and excluding larger clusters (1,000+ reads/cluster), which primarily contain interchromosomal interactions.

Mapping spatial interactions of other biomolecules in the nucleus

SPRITE represents a robust framework to map spatial interactions of other biomolecules (e.g., RNA and protein) and their co-occurrence in 3D space. For example, SPRITE has been successfully used to map RNA–RNA and RNA–DNA contacts[27], and we anticipate that it can be extended to map protein localization using a pool of barcoded antibodies[28,29] to generate combinatorial and spatial maps of DNA, RNA and/or protein. Finally, SPRITE can be used to generate single-cell structure maps by adapting the initial lysis steps[30]. Such applications of SPRITE will extend our ability to research 3D spatial mapping of these biomolecules that have long remained a challenge.

Experimental design

Cell culture

Cells can be cultured using standard guidelines for the cell type best suited for the experimenter's needs. We have performed SPRITE on mouse F1-21 embryonic stem cells and human GM12878 lymphoblast cells. Details about cell culturing conditions can be found in ref. [14].

Adaptor and barcode design

The SPRITE adaptors and tag ligation scheme that is central to the SPRITE process are described in Box 1.

Cell number

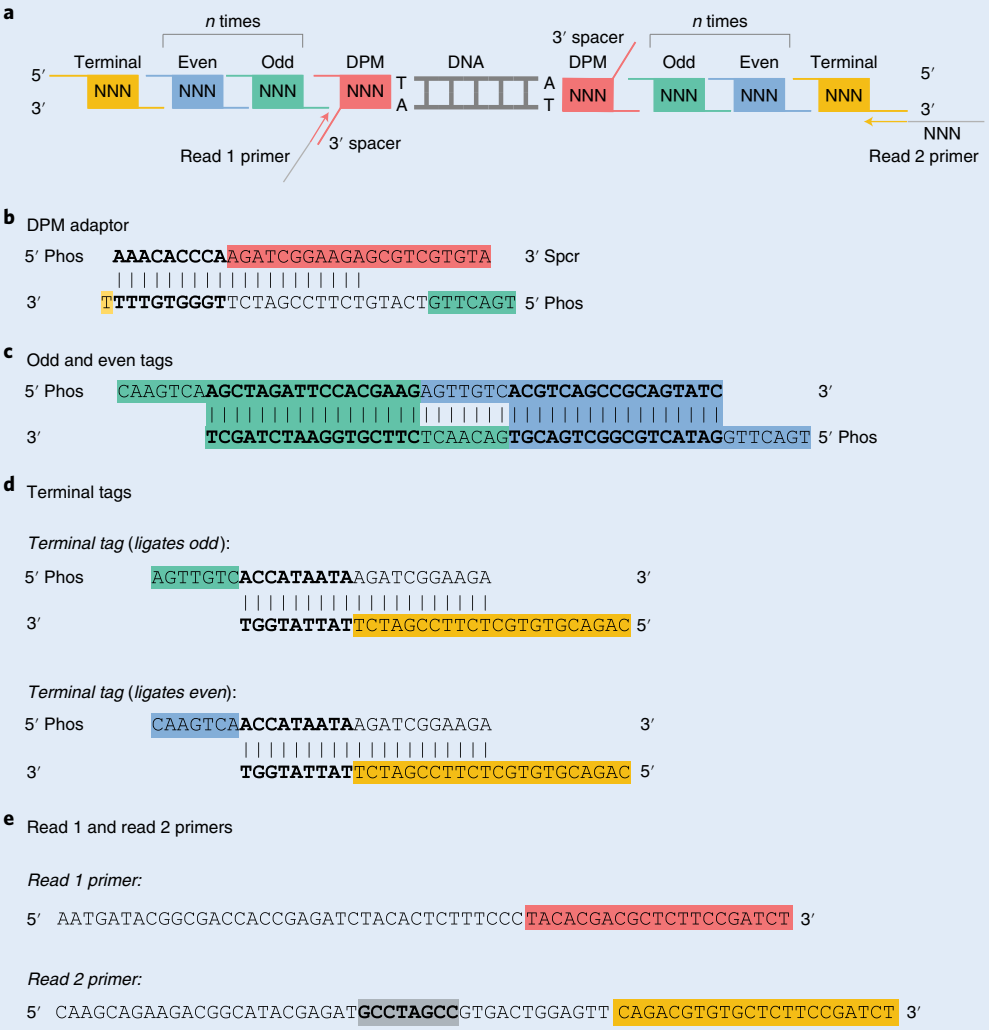
While we typically cross-link 5–10 million cells for a SPRITE experiment, the final amount of material inputted into the SPRITE split-and-pool tagging steps corresponds to DNA recovered from ~3,000 cells. However, we typically work with far more cells during the cross-linking steps because (i) loss can occur during the cross-linking and lysis procedure and (ii) DNase optimization steps require enough DNA to visualize in a gel and to test multiple concentrations of DNase enzyme. In applications where cell number is limiting, cell number can be reduced. We recommend ensuring that at least >100 ng of genomic DNA is available to test at least four DNase concentrations for DNase digestion. In our experience, 1–2 million cells have sufficient genomic DNA to perform several (>100) tests for DNase digestion. The minimum number of cells we have successfully generated SPRITE heatmaps and libraries from is ~200,000 cells.

Cross-linking and cell lysis

Cells are dual cross-linked with 2 mM DSG and 3% formaldehyde, which favors fixing large chromatin complexes and long-range interactions. We have tested reducing the formaldehyde concentration to 1% while keeping the DSG concentration to 2 mM with reduced sonication time (~30–45 s, 4–5 W; Branson Sonicator) and have obtained successful SPRITE libraries with known interactions such as TADs, compartments and chromosome territories. The fragmentation conditions recommended in this protocol have been optimized for cells cross-linked with 3% formaldehyde and 2 mM DSG. Cross-linking with 1% or 3% formaldehyde without DSG may work, but applying the same extent of sonication (1 min, 4–5 W; Branson Sonicator) will result in SPRITE libraries almost completely devoid of interactions. We recommend optimizing fragmentation if cross-linking conditions are varied to ensure interactions are maintained during sonication. After cross-linking, cells are lysed with a protocol modified from the Amit Laboratory HT-ChIP protocol[31] prior to chromatin fragmentation using sonication.

To prevent clumping of cells in suspension in these strong cross-linking conditions, we recommend resuspending the cross-linked pellet uniformly in a smaller volume (1 mL) of DSG cross-linking solution using a P-1000 micropipette before adding the full volume of cross-linking solution. If the full volume of DSG cross-linking solution is added without first resuspending the pellet, it will be challenging to completely break up the pellet, resulting in cell clumps being cross-linked together. We recommend checking nuclei integrity under a microscope before performing the full SPRITE protocol on a new cell type. Alternatively, cross-linking of adherent cells can be performed directly on the plate without trypsinization of cells and has resulted in successful SPRITE experiments (see Supplementary Methods). In this protocol, we perform trypsinization on mouse embryonic stem

Box 1 | Description of SPRITE adaptors and primers



(a) Central adaptor and tag scheme

The above figure illustrates the scheme that is central to the SPRITE process. SPRITE uses a split-and-pool strategy to uniquely barcode all molecules within a cross-linked complex. The final product contains a series of unique tags (DPM, Odd, Even and Terminal tags) ligated to each molecule, which we refer to as a barcode.

SPRITE tags are ligated using a series of oligos with compatible sticky ends. First, the DPM adaptor is ligated to dA-tailed DNA. The DPM tag is then ligated to an 'Odd' tag. The 'Odd' and 'Even' tags were designed so that they can be ligated to each other over multiple ligation rounds, such that, after Odd is ligated, Even ligates the Odd tag, and then Odd can ligate the Even tag. This can be repeated such that the same two plates of 96 tags can be used over multiple rounds of split-pool tagging without self-ligation of the adaptors to each other. Finally, a set of barcoded Terminal tags (Terminal Odd and Terminal Even ligate to Odd and Even tags, respectively) are ligated at the end to attach an Illumina sequence for final library amplification.

As an example, in our previous study (ref. [14]), we performed five rounds total of split-and-pool ligation in the following order: DPM, Odd, Even, Odd and Terminal tag, but SPRITE can be adapted to perform an arbitrary number of rounds of split-pool barcoding. Additionally, to reduce the cost of adaptors purchased to perform SPRITE, a single DPM and single terminal tag can be used. In exchange for omitting these two split-and-pool rounds of DPM and terminal ligations, two additional rounds of Odd and Even split-pool tagging across 96 wells must be performed instead to generate enough combinatorial complexity. Specifically, the experimenter would perform DPM ligation in a single tube before starting split-pool barcoding and terminal ligation in a single tube following split-and-pool barcoding. Increasing the sequencing length on read 2 is also required to sequence through the additional barcodes.

(b) DPM adaptor

The DPM adaptor is the first tag ligated to dA-tailed genomic DNA. Each DPM tag contains: (i) a 5′ phosphate group on both ends for ligation to dA-tailed genomic DNA and to the Odd tags, (ii) a single dT overhang for ligation to dA-tailed genomic DNA (yellow), (iii) a nine-nucleotide sequence unique to each of the 96 DPM tags (bold region), and (iv) a seven-nucleotide sticky-end overhang that ligates to the Odd set of adaptors (green region). Each DPM tag also contains a sequence that is complementary to the universal read 1 Illumina primer, which is used for library amplification (red region).

Box 1 | Description of SPRITE adaptors and primers (Continued)

Because the DPM tag will ligate to both ends of the double-stranded DNA molecule, DPM was designed such that the barcode sequence will only be read from one sequencing read (read 2), rather than both reads (reads 1 and 2). To achieve this, we included a 3′ spacer on the top strand of the DPM adaptor, which prevents the top strand of the Odd tag from ligating to genomic DNA. This modification is also critical for successful amplification of the barcoded DNA by preventing hairpin formation of the single stranded DNA during the initial PCR denaturation, because otherwise both sides of the tagged DNA molecule would have complementary barcode sequences.

(c) Odd and Even tags

After DPM ligation, two sets of 96 'Odd' and 'Even' tags are ligated. The tags are named as such because Odd tags can be ligated in the first, third, fifth, etc., rounds of the SPRITE process and the Even tags can be ligated in the second, fourth, sixth, etc., rounds of SPRITE. The Odd tags contain a seven-nucleotide sticky end (5′ CAAGTCA 3′) that anneals to the Even tags (5′ TGACTTG 3′), and the Even tags have a distinct seven-nucleotide sticky end (5′ AGTTGTC 3′) that anneals to the Odd tags (5′ GACAACT 3′).

The dsDNA molecule is an example of an Odd tag (green) and an Even tag (blue) ligated together. The following points are important to note: (i) the 5′ overhang on the top strand of Odd ligates either to the DPM adaptor (green sequence on DPM bottom strand) or the 5′ overhang on the bottom strand of the Even tag; (ii) the bolded regions on each tag are unique 17 nucleotide sequences for each of the 96 tags (192 total, accounting to Odd tags and Even tags); and (iii) both tags have 5′ phosphate groups to allow for sequential tag ligation. The 5′ overhang on the bottom strand of each of the Odd and Even tags can be ligated to a terminal tag (designed with a complementary overhang for Odd or Even) to attach the Illumina sequence for library amplification.

(d) Terminal tag

The Terminal tags contain a seven-nucleotide sticky end that ligates to the Odd tags (green region), though a Terminal tag can also be designed to ligate to an Even tag (blue region). The Terminal tag only contains a modified 5′ phosphate on the top strand. The bottom strand contains a priming region (yellow) that contains part of the Illumina read 2 sequence, which allows for priming and incorporation of the full-length barcoded read 2 Illumina adaptor. Each terminal tag contains a unique sequence of 9–12 nucleotides (bold). Illumina sequencers produce poor-quality scores when they encounter sequences of low base diversity, such as the common sticky ends of our SPRITE tags in read 2 (e.g., Odd, Even, Terminal). Therefore, to increase base diversity, we add a mixture of zero, one, two or three additional bases (9, 10, 11, 12 nt total) to the Terminal adaptor unique sequences to offset the position of the common sticky end sequence during read 2 (see examples in Supplementary Table 1). This can prevent low-diversity reads during sequencing and makes the SPRITE sequencing compatible on a variety of Illumina Sequencers with one-color, two-color and four-color chemistry.

Final library amplification primers

(e) Read 1 and read 2 primers

DNA libraries are amplified using common primers that incorporate the full Illumina sequencing adaptors. The read 1 primer amplifies the top strand of the DPM tag on DNA (red region), and adds the Illumina read 1 sequence to each molecule. The read 2 primer amplifies the terminal tag on DNA (yellow region) and adds the Illumina read 2 sequence to the molecule.

The DPM adaptor is designed with a 3′ spacer to aid in final library amplification. If the 3′ spacer is absent, each strand will form a hairpin loop during the initial denaturation due to reverse complementarity with the barcode sequences on the other side of the target DNA molecule. Instead, the 3′ spacer allows the barcodes to only ligate to the 5′ end of each single-stranded DNA sequence, and not the 3′ end, preventing these hairpins from forming.

The read 2 (2P_barcoded) primer (shown above is one of 96 available barcoded read 2 primers) contains an eight-nucleotide barcode (bold, gray region) within the primer. This barcode is read from the Illumina sequencer during the indexing priming step and effectively serves as an additional round of tag addition during SPRITE. Dilution of the sample into multiple wells is performed at the final step of SPRITE prior to ProK elution from NHS beads. Each dilution of the sample prior to ProK elution isolates a subset of the tagged complexes into different wells. Each aliquoted sample is amplified with a different 2P_barcoded primer.

Both the read 1 and read 2 primers are ~60 nucleotides each, yet the sequences they anneal to initially are ~20 nucleotides. For this reason, we set two different annealing temperatures during the final library PCR. The first annealing temperature (68 °C) is for the first four cycles until enough copies are made with fully extended primer regions. After these four cycles, the annealing temperature is raised for a remaining five cycles (70 °C).

Sequencing recommendations

Read 1 begins with the DPM tag and is followed by the associated genomic DNA sequence. Therefore, at least 50 bp from read 1 must be sequenced to allow identification of the 10 bp DPM tag sequence and at least an additional 40 bp for sequencing the genomic DNA sequence. Longer reads will improve alignment to genomic DNA, especially allele-specific alignments.

Read 2 sequences the tags ligated to genomic DNA. Therefore, read 2 must be long enough to sequence through all ligated tags. For example, if four rounds of barcoding were performed after DPM ligation, we recommend sequencing at least 90–95 bp to completely sequence through each tag.

Length of SPRITE barcode post-DPM = $\text{Odd}_{\text{len}} + \text{Even}_{\text{len}} + \text{Odd}_{\text{len}} + \text{Terminal}_{\text{len}}$

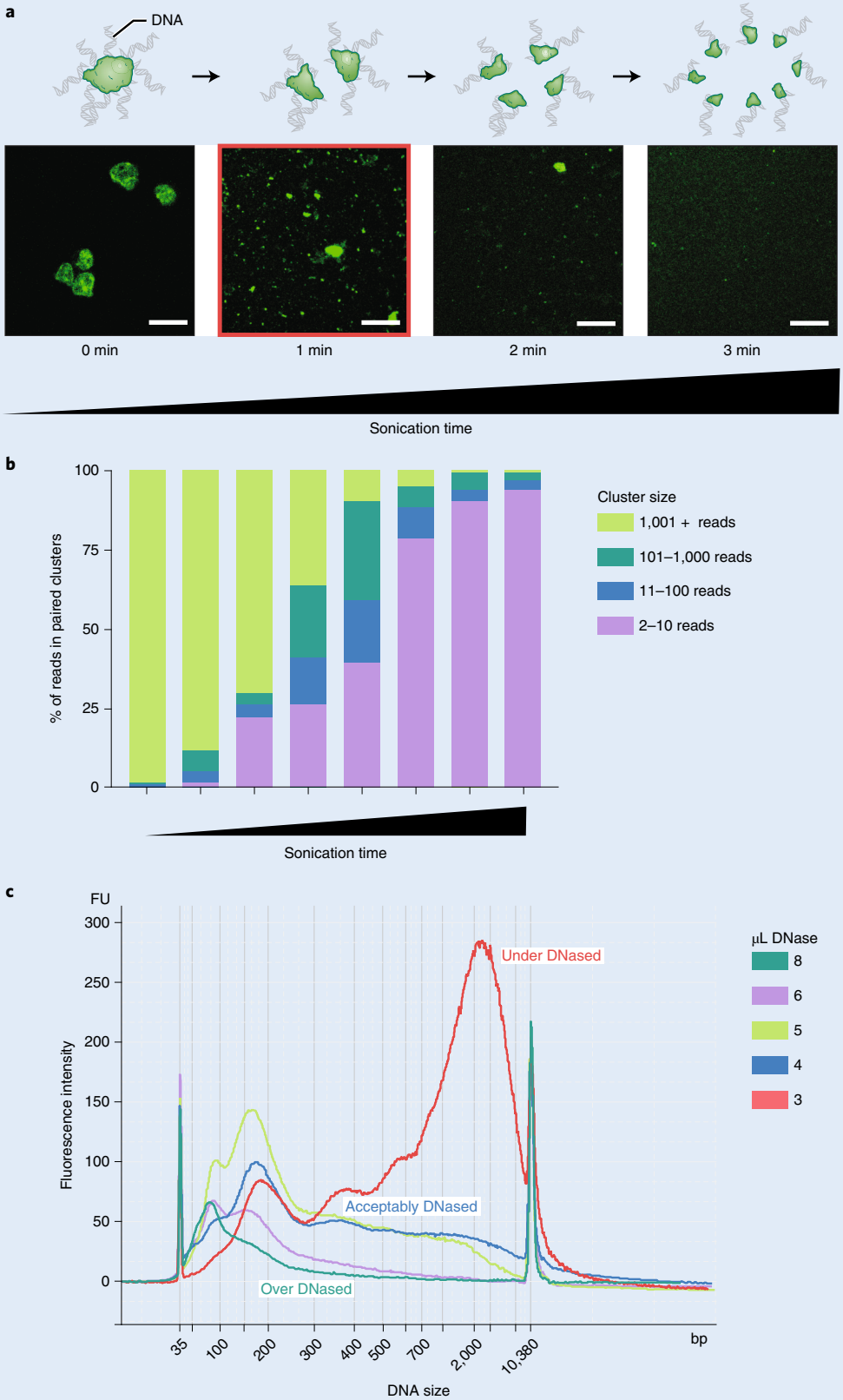
Length of SPRITE barcode post-DPM = 24 bp + 24 bp + 24 bp + (10 to 15 bp) = 82 to 87 bp

(mES) cells prior to cross-linking in order to break up large mES cell colonies into a single-cell suspension.

Chromatin fragmentation

SPRITE fragmentation is performed using a light sonication followed by DNase digestion to obtain DNA sizes suitable for DNA amplification. Optimization of fragmentation conditions (extent of sonication, amount/timing of DNase treatment) is a critical step in establishing the protocol for the first time. The length of sonication might vary from 30 s to several minutes depending on the sonicator used. We have found that sonication time is critical to break up nuclei into smaller structures containing many cross-linked DNA molecules. Oversonication results in libraries devoid of

Box 2 | QC 1: using microscopy to visualize sizes of cross-linked complexes after sonication and fragmentation ● Timing 4 h



Box 2 | QC 1: using microscopy to visualize sizes of cross-linked complexes after sonication and fragmentation ● Timing 4 h (Continued)

Optimizing fragmentation of DNA is an important step in the SPRITE protocol. The goal is to break down intact nuclei into smaller cross-linked complexes of interacting DNA molecules (**a**, scale bar 10 μm). Care should be employed to avoid overfragmentation or underfragmentation as underfragmentation using sonication will result in large SPRITE clusters with almost no small cross-linked complexes, and overfragmentation will result in almost exclusively small clusters and unpaired interactions (**b**). One way to assess the distribution of cross-linked complexes is to directly visualize them by microscopy. Successful SPRITE libraries described here have been obtained after 1 min of sonication using 4–5 W of power (**a**, red box) followed by DNAse digestion to fragment the DNA down to a sequenceable length (~100 bp to 1 kb) (**c**). Below we describe a microscopy test that can be employed to qualitatively evaluate successful fragmentation of nuclei into smaller complexes prior to deciding whether to proceed with the SPRITE procedure.

The three samples we recommend preparing for this QC are as follows:

1 No cells (to measure background staining)
2 Positive control: nonsonicated intact cells following permeabilization from Step 36
3 SPRITE sample post-sonication from Step 39

DNA stain

1 Draw three small circles on one side of a mounting glass slide with a fine-tip pen. These circles will mark where to spot cells.
2 Dilute 0.1 mg/mL poly-D-lysine stock 1 to 5 in ddH_2O to a final concentration of 0.02 mg/mL.
3 On the other side of the slide, add 5 μL of 0.02 mg/uL poly-D-lysine inside each of the circles to coat the slide.
4 Dry slides at RT for 1 h or until poly-D-lysine has completely dried.
5 Place 2 μL of sonicated cell lysate on top of poly-lysine spot.
▲ **CRITICAL STEP** 2 μL of SPRITE cell lysate corresponds to 33,000 cells. We have successfully imaged down to 2,000 cells with this procedure.
6 Dry slides at RT for 1 h or overnight.
7 Incubate dried slides for 1 h in Coplin jars with one of the following DNA dye solutions to visualize the genomic DNA in intact cells or fragmented nuclei:
- Hoechst 33342 solution at a recommended dilution in 1× PBS of 1:2,000 (Ex350/Em461)
- SYBR Gold at a recommended dilution in 1× PBS of 1:10,000 (Ex300/Ex494/Em537)
- YOYO-3 Iodide at a recommended dilution in 1× PBS of 1:10,000 (Ex612/Em631)

8 Wash slide by incubating in a Coplin jar filled with 1× PBS for 10–30 min.
9 Remove the slide from the Coplin jar.
10 Mount with mounting medium (e.g., ProLong Gold Antifade Mountant) and place a cover glass on top. Optional: mounting medium containing DAPI may also be used as an additional stain for DNA.
11 Image slides with a confocal microscope. We recommend imaging using a 100× oil objective on a confocal microscope (e.g., LSM800) acquiring *z*-sections (512 × 512) to image the nuclei. See representative images in **a** for examples of intact nuclei (presonication), proper cross-linked complex sizes following light sonication (red box) and two samples devoid of visible large cross-linked complexes (oversonicated).

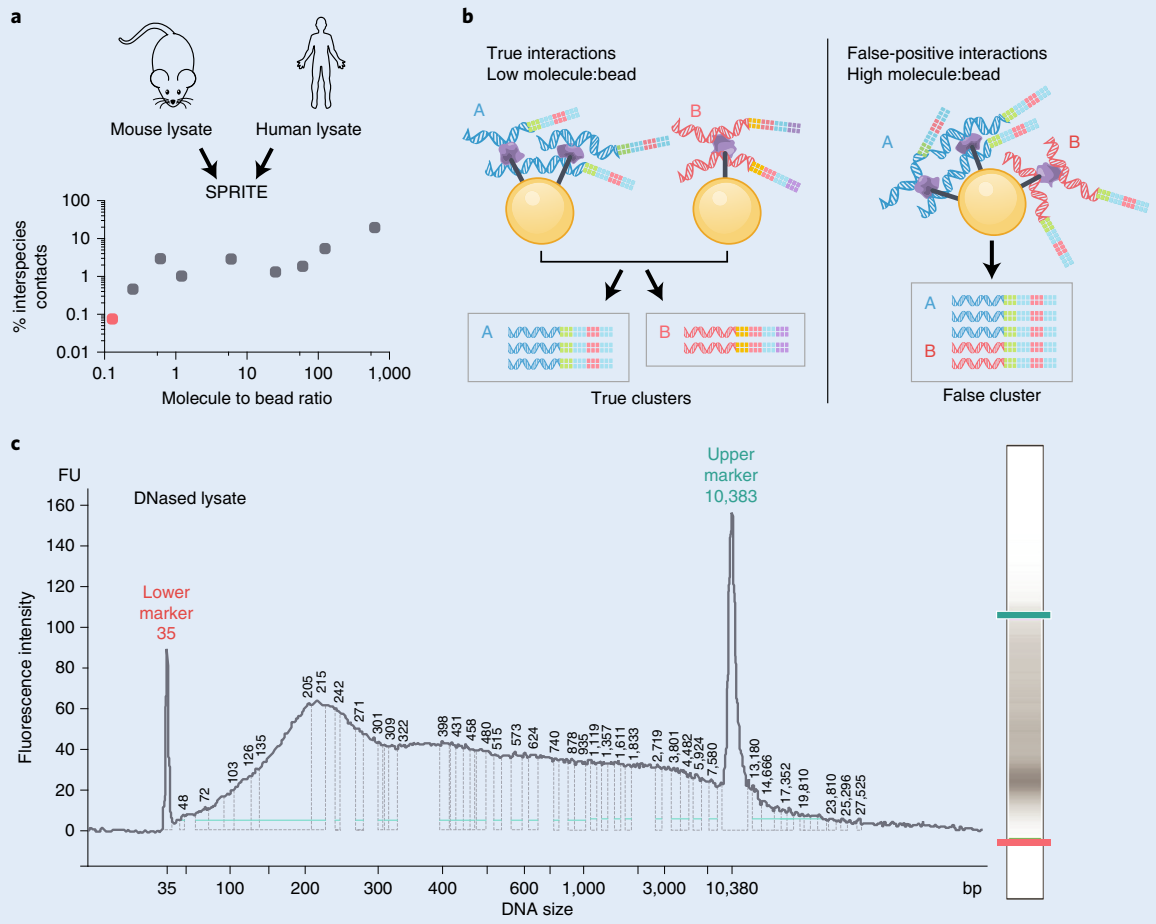
any multiway DNA interactions (singlet SPRITE clusters), whereas undersonication can result in libraries where entire nuclei are tagged rather than smaller-scale structures in the nucleus. See Box 2 for details on how to assay cell fragmentation using microscopy prior to SPRITE and the relationship between sonication and SPRITE cluster sizes. We also note that certain ChIP-seq protocols and other chromatin fragmentation protocols include high-speed spins following sonication. However, as this will result in loss of SPRITE clusters and cause loss of multiway DNA interactions in SPRITE, such a procedure is strongly discouraged.

DNase treatment might vary in Turbo DNase concentration, depending on cell number, ploidy, cross-linking strength and the desired DNA fragment size. To optimize DNase timing and amount, treat samples with varying enzyme concentration for 20 min, immediately quench with EDTA and egtazic acid (EGTA) on ice, and assay the resulting DNA sizes as described in the protocol (Steps 41–53). We recommend optimizing DNA fragment sizes between 100 bp and 1 kb (Box 2, panel c). If an appropriate combination of solubilization and DNA fragment sizes cannot be obtained by varying the extent of sonication or the amount of DNase, then the amount of lysate and the strength of the cross-linking may be reduced instead.

NHS coupling

Cross-linked lysate is coupled to NHS-activated magnetic beads overnight. It is important to measure the molarity of DNA molecules in 50% of the DNAse-treated lysate prior to NHS-bead coupling to ensure that multiple cross-linked complexes do not become covalently coupled to the same bead, which would result in spurious interactions (Box 3). The NHS beads allow us to perform several enzymatic steps by adding buffers and enzymes directly to the beads and performing rapid buffer exchange between each step on a magnet. All enzymatic steps are performed with shaking at 1,600 rpm using an Eppendorf Thermomixer to avoid bead settling and aggregation. All enzymatic steps are inactivated by adding 1 mL modified RLT buffer, and then the NHS beads are buffer

Box 3 | Molecule-to-bead calculations for NHS coupling with reduced rates of false-positive interactions



Coupling the correct number of molecules to beads is critical for ensuring a successful SPRITE dataset. When performing SPRITE for the first time, it may be worthwhile to do a human–mouse (or other interspecies) mixing experiment to assess the frequency of cross species pairs as a proxy for nonspecific pairing (**b**). In a human–mouse mixing experiment, we have found that using a molecule-to-bed ratio that is too high results in an increased rate of human and mouse interspecies false-positive interactions (**a**). Panel **a** adapted with permission from ref. [14]. Specifically, molecule-to-bead ratios ≥1 promote attachment of multiple complexes on the same bead. Instead, low molecule-to-bead ratios (e.g., ≤1 molecule per bead) ensure that a maximum of one cross-linked complex per bead is coupled to obtain a reduced number of interspecies pairing.

In a typical SPRITE experiment, we aim to bind at a 1:4 to 1:1 ratio of DNA molecules:beads. Assuming that we have 50% NHS coupling efficiency, the coupling ratio is then 1:8 to 1:2 molecules per bead. Generally, we bind between 9.5 and 19 billion molecules to 19 billion beads, corresponding to 2 mL of Pierce NHS-activated beads. We note that we use molecules as an overestimate to reduce noise per bead. This is because there should be far more molecules than cross-linked complexes in the SPRITE lysate. As a result, the complex-to-bead ratio should be even lower than stated above.

The number of DNA molecules in the lysate is estimated from the DNA concentration of the lysate after DNase digestion, reverse cross-linking and column purification. The large volume of lysate (600 µL) is recommended for multiple optimization trials of DNAse fragmentation conditions. The 10 µL of lysate in each DNase titration is needed to ensure there is enough DNA to accurately quantify concentration and average size of DNA without any amplification. We measure DNA concentration after lysate is DNase-treated, cleaned and concentrated in 10 µL H_2O in Steps 41–47 of the protocol. Steps 48 and 49 from the protocol post-DNAse digestion determine the concentration and size distribution of DNA (**c**) in each sample using Qubit HS DNA dye and the Agilent DNA Bioanalyzer or TapeStation, respectively. We have provided Supplementary Table 1 to estimate the number of molecules present in the remaining 10 µL of lysate and calculate the volume of lysate to couple to NHS beads. We have also provided Extended Data Fig. 1 for an example DNA distribution from fragmenting DNA from human GM12878 cells.

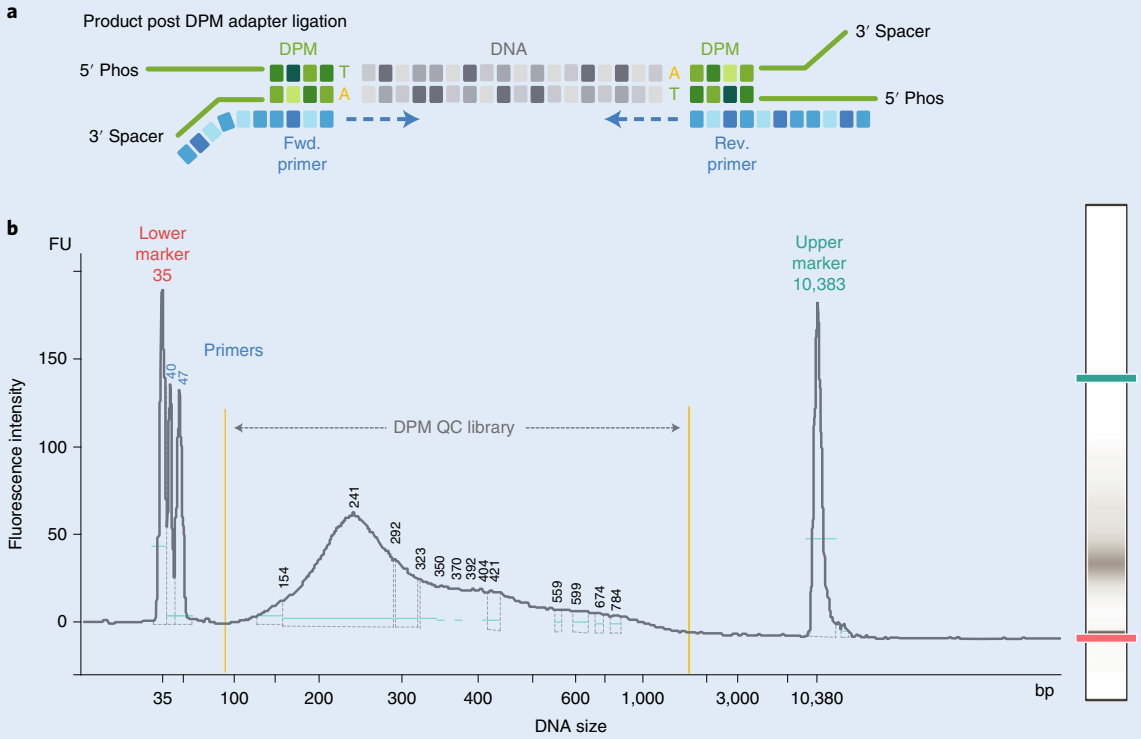
exchanged with three washes of 1 mL SPRITE wash buffer. Washing steps are shown in detail, including recommendations on how to avoid bead loss during each step, in the video protocol (Supplementary Video 1).

DPM adaptor ligation

After chromatin fragmentation, DNA is blunt-ended and dA-tailed using DNA end repair and dA-tailing enzymes. This single dATP is necessary to ligate the DPM adaptor in the next step. We

Box 4 | QC 2: determine ligation efficiency of the DPM adaptor ● Timing 4 h

After DPM ligation, we recommend removing a 5% aliquot of the ligated material and performing a PCR to amplify DPM-ligated material. This will ensure that DPM was successfully ligated prior to proceeding with the subsequent split-and-pool ligation steps.



1 Continuing from Step 94, place sample on magnet, remove the SPRITE wash buffer and resuspend beads in 50 µL of SPRITE ProK buffer.
2 Add ProK enzyme and additional SPRITE ProK buffer to elute DNA by reversal of cross-links through heating and ProK.

Stock solution	Volume
Sample on beads in SPRITE ProK buffer	50 µL
SPRITE ProK buffer	45 µL
ProK	5 µL
Total	100 µL

3 Incubate at 65 °C for a minimum of 2 h, shaking at 1,600 rpm.
▲ CRITICAL STEP You may reverse cross-link for 55 °C for 1 h, then increase the temperature to 65 °C for an overnight incubation (>12 h). We highly recommend reverse cross-linking for at least 2 h.
4 Place the microcentrifuge tube on a magnet and capture the beads. Transfer the supernatant containing the DNA to a clean 1.7 mL microcentrifuge tube.
5 Rinse the beads by pipetting 25 µL of H_2O into the tube containing the beads. Vortex, briefly centrifuge, and recapture the beads. Transfer the 25 µL of H_2O that now contains any residual nucleic acid and combine with the previous supernatant in the new sample tube. Discard the beads.
6 Clean the DNA by following the protocol provided in the Zymo RNA Clean and Concentrator Kit (>17 nt), binding with two volumes of RNA binding buffer and one total volume of 100% ethanol. Elute twice in 20 µL H_2O for a combined eluate of 40 µL.
7 Mix the following reagents to amplify the DNA molecules that are ligated to DPM. The forward primer should prime off the 5′ end of the DPM adaptor, and the reverse primer should prime off the 3′ end of the DPM adaptor.

Stock solution	Volume
Sample (cleaned)	20 µL
DPMQC forward primer (10 µM)	2.5 µL
DPMQC reverse primer (10 µM)	2.5 µL
NEB 2× Q5 Hot-Start Master Mix	25 µL
Total	50 µL

Box 4 | QC 2: determine ligation efficiency of the DPM adaptor ● Timing 4 h (Continued)

8 Amplify with the following PCR program:

Cycle	Initial denaturation	Denature	Anneal	Extend	Final extension	Hold
1	98 °C, 120 s					
2–17 (16 cycles)		98 °C, 10 s	66 °C, 30 s	72 °C, 40 s		
18					72 °C, 120 s	
19						4 °C, ∞

9 The next step is to clean the PCR reaction and size select for your target DNA molecules. Our DPM adaptors are 30 base pairs each, and our target DNA molecules no less than 50 base pairs. Agencourt AMPure XP beads size select while cleaning the PCR reaction of unwanted products.

10 Add 2.0× volume (100 µL) of AMPure XP beads to the sample for a total volume of 150 µL and mix thoroughly by pipetting up and down at least five times.

11 Incubate for 10 min at RT.

12 Place the tube on an appropriately sized magnet to capture the beads and the bound DNA. Wait a few minutes until all the beads are captured.

13 Remove the supernatant and discard.

14 Wash beads twice with 200 µL of 80% ethanol by pipetting ethanol into the tube while beads remain captured on the magnet, moving the tube to the opposite side of the magnet so that beads pass through the ethanol, and then removing the ethanol solution.

15 Quickly spin down the beads in a microcentrifuge, recapture on magnet and remove any remaining ethanol.

16 Air-dry beads while the tube is on the magnet until completely dry.

17 Elute the amplified DNA by resuspending the beads in 12 µL of H_2O. Place the solution back on the magnet to capture the beads. Remove the eluted amplified DNA and place in a clean microcentrifuge tube.

18 Determine the concentration of amplified DNA by measuring 1–2 µL of the cleaned PCR library following the directions provided with the Qubit dsDNA HS Assay Kit.

19 Determine the size distribution of DNA with the DPM adaptor in each sample following the directions provided with either the High Sensitivity DNA Kit for the Agilent Bioanalyzer or the High Sensitivity (HS) D1000 ScreenTape for the Agilent 2200 TapeStation. The average size should be roughly similar to the average size of the input lysate (~200–400 base pairs).

20 Estimate the number of unique molecules pre-amplification using Supplementary Table 1 to confirm that DPM has been successfully ligated. ▲ **CRITICAL STEP** The 5% aliquot should contain, at the very minimum, 10–15 million unique DNA molecules to proceed with SPRITE. If the aliquot contains less than this number, there will not be enough unique reads to sequence the SPRITE library. If this is the case, consider whether a mistake was made during ligation of the DPM adaptor or during one of the critical steps of cross-linking, lysis or NHS bead coupling.

DPM primers for QC of DPM ligation

These primers are used to verify that the DPM adaptor has been successfully ligated to DNA in the lysate. If libraries are undetectable at this step after 16 cycles of PCR on a Bioanalyzer trace, we strongly recommend not to proceed as subsequent ligation of tags and amplification of tagged DNA during the SPRITE protocol will be unsuccessful.

DPMQCprimerF 5′ TACACGACGCTCTTCCGATCT 3′

DPMQCprimerR 5′ TGACTTGTCATGTCTTCCGATCT 3′

The forward and reverse primers amplify the top strand and bottom strand of the DPM adaptor, respectively (see Box 1).

highly recommend performing a DPM ligation quality control (QC) (see Box 4) prior to starting split-and-pool barcoding to ensure DPM ligation was successful and that there are enough DNA molecules ligated with DPM for the subsequent SPRITE tag ligations.

Split-and-pool barcoding and final library amplification

The SPRITE method works by splitting cross-linked lysate across a 96-well plate (Fig. 2). Each well of the 96-well plate contains a unique tag (DPM) to which the DNA molecules are ligated. The ligation reactions are stopped, pooled and split again into a new 96-well plate containing different, unique tags. After DPM ligation, the second round of ligation is with the plate of Odd tags, followed by ligation of Even tags on the third round. This is repeated over multiple rounds, where the Odd and Even tags are designed to be alternated over multiple rounds of ligation and are named as such as they ligate the first, third, fifth, etc., and second, fourth, sixth, etc., rounds, respectively. If n rounds of tag ligation are performed, 96^n unique barcodes are generated.

We typically ligate five tags, creating >8 billion unique barcodes. To reduce the cost of barcoded oligos purchased for SPRITE, a single Odd and a single Even 96-well plate of tags may be purchased and ligated over multiple rounds, rather than purchasing an additional set of barcoded DPM and Terminal barcode. In this case, we recommend purchasing a few DPM and Terminal barcodes and performing those ligations in a single tube (rather than across 96 wells) at their respective steps. This is necessary to add on the required sticky-end sequence for either initiating the split-and-pool

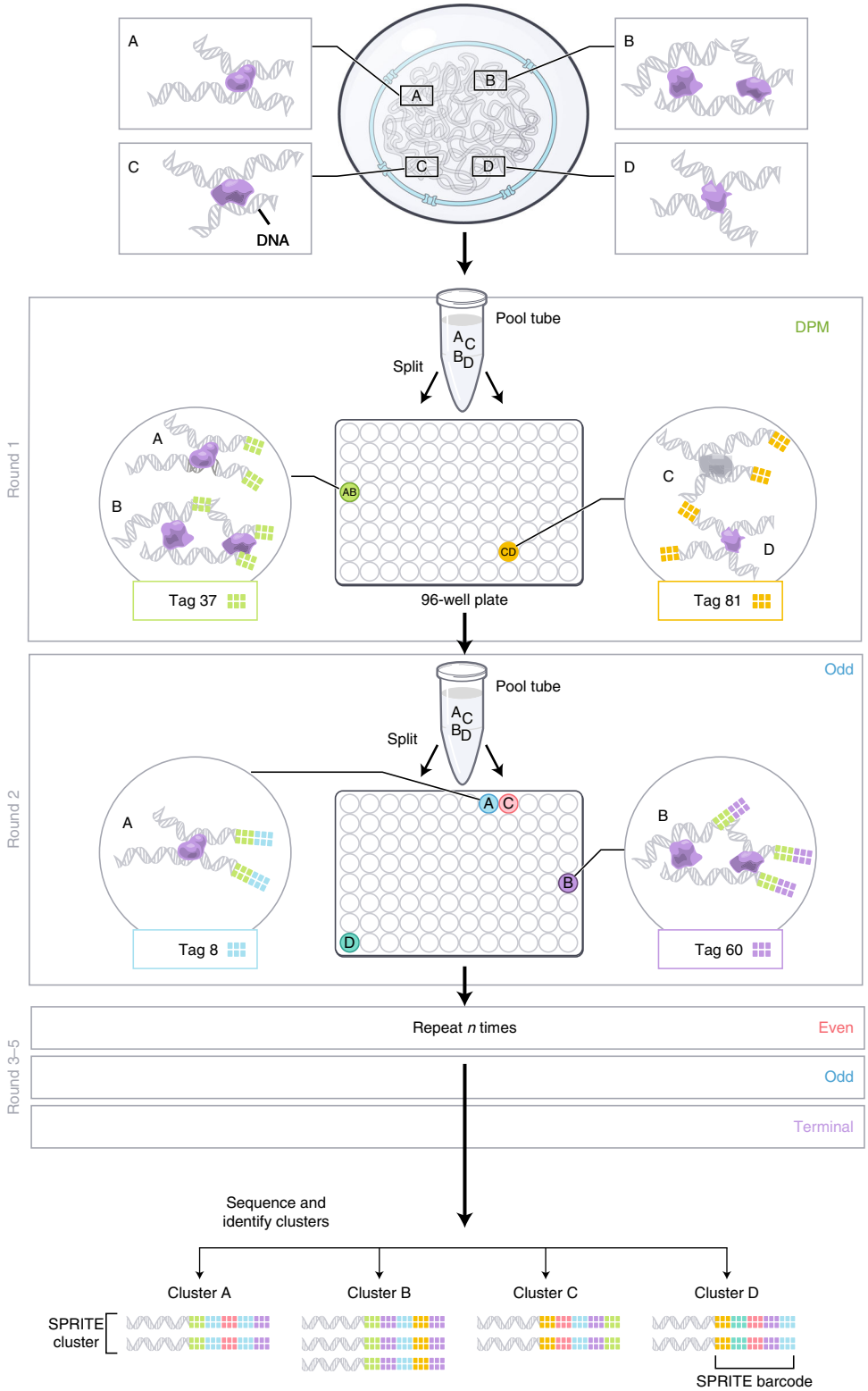
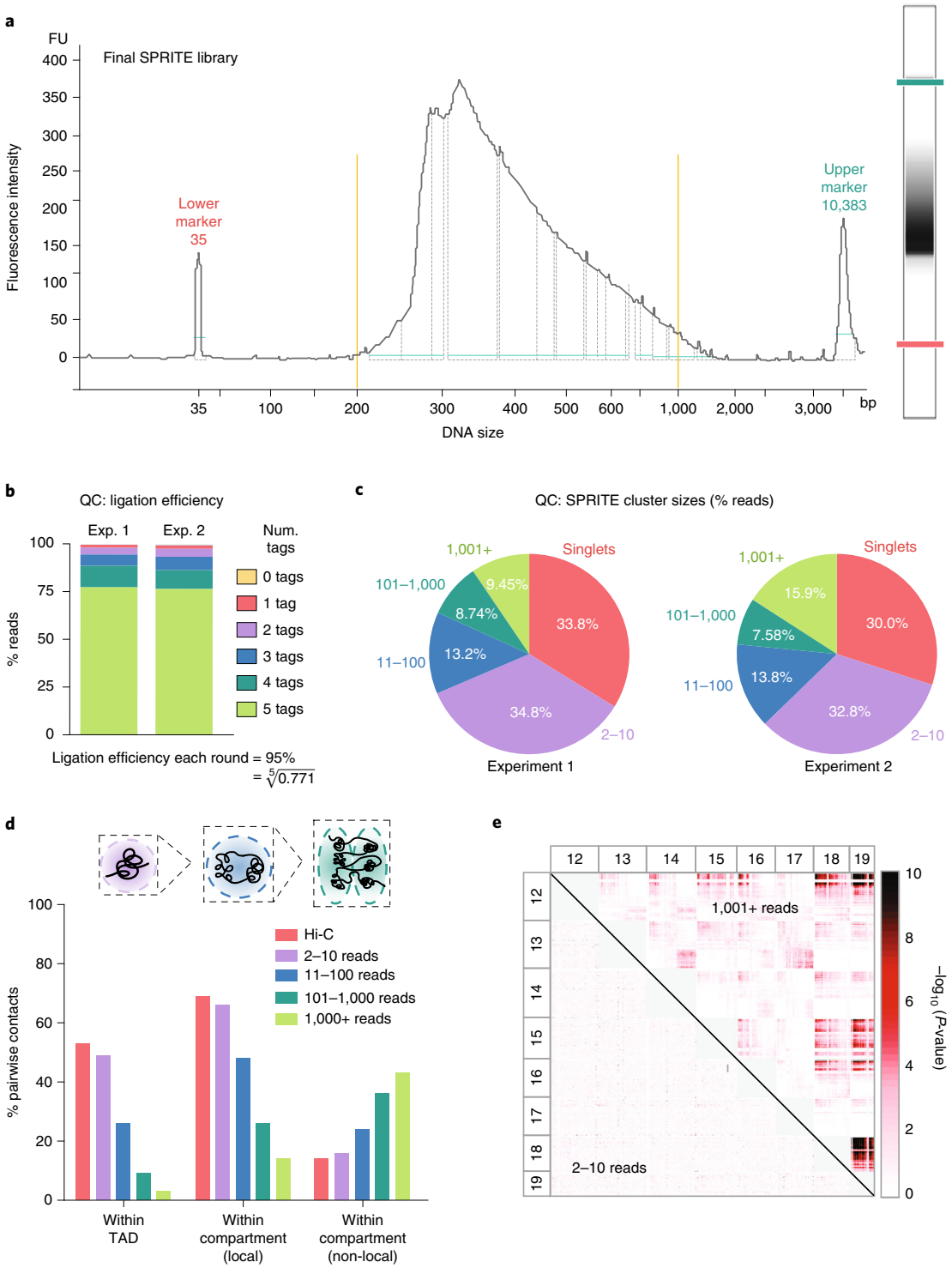


Fig. 2 | Schematic of split-pool procedure. Split-and-pool barcoding works by splitting cross-linked complexes across a 96-well plate containing 96 unique tags, ligating a unique sequence (colored tag) to each DNA molecule, and pooling all cross-linked complexes into a single tube. This split-and-pool process is repeated over multiple rounds, sequentially adding an additional tag each round. Because all molecules within a cross-linked complex are covalently attached, they will sort across the same wells during each round of the split-and-pool process and will obtain the same series of tags, which we refer to as a SPRITE barcode. Genomic DNA fragments and their associated barcodes are then sequenced. All reads sharing the same SPRITE barcodes are matched to generate SPRITE clusters.



ligations (DPM adaptor) or performing Illumina sequencing (Terminal adaptor), without purchasing an entire 96-well plate of sequences for split-and-pool barcoding.

SPRITE requires highly efficient ligations to barcode all DNA molecules within a cross-linked complex. For example, a low ligation efficiency of ~20% during each round would compound to a ligation efficiency of 0.032% ($0.2^{5\ \text{rounds}}$) of all five tags during split-pool barcoding. We have optimized the ligation conditions to be highly efficient (~95%) each round. Specifically, all SPRITE tags contain 7 nt 'sticky ends', which we found promotes highly efficient ligation at 20 °C due to an

◀ **Fig. 3 | Summary of alignment statistics. a**, An example Bioanalyzer profile of a final SPRITE library after PCR amplification. **b**, A summary of ligation efficiency statistics is outputted as a QC step from the SPRITE pipeline to confirm tags have successfully ligated to each DNA molecule. The distribution of reads containing zero, one, two, three, four or five SPRITE tags is shown for two independent SPRITE experiments. Ligation efficiency at each round (95%) is calculated by taking the fifth root of the fraction of reads containing five tags (77.1%). **c**, SPRITE cluster sizes are outputted as a QC step from the SPRITE pipeline to confirm interactions have been successfully detected. Individual SPRITE clusters contain all reads sharing the same barcode. The number of reads sharing the same barcode within an individual cluster can range from 1 read per cluster (singlets; molecules not interacting with other molecules, red), 2–10 reads per cluster (purple), 11–100 reads per cluster (blue), 101–1,000 reads per cluster (dark green) to >1,000 reads per cluster (light green). The percentage of reads that correspond to different SPRITE cluster sizes is shown for two independent experiments generated on 3% FA-DSG samples sonicated for 1 min, 4–5 W as described in the procedure. Successful experiments typically detect a distribution of cluster sizes similar to those shown here. **d**, The number of DNA molecules in a SPRITE cluster reflects the distance at which DNA molecules are interacting in the nucleus. Specifically, smaller SPRITE clusters primarily capture close-range (within TAD) interactions, whereas larger clusters capture longer-distance interactions within A or B compartments (local and nonlocal). Intrachromosomal contacts between Hi-C and SPRITE are calculated for various SPRITE cluster sizes. **e**, Interchromosomal contacts were computed for different SPRITE cluster sizes, and *P*-values were generated to identify significant interchromosomal interactions. SPRITE clusters containing >1,000 reads per cluster are highly enriched for interchromosomal interactions that occur between chromosomes that organize around the nucleolus (12, 15, 16, 18 and 19 in mES cells), whereas clusters containing 2–10 reads per cluster are depleted for interchromosomal interactions. Panel **a** reproduced with permission from ref. [14], Elsevier; panels **c**–**e** adapted with permission from ref. [14], Elsevier.

optimal melting temperature (T_m). In a successful SPRITE experiment, we typically achieve >75% total reads tagged with all five barcodes identified, which corresponds to a ligation efficiency of ~95% at each round ($0.95^{5\ \text{rounds}} = 0.75$) (Fig. 3b).

A final step of barcoding involves splitting the beads after ligation of the terminal tag into smaller aliquot sizes, reverse cross-linking overnight, column purification and PCR amplification of the barcoded DNA using distinct Illumina sequencing primers for each SPRITE aliquot. This effectively serves as another round of split-pool barcoding as each library is tagged with a unique barcoded Illumina primer, further reducing the probability that molecules in different clusters obtain the same barcodes.

SPRITE data processing, cluster generation and heatmap generation

SPRITE libraries (Fig. 3a) can be sequenced on any high-throughput sequencing platform. The key parameters are ensuring that read length is long enough to read all barcodes and their associated genomic DNA regions. We typically sequence SPRITE libraries using Illumina paired-end sequencing on the HiSeq 2500 or NextSeq 500 with 100 × 100 bp, but read length can be adjusted based on the sequencing kit available and number of barcodes ligated. The reads have the following structure: read 1 contains genomic DNA and the DPM tag, and read 2 has the remaining Odd, Even and Terminal tags.

The SPRITE computational pipeline is described below (Fig. 4a).

Adaptor trimming

Illumina adaptors are removed using Trim Galore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). To remove DPM sequences from the 5′ end of read 1 and to account for short genomic DNA fragments that could lead to readthrough of tag sequences on the 3′ end of read 1, we perform a second round of trimming using Cutadapt[32] following barcode identification.

Barcode identification

This step identifies the barcodes on each of the sequenced reads. SPRITE barcodes are identified by parsing the first DNA tag sequence from the beginning of read 1 and the remainder of the tags from read 2. We identify these tag sequences by searching each read for the sets of DPM, Odd, Even and Terminal tags that were ligated to these samples. We allow for up to two mismatches in each Odd and Even tag to account for possible sequencing errors. Because the tags were designed to contain at least four mismatches to any other tag sequence, this enables robust barcode identification. For DPM and terminal tag alignments, we do not tolerate any mismatches due to their shorter unique barcode sequences. After barcode identification, the barcode string is appended to each read name in the FASTQ file. We exclude any reads that do not contain a full set of all ligated tags (DPM, Odd, Even, Odd, Terminal) in the order expected from the experimental procedure. Ligation efficiency corresponding to the percentage of reads (ideally >75%) containing a full set of SPRITE barcodes is calculated.

Cutadapt DPM trimming

To ensure reads are properly aligned to genomic DNA, DPM sequences are trimmed from the beginning of read 1.

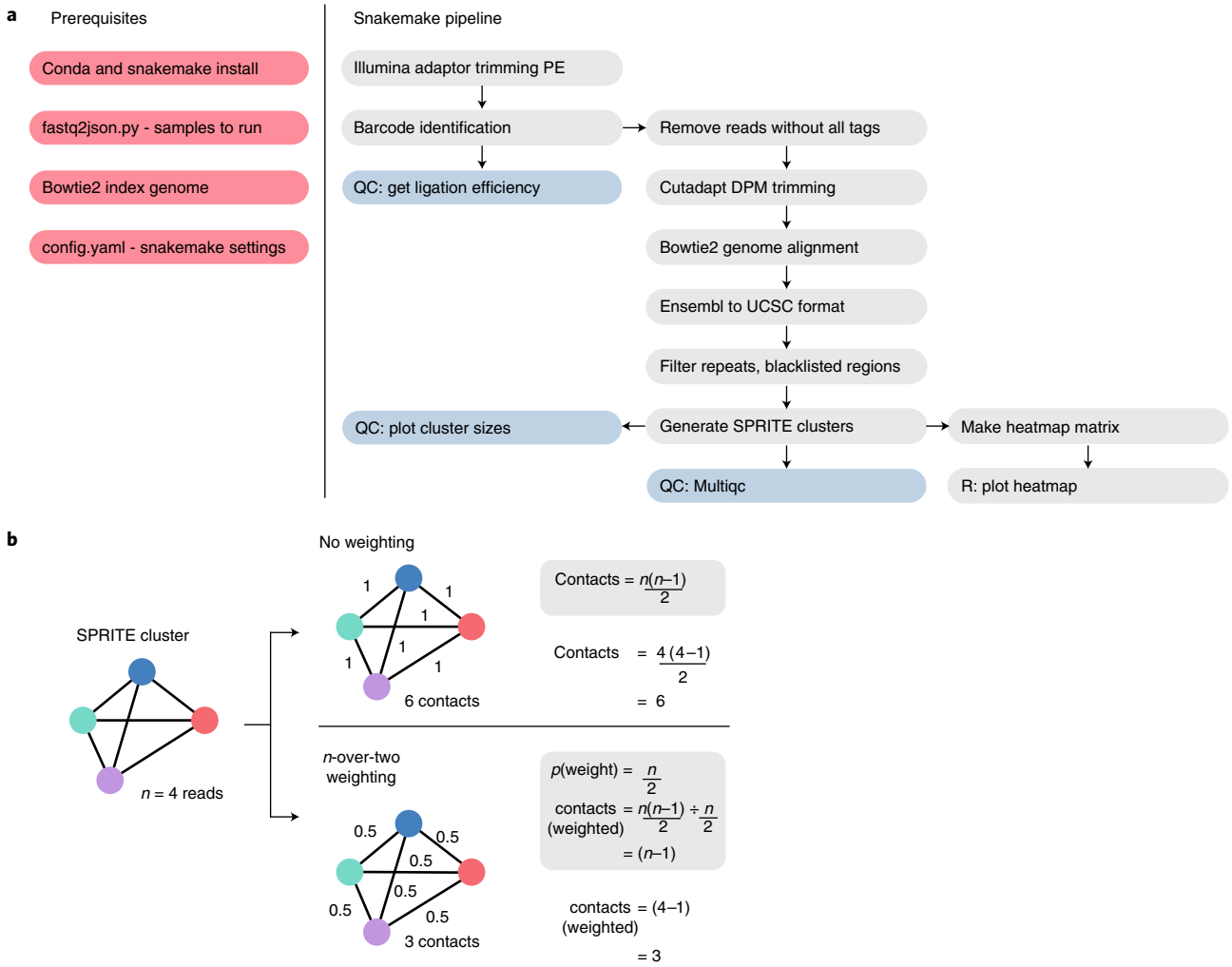


Fig. 4 | Computational SPRITE Pipeline. a, The SPRITE Snakemake pipeline works as follows: the prerequisites (red boxes) before starting include installing conda and Snakemake. Then, run fastq2json.py to compile the paths of all SPRITE fastq files into a .json file. Generate and index a Bowtie2 genome for genome alignments (mm10 and hg38 are currently supported). For each experiment, modify the config.yaml file to input the total number of tags corresponding to the number of ligation rounds and also the location of the reference genome index for alignment. The pipeline is then launched (gray boxes) where (1) adaptor trimming is performed, (2) barcodes are identified, (3) reads without all SPRITE tags are removed, (4) the DPM tag is trimmed from the beginning of read 1, (5) all reads are aligned to the genome using Bowtie2, (6) chromosome annotations are converted from Ensembl to UCSC format, (6) all regions that do not fall within repeat-masked or blacklisted genomic bins are retained, (7) all reads sharing the same barcodes are matched and collapsed into SPRITE clusters and (8) heatmaps are generated. Certain QC files are outputted along the way to quantify ligation efficiency and plot SPRITE cluster sizes, and all summary statistics are outputted in MulitQC. **b**, Because the number of pairwise contacts scales quadratically with the number of reads (n) contained within a SPRITE cluster, larger clusters will contribute a disproportionately large number of the contacts observed between any two bins. To account for this, we downweighted each of the possible pairwise contacts that can be enumerated from a single SPRITE cluster by the cluster size in which it is observed. This is achieved by downweighting each pairwise contact by $2/n$. In this way, the total contribution of pairwise contacts from a cluster is proportional to the minimally connected edges in the graph and will each contain have $n - 1$ contacts. This also ensures that the number of pairwise contacts contributed by a cluster is linearly proportional to the number of reads within a cluster. Here, we show an example of SPRITE cluster weighting using a cluster of four reads, where the total number of contacts is six. Our SPRITE cluster weighting scheme computes all six possible contacts and downweights each contact by $n/2$, such that the total number of contacts sums to three.

Bowtie 2 genome alignment

We align each trimmed read to the mouse or human reference genomes (mm10 for mouse and hg38 for human) using Bowtie2. The corresponding SAM file is sorted and converted to a BAM file using SAMtools.

Filter repeats, blacklisted regions

We filter the resulting BAM file for low-quality reads, multimappers and repetitive sequences. First, we remove all alignments with a MAPQ score <20. Then, we remove all alignments overlapping genomic regions that are blacklisted[33] or masked by Repeatmasker (UCSC, milliDiv <140) using

bedtools[34]. Researchers interested in mapping repetitive elements or other regions that we routinely filter can adapt our pipeline to address their unique analytical needs.

Generate SPRITE clusters

To define SPRITE clusters, all reads that share the same barcode sequence are matched and grouped into a single cluster. To remove possible PCR duplicates, all reads within a cluster containing the same genomic position are removed. We generate a SPRITE cluster file for each SPRITE aliquot separately, as interactions can only occur between DNA molecules derived from a single SPRITE aliquot rather than between different SPRITE aliquots. All subsequent analyses are performed on the SPRITE cluster file, where each cluster occupies one line of the resulting text file containing the barcode name and genomic alignments. SPRITE clusters capture various multiway contacts between multiple reads that all share the same barcode. A typical SPRITE experiment will yield >65–70% of reads with at least two interactions per cluster. These clusters range in size from 2 to 10 reads in a single SPRITE cluster (~30% of reads), and a distribution of cluster sizes ranging from 11 to 100 reads, 101 to 1,000 reads, and up to >1,000 reads in a single SPRITE cluster (10–15% of reads) (Fig. 3c). If most reads are unpaired and devoid of interactions or, conversely, primarily within large clusters of >1,000 reads in a single SPRITE cluster, consult the troubleshooting advice in Table 1.

Make heatmap matrix

The cluster file is used to generate a pairwise contact matrix that can be visualized as a heatmap. Heatmaps are generated by enumerating all pairwise contacts within each SPRITE cluster either with or without downweighting of individual contacts. Because the number of pairwise contacts enumerated from a SPRITE cluster scales quadratically with the number of reads, downweighting each contact by the cluster size ensures that larger clusters do not dominate the heatmap (Fig. 4b). We recommend using '*n*-over-two' weighting for initial heatmap generation to visualize known genome structures (e.g., TADs, compartments, chromosome territories).

Visualizing SPRITE clusters

We identify multiway interactions by counting how often multiple genomic regions simultaneously interact in individual SPRITE clusters. In Fig. 5, rows show individual SPRITE clusters where multiple regions simultaneously interact in the same cross-linked complex. Specifically, black boxes correspond to genomic bins that contain at least one read in a given SPRITE cluster. We observe multiway contacts corresponding to interactions across a range of genomic resolutions, including A compartments, TADs and chromatin loops.

Expertise needed to implement the protocol

The SPRITE procedure can be performed by any molecular biology laboratory with access to the reagents and equipment described in the protocol. Access to a high-throughput sequencing machine is needed to sequence final SPRITE libraries. Furthermore, using the SPRITE computational pipeline requires modest experience using the Unix operating system to install and run data analysis software.

Limitations

One of the limitations of SPRITE is the high sequencing depth needed to identify DNA structural patterns like chromosome loops. Many genomic architecture mapping methods are limited by the need for deep sequencing in order to detect high-resolution features. For example, we need approximately 5 million reads to visualize chromosome territories, 25 million reads to visualize compartments, 200 million reads for TADs and 1 billion reads for loop interactions. We recently adapted SPRITE to combine it with enrichment procedures, similar to those implemented for ChIA-PET[35], Hi-ChIP[36], Capture Hi-C[37,38] or PLAC-seq[35,36,39], to select for substructures of interest for the specific needs of the experiment (i.e., enhancers and promoters)[23].

A second limitation of SPRITE is the narrow window of chromatin fragmentation optimization to get meaningful contacts. Overfragmentation results in sparse clusters while underfragmentation results in clusters that are too large to generate high-confidence interactions (see Box 2, panels a, b). We highly recommend optimizing and QCing the fragmentation conditions when working with a new cell line, sonicator machine or other variables that may affect DNA size.

There is also a high upfront cost associated with SPRITE, in particular the 96-well plate of DPM adaptors, which must be ordered with a 3′ C3 spacer modification. To reduce the cost, one may

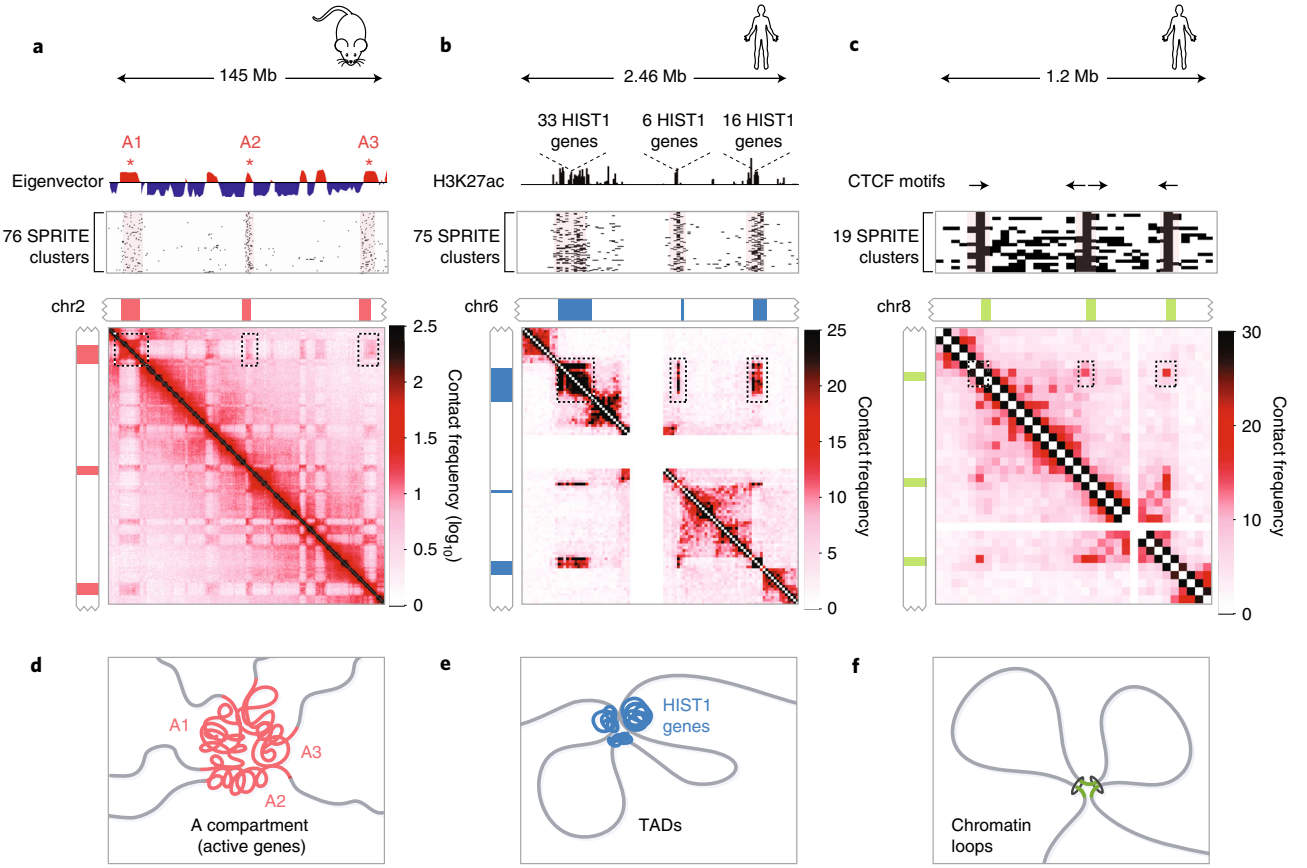


Fig. 5 | SPRITE Identifies higher-order interactions that occur simultaneously. a, Compartment eigenvector showing A (red) and B (blue) compartments on mouse chromosome 2 (top). Individual SPRITE clusters (rows) containing reads mapping to at least three distinct A compartment regions (*) (middle). Pairwise contact map at 200 kb resolution (bottom). Pink bars represent A compartment regions. **b**, H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq, ENCODE) signal across a 2.46 Mb region on human chromosome 6 corresponding to three TADs containing 55 histone genes (top). SPRITE clusters containing reads in all three TADs (middle). Pairwise contact map at 25 kb resolution (bottom). Blue bars represent histone gene regions. **c**, CTCF motif orientations at three loop anchors on human chromosome 8 (top). SPRITE clusters overlapping all three loop anchors (middle). Pairwise contact map at 25 kb resolution (bottom). Green bars represent CTCF motif sites. **d**, Schematic of multiple A compartment interactions. **e**, Schematic of higher-order interactions of HIST1 genes (green). **f**, Schematic of higher-order interactions between consecutive loop anchors. Figure adapted with permission from ref. [14], Elsevier.

modify the protocol to use a single DPM adaptor sequence, performing DPM in a single tube before commencing with split-pool barcoding. Importantly, in this scenario the experimenter will need to utilize an additional round of barcoding to generate enough combinatorial complexity to distinguish unique complexes as well as increasing the sequencing length on read 2 to sequence an additional barcode. Terminal adaptor ligation may also be performed in a single well in exchange for an extra round of Odd or Even ligation; however, for the terminal ligation we recommend purchasing four different terminal adaptors of variable lengths from the set of 96 tags provided in Supplementary Table 2 to introduce a stagger and prevent any monotemplate issues during sequencing from the identical 7 nt sticky-end sequence located within each SPRITE tag (see Box 1). We have found that this greatly improves sequencing quality scores on read 2 for the barcode identification. However, wherever possible, we recommend performing SPRITE as described with the first round of barcoding being the 96-well plate of DPM adaptors.

Materials

Biological materials

- GM12878 human lymphoblasts (Coriell Cell Repositories; RRID: CVCL_7526)
- F1 2-1 hybrid wild-type mES cell line (F1 2-1; 129 × cast; RRID: CVCL_XY63) (generously provided by K. Plath; ref. [40]) **! CAUTION** The cell lines used in your research should be regularly checked to

ensure they are authentic and are not infected with mycoplasma. The cell lines used in this protocol were regularly checked for mycoplasma infection.

Reagents

General reagents

- SPRITE tags (Supplementary Table 2; IDT, custom order) ▲ **CRITICAL** Order all plates of SPRITE tags (DPM, Odd, Even and Terminal) resuspended to 200 μM in nuclease-free water. Use extreme care to avoid cross-contamination between wells at all times when working with stock plates. Always centrifuge stock plates at 1,000*g* for 2 min prior to opening.
- Indexing SPRITE Library Amplification primers (Supplementary Table 2; IDT, custom order)
- Buffer RLT (Qiagen, cat. no. 79216) ! **CAUTION** Modified RLT buffer contains guanidine thiocyanate, which when mixed with bleach produces hydrogen cyanide gas and hydrogen chloride gas. Be careful to ensure that all liquid modified RLT buffer waste is disposed of in its own waste container. Solids that have touched modified RLT buffer such as tips and reservoirs should also be discarded in a separate solid RLT buffer container.
- Calcium chloride ($CaCl_2$) solution, 1 M (Sigma-Aldrich, cat. no. 21115)
- Manganese chloride ($MnCl_2$) solution, 1 M (Sigma-Aldrich, cat. no. M1787) ▲ **CRITICAL** It is imperative that $MnCl_2$ is used for DNase digestion and not Turbo DNase buffer that comes with the kit. Mn(II) generates the appropriate DNA ends necessary for end repair and further downstream library prep steps.
- EDTA, 0.5 M, pH 8.0 (Thermo Fisher Scientific, cat. no. 15575020)
- EGTA, 0.5 M, pH 8.0 (Fisher Scientific, cat. no. 50255957)
- Glycerol (Sigma-Aldrich, cat. no. G5516)
- Sodium chloride (NaCl), 5 M (Thermo Fisher Scientific; cat. no. AM9759)
- UltraPure DNase/RNase-free distilled water (Thermo Fisher Scientific, cat. no. 10977015)
- Triton-X-100 detergent (Sigma-Aldrich, cat. no. T8787)
- NP-40 Surfact-Amps detergent, 10% solution (Thermo Fisher Scientific, cat. no. 28324)
- *N*-Lauroylsarcosine sodium salt solution, 20% solution (Sigma-Aldrich, cat. no. L7414)
- Tris-HCl buffer, 1 M, pH 7.5 (Thermo Fisher Scientific, cat. no. 15567027)
- Tris-HCl buffer, 1 M, pH 8.0 (Thermo Fisher Scientific, cat. no. 15568025)
- Hydrochloric acid (HCl) concentrated, 12 M (VWR, cat. no. 470301-260) ! **CAUTION** Do not breathe mist or vapors upon opening the concentrated bottle. Use only in a well-ventilated area with gloves and eye and face protection. Store in a corrosion-resistant container.
- Ethanolamine (Sigma-Aldrich, cat. no. E9508)
- Sodium deoxycholate (DOC); Sigma-Aldrich, cat. no. D6750) ! **CAUTION** DOC should be handled with gloves in the chemical hood and/or with an N95 dust mask.
- Lithium chloride solution, 8 M (Sigma-Aldrich, cat. no. L7026)
- Sodium dodecyl sulfate (SDS), 20% solution (Thermo Fisher Scientific, AM9820)
- HEPES buffer, pH 7.4, 1 M (Teknova, cat. no. H1030)
- RNase-Free BSA (American Bio, cat no. AB01243-00050)
- Phosphate buffered saline (PBS), pH 7.4 (1×), no calcium, no magnesium, liquid (Thermo Fisher Scientific, cat. no. 10010049)
- Disuccinimidyl glutarate (DSG), 50 mg (Thermo Fisher Scientific, cat. no. 20593)
- Dimethyl sulfoxide (DMSO) (Sigma-Aldrich, cat. no. D2650)
- Formaldehyde ampules, 16%, methanol-free (Thermo Scientific Pierce, cat. no. PI28908) ! **CAUTION** Formaldehyde is highly toxic if swallowed or inhaled and causes burns upon skin contact. Work with formaldehyde should be done in a fume hood. Dispose of all formaldehyde waste in a secondary waste container. Store in a corrosion-resistant container.
- Glycine, >99% (Sigma-Aldrich, cat. no. G7403-250G)
- Protease cocktail inhibitor tablets (Sigma-Aldrich, cat. no. 04693159001)
- TURBO DNase, 2 U/μL (Thermo Fisher Scientific, cat. no. AM2238)
- Proteinase K, Molecular Biology Grade (ProK), 800 U/mL (New England Biolabs, cat. no. P8107S)
- Pierce NHS-activated magnetic beads (Thermo Fisher Scientific, cat. no. 88827) ▲ **CRITICAL** We strongly recommend sealing opened bottles with parafilm and storing the bottle with desiccant in a 50 mL conical tube. Magnetic beads are moisture sensitive.
- NEBNext End Repair Module (contains end repair enzyme mix and reaction buffer; New England Biolabs, cat. no. E6050)

- NEBNext dA-Tailing Module (New England Biolabs; cat. no. E6053)
- Instant Sticky-End Ligase Master Mix (New England Biolabs; cat. no. M0370)
- NEBNext quick ligation reaction buffer (New England Biolabs; cat. no. B6058S)
- 1,2-Propanediol (Sigma-Aldrich, cat. no. 398039) ▲ **CRITICAL** Protect 1,2-propanediol from light by placing it inside a drawer.
- Q5 Hot Start High-Fidelity 2× Master Mix (New England Biolabs, cat. no. M0494)
- Agencourt AMPure XP magnetic beads (Beckman Coulter, cat. no. A63880)
- High Sensitivity DNA Kit for the Agilent Bioanalyzer (Agilent Technologies, cat. no. 5067-4626)
- High Sensitivity D1000 ScreenTape and reagents (sample buffer and ladder) for the Agilent 2200 TapeStation (Agilent Technologies, cat. nos. 5067-5584 and 5067-5585)
- Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, cat. no. Q32854)
- (Optional) Poly-D-lysine hydrobromide (Sigma, cat. no. P6407-5MG)
- (Optional) ProLong Gold Antifade Mountant with or without DAPI (Thermo Fisher Scientific; cat. no. P36935)
- (Optional) Micro Slides (VWR Superfrost Plus Micro Slide, Premium, cat. no. 48311-703)
- (Optional) Hoechst 33342 solution (Thermo Fisher Scientific, cat. no. 62249)
- (Optional) SYBR Gold Nucleic Acid Gel Stain (Thermo Fisher Scientific, cat. no. S11494)

(Optional) YOYO-3 iodide (Thermo Fisher Scientific, cat. no. Y3606) ▲ **CRITICAL** Here we provide multiple dye options for staining DNA (Hoechst 33342, SYBR Gold, and YOYO-3 iodide), depending on the availability of the particular dye in your laboratory.

mES cell culture specific reagents

- Trypsin-EDTA (0.25%), phenol red (Thermo Fisher Scientific, cat. no. 25200-056)
- Chicken serum, USA origin, sterile-filtered, cell culture tested (Sigma, cat. no. C5405-100ML)
- DMEM/F-12, 1:1 (Thermo Fisher Scientific, cat. no. 11330-057)
- Gibco BSA Fraction V, 7.5% (Thermo Fisher Scientific, cat. no. 15260037)

GM12878 cell culture specific reagents

- Penicillin–streptomycin, 10,000 U/mL (Thermo Fisher Scientific; cat. no. 15140122)
- Seradigm fetal bovine serum (VWR, cat. no. 97068-091)
- RPMI 1640 medium (Thermo Fisher Scientific, cat. no. 11875119)
- L-Glutamine, 200 mM (Thermo Fisher Scientific, cat. no. 25030081)

Equipment

- Cell culture incubator (37 °C at 5% CO_2 for all cell lines used in this protocol)
- Branson needle-tip sonicator (102C), 1/8″ Doublestep tip, 3 mm diameter (Branson Ultrasonics, cat. no. 101-148-063) ▲ **CRITICAL** Another sonicator may also be used. For example, we have also used a Covaris M220 Focused-ultrasonicator (Covaris, cat. no. 500295). Exact conditions will need to be optimized when using a different sonicator.
- Eppendorf 96-well twin.tec PCR plates, 96-well, semi-skirted (Eppendorf, cat. no. 10049-108)
- Pipetting reservoirs (VWR, cat. no. 10015-232)
- PCR tubes, 0.2 mL (USA Scientific, cat. no. 1402-4700)
- PCR 12-tube strip with 12-cap strips, 0.2 mL (USA Scientific, cat. no. 1402-2408)
- Lo-bind microcentrifuge tubes, 1.5 mL (Eppendorf, cat. no. 022-43-108-1)
- Centrifuge tubes, 15 ml and 50 ml (Genesee Scientific Corporation, cat. nos. 28-103 and 28-108)
- RNA Clean and Concentrator-5 Kit with Capped Columns (Zymo Research, cat. no. R1018)
- E-Gel Power Snap electrophoresis system (Thermo Fisher Scientific, cat. no. G8100)
- E-Gel EX Agarose Gels, 2% (Thermo Fisher Scientific, cat. no. G40100)
- Benchtop microcentrifuge, Myfuge 12 mini microcentrifuge (Benchmark Scientific, cat. no. C1012)
- Plate centrifuge (Eppendorf, cat. no. 5430/5430R)
- Qubit Fluorometer 2.0 (Thermo Fisher Scientific, cat. no. Q32866)
- Thermal mixer C (Thermomixer C; Eppendorf, cat. no. Z605271)
- Eppendorf SmartBlock 1.5 mL and PCR 96 thermoblocks (Eppendorf, part nos. 5360000038 and 5306000006)
- Magnetic racks for 1.5 mL, 15 mL tubes (DynaMag-2 and DynaMag-15 magnets; Thermo Fisher Scientific, cat nos. 12321D and 12301D)
- Magnetic rack for 96-well plate (DynaMag-96 Side Magnet; Thermo Fisher Scientific, cat. no. 12331D)

- Thermal cycler (Eppendorf Mastercycler pro; Eppendorf, cat. no. 95043-592)
- Agilent Bioanalyzer (Agilent Technologies, model no. G2939A)
- Agilent 2200 TapeStation (Agilent Technologies, manual part no. G2964-90003 Rev. C)
- HulaMixer Sample Mixer (Thermo Fisher Scientific; 15920D)
- Vortex mixer (Analog vortex mixer; VWR, model no. 58816-123)
- Sequencer (Illumina, model no. HiSeq 2500)
- Single channel pipettes (P2, P20, P200, P1000, Rainin, cat. no. 30579367)
- 12-well multichannel pipettes (P20, P200, Rainin, cat. nos. 17013808 and 17013810)
- Low-retention pipette tips, 20 µL, 200 µL and 1,000 µL (Rainin, cat. nos. 30389226, 30389240 and 30389213)

Computer, programs and source code

- The SPRITE pipeline has been tested on a high-performance computing cluster running CentOS 7 and a local environment with 30 GB of random-access memory, an i7-8750H CPU running Ubuntu 18.04.3 LTS. Local runtime for fastq files with ~45 million reads was ~7 h
- Snakemake pipeline software (https://snakemake.readthedocs.io/en/stable/)
- Conda package (https://docs.conda.io/projects/conda/en/latest/) or miniconda package (https://docs.conda.io/en/latest/miniconda.html)
- Python 3.7.3 (https://www.python.org/)
- Java 8 (https://www.java.com/en/download/)
- R software v3.6.1 (https://www.r-project.org/)
- Hi-Corrector v1.2 software (https://github.com/jasminezhoulab/Hi-Corrector)
- Bowtie2 v2.3.5 (https://github.com/BenLangmead/bowtie2/releases/tag/v2.3.5)
- Bedtools v2.29.0 (https://github.com/arq5x/bedtools2/releases/tag/v2.29.0)
- Multiqc v1.6 (https://github.com/ewels/MultiQC/releases/tag/v1.6)
- Samtools v1.9 (https://github.com/samtools/samtools/releases/tag/1.9)
- Trim Galore! V0.6.2 (https://github.com/FelixKrueger/TrimGalore/releases/tag/0.6.2)
- Cutadapt v2.5 (https://github.com/marcelm/cutadapt/releases/tag/v2.5)
- Pigz v2.3.4 (https://github.com/madler/pigz/releases/tag/v2.3.4)
- Fastqc v0.11.8 (https://github.com/s-andrews/FastQC/releases/tag/v0.11.8)
- Python packages:
  - Pysam v0.15.0.1 (https://github.com/pysam-developers/pysam/releases/tag/v0.15.0.1)
  - Numpy v1.17.2 (https://github.com/numpy/numpy/releases/tag/v1.17.2)
- R packages
  - Ggplot2 v3.1.1 (https://github.com/tidyverse/ggplot2/releases/tag/v3.1.1)
  - Gplots v3.0.1.1 (https://rdocumentation.org/packages/gplots/versions/3.0.1.1)
  - Readr v1.3.1 (https://github.com/tidyverse/readr/releases/tag/v1.3.1)
  - Optparse v1.6.2 (https://rdocumentation.org/packages/optparse/versions/1.6.2)

Reagent setup

▲ **CRITICAL** We recommend storing all buffers for up to 2 months at room temperature (RT, 22 °C) (except SPRITE lysis buffers, which should be stored at 4 °C) unless otherwise stated here or in a manufacturer's protocol.

10% Triton X-100 (vol/vol)

Prepare 50 mL of 10% (vol/vol) Triton X-100 by diluting 5 mL of Triton X-100 with 45 mL of nuclease-free water. Store at RT for up to 1 year.

10% DOC (wt/vol)

Prepare 10 mL of 10% (wt/vol) DOC by dissolving 1 g of DOC in 10 mL of nuclease-free water. Protect from light and store at RT for up to 1 month. ▲ **CRITICAL** DOC is light sensitive. Cover dissolved DOC with aluminum foil and place inside a drawer to protect DOC from light.

0.5 M DSG

Prepare 0.5 M DSG by adding 306 µL dimethyl sulfoxide to a bottle containing 50 mg DSG. Store remaining DSG at −20 °C for up to 1 year.

Trypsin versene phosphate buffer (TVP)
Prepare 1 mM EDTA, 0.025% trypsin (vol/vol), 1% Sigma chicken serum (vol/vol) in 1× PBS pH 7.4. Store at 4 °C for up to 6 months.

Post-TVP wash solution
Prepare DMEM/F-12 medium supplemented with 0.03% Gibco BSA Fraction V (vol/vol). Store at 4 °C for up to 6 months.

2 mM DSG cross-linking solution
Make an appropriate amount of fresh 2 mM DSG cross-linking solution by adding 16 µL of 0.5M DSG per 4 mL of RT 1× PBS pH 7.4.

3% formaldehyde cross-linking solution (vol/vol)
Prepare a fresh 3% formaldehyde solution by combining 750 µL of 16% formaldehyde for every 3.25 mL RT 1× PBS pH 7.4.

0.5% BSA scraping buffer (wt/vol)
Dissolve 0.5% BSA (wt/vol) in 1× PBS pH 7.4. Store at 4 °C for up to 1 month.

Cell lysis buffer A
Prepare 10 mL of cell lysis buffer A by mixing 500 µL of 1 M HEPES pH 7.4, 20 µL of 0.5 M EDTA, 20 µL of 0.5 M EGTA, 280 µL of 5 M NaCl, 250 µL of 10% (vol/vol) Triton X-100, 500 µL of 10% (vol/vol) NP-40, 1 mL of 100% glycerol and 7.43 mL of nuclease-free water. Store for up to 1 month at 4 °C.

Cell lysis buffer B
Prepare 10 mL of cell lysis buffer B by mixing 500 µL of 1 M HEPES pH 7.4, 30 µL of 0.5 M EDTA, 30 µL of 0.5 M EGTA, 400 µL of 5 M NaCl and 9.04 mL of nuclease-free water. Store for up to 1 month at 4 °C.

Cell lysis buffer C
Prepare 10 mL of cell lysis buffer C by mixing 500 µL of 1 M HEPES pH 7.4, 30 µL of 0.5 M EDTA, 30 µL of 0.5 M EGTA, 200 µL of 5 M NaCl, 100 µL of 10% (wt/vol) DOC, 250 µL of 20% *N*-lauroylsarcosine (wt/vol) and 8.89 mL of nuclease-free water. Store for up to 1 month at 4 °C. ▲**CRITICAL** EDTA and EGTA may be omitted from cell lysis buffer C depending on the application (e.g., if using a cell type in which more complete DNAse digestion is needed, as EDTA and EGTA can inhibit digestion). Enzyme concentrations recommended here for DNA fragmentation have been optimized in this protocol with EDTA and EGTA included in the lysis buffers. ▲**CRITICAL** SDS can be added to permeabilize nuclei. We note, however, that we optimized our protocol without SDS in our lysis buffers.

100× $MnCl_2$/$CaCl_2$ mix
Make a 100× solution of $MnCl_2$ and $CaCl_2$ by mixing 250 µL of 1 M $MnCl_2$, 50 µL of 1 M $CaCl_2$ and 700 µL of nuclease-free water. Store at RT for up to 1 year.

10× SPRITE DNase buffer
Prepare 1 mL of 10× SPRITE DNase buffer by mixing 200 µL of 1 M HEPES pH 7.4, 200 µL of 5 M NaCl, 50 µL of 10% (vol/vol) NP-40, 100 µL of 100× $MnCl_2$/$CaCl_2$ mix and 450 µL of nuclease-free water. Store at RT for up to 1 year.

25× DNase stop solution
Prepare 1 mL of 25× DNase stop solution by mixing 500 µL of 0.5 M EDTA, 250 µL of 0.5 M EGTA and 250 µL of nuclease-free water. Store at RT for up to 1 year.

Hoechst dye stock solution
Prepare the Hoechst dye stock solution by dissolving the contents of one vial (100 mg) in 10 mL of deionized water to create a 10 mg/mL (16.23 mM) solution. ▲**CRITICAL** Hoechst dye has poor solubility in water, so sonicate as necessary to dissolve. The 10 mg/mL Hoechst stock solution may be stored at 2–6 °C for up to 6 months or at ≤−20 °C for longer periods.

Hoechst staining solution
Prepare the Hoechst staining solution by diluting the Hoechst stock solution 1:2,000 in 1× PBS.

SPRITE ProK buffer
Prepare 10 mL of SPRITE ProK buffer by mixing 200 µL of 1 M Tris-HCl pH 7.5, 200 µL of 5 M NaCl, 500 µL of 10% (vol/vol) Triton X-100, 100 µL of 20% (wt/vol) SDS, 200 µL of 0.5 M EDTA, 200 µL of 0.5 M EGTA and 8.6 mL of nuclease-free water.

NHS coupling buffer
Prepare 1 mL of NHS coupling buffer by adding 5 µL of 20% (wt/vol) SDS to 995 µL 1× PBS. ▲ **CRITICAL** Do not place this on ice as doing so will result in SDS precipitation.

1 mM HCl solution
Prepare 1 mL of 1 mM HCl by diluting 1 µL of 1 M HCl in 999 µL ultrapure water.

Modified RLT buffer
Prepare 50 mL of modified RLT buffer by mixing 47.8 mL of Qiagen Buffer RLT with 100 µL of 0.5 M EDTA, 100 µL of 0.5 M EGTA, 500 µL Tris-HCl pH 7.5, 500 µL of 10% (vol/vol) Triton X-100, 500 µL of 10% (vol/vol) NP-40 and 500 µL of 20% (wt/vol) *N*-lauroylsarcosine.

SPRITE wash buffer
Prepare 50 mL of SPRITE wash buffer by mixing 1 mL of 1 M Tris-HCl pH 7.5, 500 µL of 5 M NaCl, 1 mL of 10% (wt/vol) DOC, 1 mL of 10% (vol/vol) Triton X-100, 1 mL of 10% (vol/vol) NP-40 and 45.5 mL of nuclease-free water.

Homemade SPRITE Ligation Master Mix (3.125×)
Prior to pipetting the viscous NEB Quick Ligation Reaction and Instant Sticky-End Ligation Master Mix, thaw all reagents on a rotator at RT until they are completely in solution. Combine 1,600 µL of 5× NEBNext quick ligation reaction buffer, 600 µL of 1,2-propanediol and 1,000 µL 2× Instant Sticky-end Ligation Master Mix. Store at −20 °C for over 2 weeks. This solution can undergo multiple freeze–thaws.

10× annealing buffer
Prepare 10× annealing buffer by mixing 250 µL of 8 M LiCl, 100 µL of 1 M Tris-HCl pH 7.5 and 650 µL of nuclease-free water. Store at RT for up to 1 year.

Poly-D-lysine hydrobromide
Dilute in water to 0.1 mg/mL following the Sigma recommendations.

2.5 M glycine stop solution
Dissolve 9.38 g of glycine in nuclease-free water to 50 mL. Store up to 1 month at RT.

90 µM annealed SPRITE tags (DPM, Odd, Even and Terminal stock plates)
Generate stock plates of all annealed SPRITE adaptors prior to starting split-and-pool steps. To generate a stock 96-well plate of SPRITE tags, anneal each well of the 'top' plate with its corresponding 'bottom' plate (e.g., well A1 'DPM_Top' with well A1 'DPM_Bot'). Specifically, combine 18 µL of each 200 µM 'top' strand with 18 µL of the corresponding 200 µM 'bottom' strand in a new semi-skirted LoBind 96-well plate. Add 4 µL of 10× annealing buffer to each well of the plate. Carefully seal the plate with foil and vortex gently for 30 s to mix. Centrifuge the plate at 1,000*g* for 2 min at RT. Anneal by heating the plate to 95 °C for 5 min and slowly cooling to 25 °C in a thermal cycler with a heated lid (ramp -1 °C/s). Stored at −20 °C, stock plates are stable for multiple years.

4.5 µM annealed SPRITE tags (Diluted DPM, Odd, Even and Terminal stock plates)
Generate 4.5 µM working stock plates by diluting 5 µL of each 90 µM annealed stock plate with 95 µL of 1× annealing buffer into a new semi-skirted LoBind 96-well plate. Carefully seal the plate with foil and vortex gently for 30 s to mix. Centrifuge the plate at 1,000*g* for 2 min. Stored at −20 °C, stock plates are stable for multiple years.

4.5 μM SPRITE plates (working stocks)

Centrifuge the annealed 4.5 μM DPM, Odd, Even and Terminal stock plates (SPRITE tags described above) at 1,000*g* for 2 min at RT. Carefully remove the foil seal. Generate multiple working stock plates by aliquoting 2.4 μL of each well of the 4.5 μM SPRITE stock plates into its corresponding well of a new semi-skirted LoBind 96-well plate. We recommend making multiple 2.4 μL aliquot 4.5 μM plates of DPM, Odd, Even and Terminal that can be stored for multiple SPRITE experiments. Carefully seal the plate with foil. Centrifuge the plate at 1,000*g* for 2 min. Stored at −20 °C, SPRITE plates are stable for multiple years.

Cell culture of mES cells

mES cell lines were cultured in serum-free 2i/LIF medium and maintained at an exponential growth phase as previously described (ref. [40]). Cells are plated at a density of ~20,000–30,000/cm^2 and passaged every 3–4 d. SPRITE maps were generated in female F1 2-1 ES cells, an F1 hybrid wild-type mouse ES cell line derived from a 129 × *castaneous* cross.

Cell culture of human GM12378 cells

Human GM12878 cells, a female lymphoblastoid cell line obtained from Coriell Cell Repositories, were cultured in RPMI 1640 (GIBCO, Life Technologies), 2 mM L-glutamine, 15% (vol/vol) fetal bovine serum (Seradigm) and 1× penicillin–streptomycin and maintained at 37 °C under 5% CO_2. Cells were seeded every 3–4 d at 200,000 cells/mL in T-25 flasks, maintained at an exponential growth phase, and passaged or harvested before reaching 1 million cells/mL.

Procedure

Cell culture and cross-linking ● Timing 3 h

▲ **CRITICAL** The following steps describe cross-linking cells in suspension to obtain a single-cell suspension of mES cells, which grow in large round colonies. See Supplementary Methods for cross-linking adherent cells directly on the plate.

1 Seed adherent cells on 10–15 cm plates and culture under recommended conditions. We culture mES cells in serum-free 2i/LIF medium and maintain at an exponential growth phase as previously described[40]. This protocol details cross-linking multiple plates of cells pooled in a single suspension. It is important to maintain consistency in lysate batches by ensuring that approximately the same number of cells are pooled, cross-linked and frozen in each experiment. We typically freeze 5–10 million cells per pellet in a 1.7 mL microcentrifuge tube at Step 23.

2 TVP trypsin solution is used to lift mES cells. An hour before starting, warm TVP and post-TVP wash solution at 37 °C.

3 Pre-chill one bottle (≥100 mL) of 1× PBS (without magnesium and without calcium) at 4 °C and keep one bottle (≥100 mL) of 1× PBS at RT. Store 0.5% BSA scraping buffer at 4 °C.

4 Aspirate medium from plates.

5 Wash cells gently with 10 mL of RT 1× PBS.

6 Remove PBS, then add 5 mL TVP to each 15 cm plate and rock gently for 3–5 min at 37 °C until cells begin to detach from the plate

? TROUBLESHOOTING

7 Add 25 mL wash solution to each plate to inactivate the trypsin solution. Gently resuspend cells in the wash solution by pipetting up and down eight to ten times to break up the mES cell colonies and transfer from the plate to a 50 mL centrifuge tube. Rinse the plate with extra wash solution, and add to the 50 mL centrifuge tube. The same cells from different plates can be pooled into the same tube.

8 Centrifuge the cells for 3 min at 330*g* at RT. Wash cells by resuspending in 4 mL RT 1× PBS per 10 million cells and transfer to a 15 mL conical tube. At this step, take an aliquot of the cells to estimate cell count for the SPRITE lysis and sonication steps.

9 Centrifuge the cells for 3 min at 330*g* at RT.

10 Remove and discard the supernatant. Resuspend cells in 2mM DSG cross-linking solution (4 mL per 10 million cells). Rotate gently at RT for 45 min.

▲ **CRITICAL STEP** It is vital that at the beginning of the cross-linking process the pellet is uniformly in suspension. To achieve this, completely resuspend the pellet in 1 mL of DSG cross-linking solution using a P-1000 micropipette. After the pellet is completely resuspended, add the remaining volume of DSG cross-linking solution. If the full volume of DSG cross-linking solution is added

without first resuspending the pellet, it may be challenging to completely break up the pellet, resulting in cell clumps being cross-linked together.

11 Centrifuge cells for 4 min at 1,000*g* at RT. Discard supernatant.

12 Wash cells with 4 mL RT 1× PBS per 10 million cells.

▲ **CRITICAL STEP** As before, each time you resuspend the cell pellet, whether to wash or to put in formaldehyde or scraping buffer, ensure that the pellet is completely resuspended in the solution. Achieve this by first resuspending the pellet in 1 mL of the appropriate solution using a P-1000 micropipette, then adding the remaining volume.

13 Centrifuge cells for 4 min at 1,000*g* at RT. Discard supernatant.

14 Resuspend cell pellet in freshly prepared 3% formaldehyde solution (4 mL per 10 million cells). Rock gently at RT for 10 min.

▲ **CRITICAL STEP** We strongly recommend preparing a new working solution of 3% formaldehyde fresh every time by opening a new ampule to minimize methanol conversion. Do not use an ampule opened more than ~30 min before cross-linking the cells.

15 Add 200 µL 2.5 M glycine stop solution per 1 mL of cell suspension to quench formaldehyde solution.

▲ **CRITICAL STEP** Ensure that the 2.5 M glycine stop solution was made within a month from its use.

16 Rock gently at RT for 5 min.

17 Centrifuge cells at 4 °C for 4 min at 1,000*g*.

18 Discard the formaldehyde supernatant in an appropriate waste container. From here, keep cells at 4 °C.

19 Resuspend the cell pellet in 1 mL of cold (4 °C) 0.5% BSA scraping buffer per 10 million cells and gently rock for 1–2 min.

20 Pellet cells at 4 °C for 4 min at 1,000*g*. Discard the supernatant in a formaldehyde waste container.

21 Resuspend the cell pellet in 1 mL of cold (4 °C) 0.5% BSA scraping buffer again and gently rock for 1–2 min. Pellet as before and discard supernatant.

22 Resuspend the pellet in 1 mL of cold (4 °C) 0.5% BSA scraping buffer per 10 million cells.

23 Aliquot 10 million cells each into 1.7 mL microcentrifuge tubes and pellet at 4 °C for 5 min at 2,000*g*. Remove and discard the supernatant.

■ **PAUSE POINT** Flash-freeze pellets in liquid nitrogen and store pellets at −80 °C for up to 5 years.

Cell lysis ● Timing 1.5 h

24 Chill lysis buffers A, B and C on ice.

25 If using an electronic chiller for the sonication chamber, pre-chill to 4 °C.

26 Thaw a cell pellet (10 million cells frozen per 1.7 mL tube) on ice for 2 min.

▲ **CRITICAL STEP** Fewer cells may be used per SPRITE experiment. As little as 200,000 cells at this step (fragmented with less DNase) have yielded successful SPRITE datasets.

27 Add 700 µL of lysis buffer A supplemented with 1× Proteinase Cocktail Inhibitor (PIC) to each pellet and resuspend fully by pipetting up and down. Be careful not to introduce bubbles while pipetting.

? TROUBLESHOOTING

28 Incubate mixtures on ice for 10 min.

29 Pellet cells at 4 °C for 8 min at 850*g*.

30 Discard the supernatant, taking care not to disturb the pellet.

31 Add 700 µL of lysis buffer B supplemented with 1× PIC to each cell pellet and resuspend fully by pipetting up and down. Be careful not to introduce bubbles while pipetting.

32 Incubate mixtures on ice for 10 min.

33 Pellet cells at 4 °C for 8 min at 850*g*.

34 Discard the supernatant, taking care not to disturb the pellet.

35 Add 550 µL of lysis buffer C supplemented with 1× PIC to each 10 million nuclei pellet and resuspend fully by pipetting up and down. Be careful not to introduce bubbles while pipetting.

36 Incubate the mixture on ice for 8 min. (Optional) QC an aliquot of the nonsonicated lysate by visualizing cross-linked complexes by microscopy (Box 2).

37 Sonicate each sample at 4–5 W for 1 min (ON only): one pulse for 0.7 s ON, 3.3 s OFF. During and after sonication, keep lysate at 4 °C. A Branson needle-tip sonicator kept at 4 °C was used for this protocol.

38 If sonicating more than one pellet of the same cell type and conditions and you do not wish to keep these samples as biological replicates, pool all lysates together and split again into 10 million cell aliquots (based on cell count estimated at cross-linking steps Steps 8 and 23). This ensures that all samples in each tube are equally processed and DNase digested in the subsequent steps.

39 (Optional) QC the sonicated lysate by visualizing cross-linked complexes by microscopy (Box 2).

40 Proceed directly to DNA fragmentation.

▲ **CRITICAL STEP** At times, you may have to go back to the lysate stock to optimize the DNA fragmentation. We recommend storing several 10 µL aliquots of the lysate stock to minimize freeze–thaw cycles. Multiple freeze–thaw cycles of the lysate have shown a decreased quality of coupling observed in SPRITE datasets. Discard lysates after three freeze–thaws.

■ **PAUSE POINT** Flash-freeze lysates in liquid nitrogen and store pellets at −80 °C for up to 5 years.

DNA fragmentation ● Timing 4 h

41 If starting from frozen lysate, thaw one tube of lysate on ice.

42 Fragment the DNA with DNase. To obtain the desired DNA size distribution, perform several DNAse digestion reactions on 10 µL of lysate with varying DNase concentrations, as outlined in the table below.

Stock solution	Volume
10× SPRITE DNase buffer	2 µL
Lysate	10 µL
Turbo DNase from Thermo Fisher	0.8/1/2/3/4 µL
H_2O	7.2/7/6/5/4 µL
Total	20 µL

▲ **CRITICAL STEP** We typically perform SPRITE on DNA that has a size distribution of 50–1,000 base pairs with an average size between 200 and 300 base pairs. In general, a reaction containing 20% Turbo DNase enzyme volume achieves this (Box 2, panel c). For example, for a 20 µL reaction, 4 µL of Turbo DNase would achieve an ideal size distribution. However, a DNase titration is essential for determining the ideal condition before proceeding to SPRITE.

▲ **CRITICAL STEP** Do not use 10× Turbo DNase buffer provided with the enzyme. Instead, use the 10× SPRITE DNase buffer components listed above. Using the incorrect 10x Turbo DNase buffer will result in nicked DNA that cannot be amplified.

? TROUBLESHOOTING

43 Incubate at 37 °C for 20 min.

44 Add 1 µL of 25X DNase stop solution to each sample to terminate the reaction. Mix well by pipetting and immediately place on ice.

45 Reverse the cross-links using half of each sample (10 µL) by adding the solutions listed below. Flash-freeze the remaining half of the DNase-treated sample (10 µL) and store at −80 °C until ready to couple to NHS beads.

Stock solution	Volume
DNase-treated lysate	10 µL
SPRITE ProK buffer	85 µL
ProK	5 µL
Total	100 µL

▲ **CRITICAL STEP** Reversing the cross-links on only half the DNase-treated sample and flash-freezing the remaining half ensures that there is remaining cross-linked sample ready for SPRITE. This negates the need to treat with DNase again, which may result in a different size distribution. The remaining half of the material that is treated with ProK is used to estimate the molarity of DNA molecules in the lysate, which helps to avoid spurious interactions introduced by coupling too many molecules to one bead (see Box 3).

46 Incubate at 65 °C for a minimum of 2 h. Alternatively, you may reverse cross-links at 55 °C for 1 h, then increase the temperature to 65 °C for overnight incubation (>12 h).
? TROUBLESHOOTING

47 Clean up samples by following the protocol provided by the Zymo RNA Clean and Concentrator-5 Kit (>17 nt), binding DNA by adding 2× volumes of RNA binding buffer, mixing, and adding 1× total volume of 100% ethanol. For example, for a 100 µL reaction, add 200 µL of RNA binding buffer and 300 µL of 100% ethanol. Elute in 10 µL of H_2O.

48 Determine concentration of DNA in each sample by following the directions provided with the Qubit dsDNA HS Assay Kit.

49 Determine the size distribution of DNA in each sample by following the directions provided with either the High Sensitivity DNA Kit for the Agilent Bioanalyzer or the D1000 DNA HS ScreenTape for the Agilent 2200 TapeStation (see Extended Data Fig. 1 and Box 3, panel c, for example distributions). The number of unique molecules can be calculated as shown in Box 3, panel c, and Supplementary Table 1.

50 If none of these concentrations of TURBO DNase led to ideal fragmentation, adjust concentrations of DNAse enzyme and repeat the DNase digestion (Steps 41–49) until optimal fragmentation is achieved.
▲CRITICAL STEP Increasing the amount of Turbo DNase to >5 µL (>25% of the reaction volume) may lower the efficiency of the enzyme. Instead, we recommend diluting the lysate 1:1 by combining 5 µL of lysate with 5 µL H_2O to make up a 10 µL total volume. Lysate may be diluted further if DNAse digestion is not sufficient with high concentration of enzyme.

51 (Optional) DNase treat the remaining batch of undigested cross-linked lysate (frozen at Step 40) at the identified optimal DNase concentration.

Stock solution	Volume
10× DNase buffer	110 µL
Remaining undigested lysate (Step 40)	550 µL
Turbo DNase from Thermo Fisher	*X* µL
H_2O	*X* µL to reach final volume
Total	1,100 µL

▲CRITICAL STEP DNase digesting the batch of cross-linked lysate is not necessary for SPRITE if half of the DNase-treated material was saved at Step 45. We only recommend performing this bulk DNase reaction if you expect to perform many SPRITE reactions with the same lysate.

52 (Optional) Incubate at 37 °C for 20 min.

53 (Optional) Add 44 µL of 25 x DNase stop solution to terminate the reaction.
■PAUSE POINT Flash-freeze DNase-treated lysate in liquid nitrogen and store at −80 °C for up to 5 years.

Coupling to NHS-ester beads ● Timing overnight coupling or 3 h

54 Prior to coupling, first calculate the number of molecules required to couple to NHS beads using Supplementary Table 1 and as described in Box 3.

55 Approximately 15–30 min prior to starting coupling, bring the bottle of Pierce NHS-activated beads to RT. After opening, immediately seal the bottle with parafilm and store beads with desiccant to prevent condensation.

56 Vortex the bottle containing the Pierce NHS-activated beads in *N*,*N*-dimethylacetamide until there is a uniform suspension. Being careful not to introduce moisture into the bottle of NHS beads (e.g., leaving bottle open for long periods of time), quickly transfer 2 mL of NHS beads into a clean 1.5 mL LoBind tube. Immediately after transferring the desired amount of beads, store any remaining unused NHS beads wrapped with parafilm in a 50 mL conical tube containing desiccant. Place the tube on a magnetic rack to capture the beads.
▲CRITICAL STEP It is crucial to have an optimal bead to molecule ratio for the split-and-pool barcoding to avoid spurious interactions resulting from multiple cross-linked complexes coupling to the same NHS bead. We aim to bind at a 1:4 to 1:1 ratio of DNA molecules:beads. Box 3 describes how to calculate the number of molecules to couple to NHS beads for a given SPRITE lysate in detail. To determine the microliter amount of lysate to couple, we calculate the DNA

molarity from the concentration and average size measurements obtained in the DNA fragmentation stage of the protocol. Molarity is multiplied by the lysate volume (10 µL of cross-linked, DNase-treated lysate remains in the tube) to obtain DNA molecule number.

57 Remove the *N*,*N*-dimethylacetamide and wash beads by transferring 1 mL ice-cold 1mM HCl to the beads.

58 Remove the HCl and wash beads with 1 mL ice-cold 1× PBS.

59 Remove the PBS and add 500 µL coupling buffer and 10 µL of 0.5 M EDTA to the beads. Vortex until resuspended.

60 After calculating the number of molecules to couple to NHS beads using Supplementary Table 1, dilute *X* µL of the DNase-treated cross-linked lysate (frozen at Step 45) into a total volume of 500 µL coupling buffer.

61 Combine the DNase-treated lysate solution to the NHS-activated beads in coupling buffer solution. Vortex immediately to mix well and ensure uniform coupling of lysate to beads.

62 Incubate the lysate and beads overnight at 4 °C rotating on a HulaMixer sample mixer (or 2 h at RT).

63 Place beads on a magnet and remove 500 µL of flowthrough.

64 Add 500 µL 1 M Tris pH 7.5 (3 M ethanolamine pH 9.0 can also be used) to the beads and incubate on a mixer at 4 °C for at least 45 min. This ensures that all NHS beads will be quenched with Tris and will not bind enzymes in the following steps.

65 Wash beads twice in 1 mL of modified RLT buffer.

66 Wash beads three times with 1 mL of SPRITE wash buffer.

67 Spin the beads down quickly in a microcentrifuge and place back on the magnet to remove any remaining liquid.

Phosphorylation and end repair ● Timing 1 h

68 Blunt the 5′ and 3′ ends of the DNA molecules to prevent unwanted ligation by adding the following mixture to the beads:

Stock solution	Volume
H_2O	212.5 µL
End repair reaction buffer (10X)	25 µL
End repair enzyme mix	12.5 µL
Total	250 µL

69 Incubate on a thermomixer for 15 min at 24 °C, 1,600 rpm.

70 Wash once with 1 mL of modified RLT buffer.

71 Wash three times with 1 mL of SPRITE wash buffer.

72 Spin the beads down quickly in a microcentrifuge and place back on the magnet to remove any remaining liquid.

73 Add dATP to the 3′ ends of each DNA molecule to allow for ligation of the DPM adaptor by adding the following mixture to the beads:

Stock solution	Volume
H_2O	215 µL
dA-tailing reaction buffer (10×)	25 µL
Klenow fragment (exo-)	10 µL
Total	250 µL

74 Incubate on a thermomixer for 15 min at 37 °C, 1,600 rpm. If ligating the DPM adaptor barcode on the same day, set up the reaction described in Step 79 during this incubation.

75 Wash beads once with 1 mL of modified RLT buffer.

76 Wash three times with 1 mL of SPRITE wash buffer.

77 Spin the beads down briefly in a microcentrifuge and place back on the magnet to remove any remaining liquid.

DPM, Odd, Even, Odd and Terminal adaptor ligations ● Timing 1–2 d; 8 h for SPRITE barcoding and 4 h for optional DPM QC

▲ CRITICAL There are 96 DPM, Odd, Even and Terminal adaptors that are designed to ligate onto the DNA molecules. These adaptors are kept in a 96-well stock plate at 4.5 μM. The following steps are designed for optimum efficiency during the process. All ligation steps include SPRITE wash buffer, which contains detergents to prevent beads from aggregating and/or sticking to the plastic tips and tubes, and ensures even distribution of the beads across a 96-well plate. We have verified that these detergents do not substantially inhibit ligation efficiency.

78 Make Ligation Master Mix for five rounds of SPRITE (DPM + four extra tags). Split the master mix evenly into each well of a 0.2 mL PCR 12-tube strip by pipetting 200 μL into each well. Keep on ice.

Stock solution	Volume
NEBNext quick ligation reaction buffer (5×)	1,600 μL
Instant Sticky-End Ligation Master Mix (2×)	1,000 μL
1,2-Propanediol	600 μL
Total	3,200 μL

▲ CRITICAL STEP Homemade Ligation Master Mix can be stored for over 2 weeks at −20 °C.

79 Prepare dilute SPRITE wash buffer (1:1) by mixing 3,000 μL of SPRITE wash buffer with 3,000 μL of H_2O.

80 Accounting for bead volume, add the diluted SPRITE wash buffer to the beads from Step 77 to achieve a final volume of 1.075 mL. Ensure that the beads are evenly resuspended in the buffer. Distribute the beads equally into a 12-well strip tube by aliquoting 89.6 μL of beads into each well.

81 Centrifuge the 4.5 μM DPM adaptor stock plate at 1,000*g* for 2 min before removing the foil seal. Aliquot 2.4 μL from each well of the stock plate of DPM adaptors to the corresponding well in a new LoBind 96-well plate. Ensure that there is no mixing between wells at any point of the process to avoid cross-contamination of barcodes. Use a new pipette tip for each well. After transfer is complete, seal both plates with a new foil seal.

▲ CRITICAL STEP This step can be done in advance, in bulk, so that these plates are ready to use (see 'Reagent setup').

82 Centrifuge the 96-well plate containing the 2.4 μL of aliquoted adaptors and then remove the foil seal.

83 Aliquot 11.2 μL of beads from Step 80 into each well of the 96-well plate that contains 2.4 μL of the DPM adaptors. Ensure that there is no mixing between wells at any point of the process by using a new pipette tip for each well and that there are no beads remaining in the pipette tip.

▲ CRITICAL STEP Pipetting the beads out slowly while touching the pipette tip against the wall of the well ensures that no beads remain in the tip.

84 Carefully add any remaining beads (we usually have leftover beads after diluting beads in Step 80) to individual wells on the plate in 1 μL aliquots to evenly distribute.

85 Aliquot 6.4 μL of Ligation Master Mix from Step 78 into each well; mix gently by pipetting up and down ten times while avoiding bubbles.

The final reaction components and volumes for each well should be as follows:

Stock solution	Volume
Beads + dilute SPRITE wash buffer mix	11.2 μL
DPM adaptor (4.5 uM)	2.4 μL
Ligation Master Mix	6.4 μL
Total	20 μL

▲ CRITICAL STEP Ensure that there is no mixing between wells at any point of the process by using a new pipette tip for each well.

▲ CRITICAL STEP The Ligation Master Mix is viscous and the sample can stick in the walls of the pipette tips. To avoid this, the sample should be dispensed carefully after the reaction is well mixed. Place the pipette tip against the side of the wall of each 96 well and dispense the sample slowly to minimize loss.

86 Seal the plate with a foil seal and incubate in a thermomixer for 60 min at 20 °C, shaking for 30 s at 1,600 rpm every 5 min to prevent beads from settling to the bottom of the plate.
▲ **CRITICAL STEP** Ligation time is critical for achieving high efficiency of ligation in each round of barcoding. We have tested ligation at 5, 15, 30, 45 and 60 min reaction times and found that the 60 min ligation time had the highest yield.

87 After incubation, centrifuge the plate at 1,000*g* for 2 min before removing the foil seal.

88 Pour ~10 mL of modified RLT buffer into a sterile plastic reservoir and transfer 60 µL of modified RLT buffer into each well of the 96-well plate to stop the ligation reactions. It is not necessary to use a new tip for each well.

89 Pool all 96 stopped ligation reactions into a second sterile plastic reservoir.

90 Place a 15 mL conical tube on an appropriately sized magnetic rack and transfer the ligation pool into the conical tube. Capture all beads on the magnet, disposing of all the modified RLT buffer in an appropriate waste receptacle.

91 Remove the 15 mL conical containing the beads from the magnet and resuspend beads in 1 mL SPRITE wash buffer. Transfer the bead solution to a microcentrifuge tube.

92 Wash three times with 1 mL of SPRITE wash buffer.

93 Spin the beads down briefly in a microcentrifuge and place back on the magnet to remove any remaining liquid. Resuspend the beads in SPRITE wash buffer so that the final beads and buffer volume is 1 mL.

94 (Optional) DPM QC (Box 4): vortex well and remove a 5% aliquot of beads by transferring 50 µL of beads solution to another 1.7 mL tube. This step is highly recommended for first-time users to ensure DPM ligation and all prior coupling and enzymatic steps were successful.

95 Place the remaining 95% of beads (or 100% of beads if no aliquot removed for DPM QC) back on the magnetic rack and store beads in 1 mL of modified RLT buffer.
■ **PAUSE POINT** Keep beads at 4 °C overnight during DPM QC or after any round of split-pool barcoding. Do not freeze beads.

96 Remove modified RLT buffer from the remaining 95% aliquot and wash three times with 1 mL of SPRITE wash buffer.

97 Spin the beads down briefly in a microcentrifuge and place back on the magnet to remove any remaining liquid.

98 Repeat Steps 80–93 for the remaining four SPRITE rounds (Odd, Even, Odd, Terminal). After pooling each round, wash beads once in 1mL modified RLT buffer and three times in 1mL SPRITE wash buffer.
■ **PAUSE POINT** The split-pool barcoding can be paused after any round if beads are stored overnight in modified RLT buffer at 4 °C (described in Step 95).

Final library preparation ● Timing ~12–16 h overnight + 3 h

99 Resuspend the beads in SPRITE ProK Buffer so that the combined volume of beads and buffer is 1 mL total.

100 Remove eight 5% aliquots into clean 1.7 mL microcentrifuge tubes and elute the barcoded DNA from the beads by resuspending the beads in the reagents listed below. This serves as an additional round of barcoding so that the total number of unique barcode combinations exceeds the number of DNA molecules in the sample. The remaining lysate on beads can either be stored in modified RLT buffer at 4 °C (for up to 1 week) or also eluted in 5% aliquots.

Stock solution	Volume (per tube)
Beads in SPRITE ProK buffer	50 µL
SPRITE ProK buffer	45 µL
ProK	5 µL
Total	100 µL

▲ **CRITICAL STEP** For first-time users of the protocol, we recommend eluting, reverse cross-linking and PCR amplifying SPRITE aliquots of different sizes (e.g., 0.5%, 1%, 2%, 5%, 10%) to identify a library with 4–10 million unique molecules pre-PCR, and sequencing the library to saturation (e.g., iSeq or MiSeq). We have observed chromosome territories and a broad distribution of SPRITE clusters (Fig. 3c) for libraries that are sequenced to saturation with as few as 4 million reads.

101 Incubate at 55 °C for 1 h, then increase the temperature to 65 °C for overnight incubation (>12 h), shaking at 1,600 rpm.

102 Place the microcentrifuge tubes on a magnet and capture the beads. Transfer the supernatant from each aliquot into a new microcentrifuge tube.

103 To maximize recovery of barcoded material, rinse the beads with 25 µL of water.

104 Vortex, and recapture the beads. Transfer supernatant from each aliquot to its respective tube for a combined volume of 125 µL per tube. Discard the beads.

105 Follow the protocol provided in the Zymo RNA Clean and Concentrator Kit (>17 nt), binding with two volumes of RNA binding buffer and one total volume of 100% ethanol. Elute in 20 µL of H_2O.

106 Amplify the barcoded DNA using the read 1 primer in Box 1 and one of 96 different indexed read 2 primers (Supplementary Table 2) for each separate aliquot. This serves as an additional round of barcoding of each distinct aliquot (as described in Step 100).

Stock solution	Volume
Barcoded DNA (cleaned)	20 µL
Read 1 primer (10 µM)	2.5 µL
Read 2 primer (10 µM)	2.5 µL
2x Q5 Hot Start Master Mix	25 µL
Total	50 µL

107 Amplify using the following PCR program:

Number of cycles	Temperature	Time
1 (initial denaturation)	98 °C	180 s
4	98 °C	10 s
	68 °C	30 s
	72 °C	40 s
8	98 °C	10 s
	70 °C	30 s
	72 °C	40 s
1 (final extension)	72 °C	180 s
Hold	4 °C	∞

▲ **CRITICAL STEP** We have observed that it is best to limit the number of cycles to 16 to avoid nonlinear amplification of DNA. Here we recommend 12 total cycles. The first amplification (4 cycles) remains constant, and the second amplification varies between 8 and 12 cycles depending on aliquot size used.

108 *Clean the PCR reaction and size select for your target libraries.* The total length of the five barcodes on one amplified product is ~175 base pairs, the Illumina sequences added during PCR are ~70 base pairs, and each tagged genomic DNA molecule is no less than 50 base pairs. Agencourt AMPure XP beads are able to size select for your target libraries while cleaning the PCR reaction of unwanted products. First, add 0.7× volume (35 µL) AMPure XP beads to the sample for a total volume of 85 µL and mix thoroughly.

109 Incubate for 10 min at RT.

110 Place the beads on an appropriately sized magnet to capture the beads and the bound DNA. Wait a few minutes until all the beads are captured.

111 Remove the supernatant and discard.

112 Wash beads twice with 200 µL of 80% (vol/vol) ethanol by pipetting ethanol into the tube while beads are captured, moving the tube to the opposite side of the magnet so that beads pass through the ethanol, and then removing the ethanol solution once beads have been recaptured for each wash.

113 Quickly spin down the beads in a microcentrifuge, recapture on magnet, and remove any remaining ethanol.

114 Air-dry beads while the tube is on the magnet for 5–10 min or until the beads appear dry.

115 Elute the amplified DNA by resuspending the beads in 50 µL of H_2O. Transfer the eluted DNA to a clean PCR strip tube.

116 Repeat the size-select clean-up with 0.7× volume (35 µL) AMPure XP beads by repeating Steps 108–114, eluting DNA in 12 µL H_2O for the second elution.

▲ **CRITICAL STEP** To ensure all library material is eluted from beads, elute twice with 6 µL H_2O. Most of the material will be removed in the first elution, and any remaining material will be removed in the second. Pool the two eluates.

117 Determine the concentration of the barcoded libraries by following the directions provided with the Qubit dsDNA HS Assay Kit. Final libraries are generally between 0.3 ng/µL and 1.5 ng/µL.

? TROUBLESHOOTING

118 Determine the size distribution and average size of the barcoded libraries by following the directions provided with the High Sensitivity DNA Kit for the Agilent Bioanalyzer. Average sizes are generally ~350–450 base pairs (Fig. 3a).

119 Input values from Steps 117 and 118 into Supplementary Table 1 to estimate the number of unique DNA molecules present in the sample prior to library amplification. Use this sheet to estimate how many reads are necessary to sequence each SPRITE library to saturation (see Box 5).

120 Pool together SPRITE libraries such that the total sum of reads necessary to sequence each library to saturation is less than or equal to the number of reads available on your sequencing platform. For first-time users of the protocol, we recommend identifying a library with 4–10 million unique molecules pre-PCR, and sequencing the library to saturation (e.g., using iSeq or MiSeq) to confirm a broad distribution of cluster sizes (Fig. 3c) is obtained and chromosome territories are observed. We recommend sequencing approximately 20–25 million reads to visualize compartments, 200–300 million reads for TADs and 1 billion reads for loop interactions.

Deep sequencing ● Timing ~1 d

121 Sequence the pooled SPRITE libraries using an Illumina sequencing kit. For example, we use a TruSeq Rapid SBS v1 Kit – HS (200 cycle) kit on an Illumina HiSeq v2500 platform. For a standard SPRITE experiment (five rounds of barcoding: DPM, Odd, Even, Odd and Terminal ligations), sequence with paired-end reads with 50–90 bp for read 1 (DPM and the genomic DNA sequence), 8 bp index, and 95–125 bp for read 2 (barcodes). See Box 1 for important considerations for read lengths to ensure that all SPRITE tags and the corresponding genomic DNA are sequenced.

Computational pipeline and data analysis ● Timing ~5–8 h

▲ **CRITICAL** The following computational steps are detailed in the SPRITE pipeline instructions in the up-to-date SPRITE wiki: https://github.com/GuttmanLab/sprite-pipeline/wiki. See Box 6 for a detailed step-by-step explanation of the commands used in the SPRITE pipeline.

122 Clone the sprite-pipeline repository from github:

```
git clone https://github.com/GuttmanLab/sprite-pipeline.git
```

123 Download the following required dependencies for successfully running the protocol.

- Conda
- Java 8
- Python 3
- Snakemake
- An installation of Hi-Corrector (https://github.com/jasminezhoulab/Hi-Corrector, included in the scripts subdirectory within the sprite-pipeline directory)

124 Run fastq2json.py[41] to compile the paths of all SPRITE fastq files into a .json file using the following command:

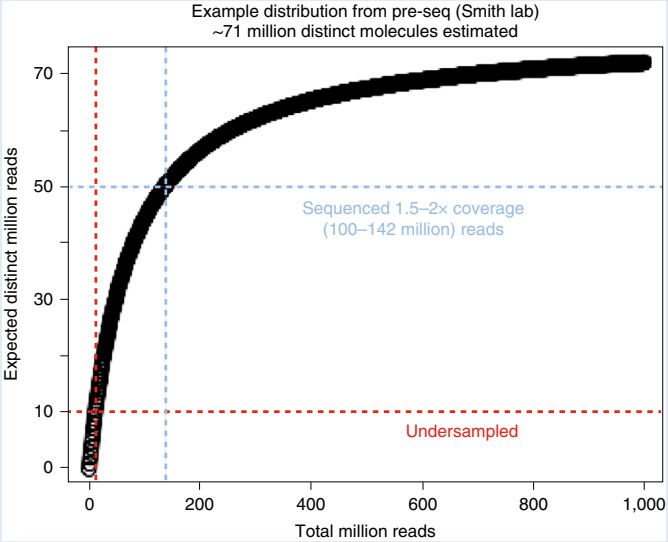
```
python fastq2json.py –fastq_dir /path/to/fastq/directory
```

▲ **CRITICAL STEP** This script will automatically identify all files ending in '_R1.fastq.gz' or '_R2.fastq.gz' within the specified fastq directory.

125 Assign values and paths to each parameter referenced in the config.yaml file (available on the SPRITE GitHub page). Users are expected to generate or download their own Bowtie2 genome index. Example bed files for masking can be found on the SPRITE GitHub. Users can download a premade Bowtie2 mouse genome index at https://ftp.ensembl.org//pub/release-95/fasta/mus_musculus/dna//Mus_musculus.GRCm38.dna_sm.toplevel.fa.gz and human genome index at ftp://ftp.ensembl.org/pub/release-97/gtf/homo_sapiens/Homo_sapiens.GRCh38.97.gtf.gz.

Box 5 | Sequencing a SPRITE library to saturation

Sequencing a SPRITE library is different from sequencing a standard DNA library, where it is standard to sequence a fraction of molecules present in the sample in order to avoid PCR duplicates. In contrast, a SPRITE library must be sequenced to saturation in order to sequence most, if not all, molecules that are interacting in a given cross-linked complex. For example, if we assume Poisson sampling during sequencing, a sequencing depth of 1× (e.g., 71 million reads to sequence 71 million unique molecules) would sample 63% of all molecules. For SPRITE libraries, we sequence with a 1.5–2× coverage, which ensures sampling of >77–86% of unique molecules contained within an individual cluster. Additional sampling of unique molecules would require sequencing hundreds of millions of more reads for modest increases in unique molecule sampling and will result in sequencing mostly PCR duplicates. Pre-seq is a computational tool developed by the Smith laboratory to estimate complexity of a sequencing library[42,44].



To determine the sequencing depth required for each sample, we approximate how many unique molecules are present in the SPRITE library by measuring the number of molecules post-PCR and then account for (i) how many rounds of amplification $2^{(cycles-1)}$ were performed, and (ii) how many SPRI cleanups were performed, each resulting in ~50% loss in molarity.

First, measure the molarity of a SPRITE library to estimate the number of molecules present in the SPRITE library following amplification. We do this in Step 119 by measuring the average size (bp) and concentration (ng/μL) of the SPRITE library. Input the numbers from Steps 117 and 118 into Supplementary Table 1 to calculate the total number of unique molecules present in the sample prior to library amplification. The step-by-step calculations for estimating this value are as follows:

a = library average size (bp)
c = library concentration (ng/μL)
n = number of cycles amplified
p = molecules post-PCR
m = library molarity (nM)
v = volume of library
u = number of unique molecules pre-PCR

1 First calculate the number of molecules present in the SPRITE library post-amplification.
$p = 10^6 \times c/(649 \times a)$
2 Calculate the number of unique molecules present in the SPRITE library.
$u = 2 \times p/(2^{(n-1)})$
$u = 4 \times p/(2^n)$
3 Calculate the sequencing depth required to achieve 2× coverage
Sequencing depth $= 2 \times u$
4 QC: use the Pre-Seq computational tool (Smith Laboratory) to estimate the complexity of the SPRITE library and whether at least 50% saturation has been successfully achieved.

126 Run the SPRITE pipeline using the following command:

```
sh run_pipeline.sh
```

127 Check ligation efficiency by opening the [sample_name].ligation_efficiency.txt in any common text editor. We routinely achieve >75% of reads containing all five barcodes (Fig. 3b).
? TROUBLESHOOTING

128 Check cluster size distribution by opening the cluster_sizes.png file. A successful SPRITE library achieves a wide distribution of cluster sizes (Fig. 3c). Unsuccessful experiments result in libraries that are primarily singlets and are therefore devoid of interactions.
? TROUBLESHOOTING

Box 6 | Computational pipeline

The automated SPRITE computational pipeline allows users to quickly and easily generate contact matrices from raw sequencing reads (Fig. 4a). The end result of the standard pipeline generates a genome-wide heatmap representing chromosome-level interactions. We note that the user will need to input experiment-specific parameters to certain files (e.g., genome for alignment, number of SPRITE tags, order of tags ligated) depending on how the experiment was performed. In the following box, users can see step by step how the SPRITE pipeline progresses from raw fastqs to cluster and heatmap generation.

Trimming

1 Remove Illumina sequencing adaptors from paired-end reads using Trim Galore!.

Barcode identification

2 The next step of the SPRITE pipeline is to identify the barcodes of the sequenced reads. This process is run by a standalone Java program called BarcodeID that accepts two input FASTQ files (paired-end sequencing) and outputs modified versions of these FASTQ files for subsequent alignment. A typical run command looks as follows:

```
java -jar BarcodeIdentification.jar \
--input1 read1.fastq.gz \
--input2 read2.fastq.gz \
--output1 read1.barcoded.fastq.gz \
--output2 read2.barcoded.fastq.gz \
--config configuration_file.txt
```

3 The inputs to this program are the two FASTQ files from a paired-end sequencing run. These files can be gzipped or uncompressed. The outputs are the same FASTQ files with the identified tags appended to the read name. The output files will be gzipped.

Example input:

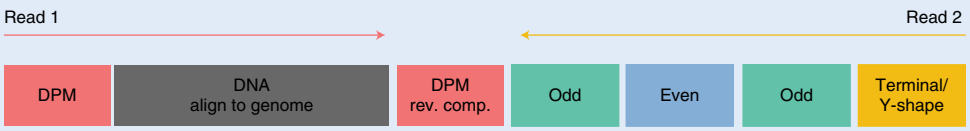
```
@HISEQ:623:HY5KHBCXX:2:2206:7231:7108
ATTGNTAGGTCGGAATTGCACGCTGTAGCGGCATGCTGATGGAGAGGAGAGACTTCTAGCTAGCTACGTGACTGATCCGCACACTGCGACACGTGATCGC
+
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
```

Example output:
Line 1: read identifier with SPRITE barcodes
Line 2: sequence
Line 3: plus sign (+)
Line 4: quality score

```
@HISEQ:623:HY5KHBCXX:2:2206:7231:7108::[DPM6A5][NYBot35_Stg][NOT_FOUND][Even2Bo19][Odd2Bo69]
ATTGNTAGGTCGGAATTGCACGCTGTAGCGGCATGCTGATGGAGAGGAGAGACTTCTAGCTAGCTACGTGACTGATCCGCACACTGCGACACGTGATCGC
+
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
```

4 In the above example, four of five tags were successfully identified (a DPM sequence named 'DPM6A5', a Y-shape sequence named 'NYBot35_Stg', an even sequence named 'Even2Bo19', and an odd sequence named 'Odd2Bo69'). No sequence in position three was found, so a NOT_FOUND tag was inserted. This is the standard behavior of the pipeline when a position cannot be identified.

5 The configuration file specifies the base sequence of every possible tag, as well as the ordering of the tags within a valid barcode. The figure shows an example tag ligation scheme used in our laboratory.



Not shown are sticky-end sequences. The first 30 lines of configuration file used to analyze this layout are below. The complete file can be found at https://github.com/GuttmanLab/sprite-pipeline/blob/master/example_config.txt.

```
# EXAMPLE CONFIG FILE FOR BARCODE ID #
READ1 = DPM
READ2 = Y|SPACER|ODD|SPACER|EVEN|SPACER|ODD
# OPTIONAL #
# Length of spacer ("sticky-end"). Defaults to 6
# SPACER = 6
# The maximum number of bases to look ahead if no tag is initially found.
# Defaults to 6.
# LAXITY = 6
# TAGS #
# Columns:
# CATEGORY NAME SEQUENCE NUM_MISMATCHES
# Valid categories are "EVEN", "ODD", "Y", "RPM", "DPM", "LIGTAG"
```

Box 6 | Computational pipeline (Continued)

```
EVEN Even2Bo1 ATACTGCGGCTGACG 2
EVEN Even2Bo5 CTAGGTGGCGGTCTG 2
EVEN Even2Bo2 GTGACATTAAGGTTG 2
```

The first few lines specify the layout in both read 1 and read 2. If tags are only in read 1 or read 2, then only the corresponding line is kept. Tag categories, such asY (referring to Terminal), ODD, EVEN, etc., are separated by pipes ('|'). The valid categories (DPM, EVEN, ODD, etc.) have no implicit meaning, and are just used to divide tags into categories. Because barcodes are experimentally added in a defined order (which is specified in the config file), we search for these barcodes in the read with this defined order. Since we know where in the read structure to expect each barcode, we search for matches to the known barcode categories at those specific positions. To do this, we include an additional category within the config file called 'SPACER', which represents the sticky ends that act as common sequence spacers between rounds of barcodes. The SPACER variable defines the length of the sticky end. The 'LAXITY' variable defines the maximum number of additional bases (beyond the sticky end itself) to search for the next barcode if it is not found at the exact position that would be expected. This is helpful because the sticky ends can occasionally contain short insertions or deletions that may shift the location of the next barcode. In our design, the SPRITE barcodes are 7 nt, but we provide a SPACER of 6 nt to account for a possible 1 nt indel in the sticky-end sequences and a LAXITY of 6 nt to account for any issues in barcode identification.

The tag section consists of many tab-delimited lines. Each line has four fields and represents a unique tag sequence. The fields, in order, are:

- the tag category (DPM, EVEN, ODD, etc.)
- the tag name
- the tag sequence
- the tag error tolerance

The tag error tolerance field allows for some tags to have base mismatches or miscalls. A value of zero means to only accept the specified sequence as a match; a value of one means to accept any sequence within one Hamming distance of the specified sequence as a match; and so on for two, three, etc.

6 Reads that do not contain a full complement of barcodes are filtered out at this stage using the get_full_barcodes.py script:

```
python get_full_barcodes.py --r1 example.R1.barcoded.fastq.gz
```

The output of this script is two fastq files, one with reads containing the full complement of barcodes (example.R1.barcoded_full.fastq.gz) and the other containing everything else (example.R1.barcoded_short.fastq.gz).

DPM trimming

7 In most cases, the DPM identity is derived from read 1, while the remaining barcodes come from read 2. As a result, the residual DPM sequences that would hinder alignment can only be removed after barcode identification with a second round of trimming using Cutadapt[32]. Cutadapt searches for the common 5′ DPM sequence (GATCGGAAGAG), as well as the 96 barcoded 3′ DPM sequence (dpm96.fasta), accounting for cases where read 1 extends through the genomic DNA sequence into the SPRITE tags ligated on the 3′ end of the fragment.

Alignment to genome

8 After the barcodes have been identified, the reads are aligned to the reference genome. In our standard ligation scheme, only read 1 contains genomic DNA and read 2 contains the SPRITE barcodes, so only read 1 is aligned. Therefore, we do not perform a paired-end alignment, despite having paired-end reads.

9 Any aligner will work. We use bowtie2 with the following parameters:

```
bowtie2 -p 10 -t --phred33 -x <bowtie2_index> -U example.R1.barcoded_full.fastq.gz | samtools view -bq 20 -F 4 -F 256 - > example.DNA.bowtie2.mapq20.bam
```

- -p specifies the number of cpu threads to use, -t prints the wall clock time required to load the index and perform the alignment, --phred33 specifies the read quality encoding, which in this case is the latest one used by Illumina sequences, -x points to the location of bowtie2 indexes and -U specifies the location of the fastq file to be aligned
- We filter reads with a MAPQ score <20, outputting only mapped reads (-F 4) and removing reads that are not the primary alignment (-F 256)

10 If the reference sequences were obtained from Ensembl, we additionally convert chromosomes to the UCSC style (chr1, chr2, etc.) with the following code:

```
python ensembl2ucsc.py -i example.DNA.bowtie2.mapq20.bam -o example.DNA.chr.bam --assembly mm10
```

Repeat masking

11 We use bedtools to discard reads that overlap the annotations in our mask file. The mask file is a composite of two separate mask files:

- UCSC repeatmasker, millidiv <140
- mm10 blacklisted regions v2

The composite mm10 and hg38 mask file that we use can be found in the main directory of the SPRITE-DNA pipeline.
The following command removes reads that align to the masked genomic regions:

```
bedtools intersect -v -a example.DNA.chr.bam -b mask_file.bed > example.DNA.chr.masked.bam
```

Cluster generation

12 To define SPRITE clusters, all reads that have the same barcode sequence are grouped into a single cluster. To remove possible PCR duplicates, all reads with the same genomic position and an identical barcode are removed. We generate a SPRITE cluster file for all subsequent analyses, where each cluster occupies one line of the resulting text file containing the barcode name and genomic alignments.

13 The SPRITE cluster generation script is as follows, where N is the number of tags in a barcode:

```
python get_clusters.py --input example.DNA.chr.masked.bam --output example.clusters --num_tags N
```

Box 6 | Computational pipeline (Continued)

Output file format: the below line represents a single cluster of size three. The first column is the barcode itself ending with the sample name. Each subsequent column contains the type of SPRITE library (DNA), the strand in square brackets, followed by the alignment chromosome, start and end coordinates:

```
DPM6A5.NYBot35_Stg.Odd2Bo71.Even2Bo19.Odd2Bo6.example DNA[+]_chr1:18884355-18884455 DNA[-]
_chr1:18834000-18834100 DNA[+]_chr1:200041887-200041900
```

Heatmap generation

Overview.

14 The final step of the workflow is to transform a file of clusters into a file of contacts. A contacts file is a text file containing a simple square matrix of values representing the contact frequency between any two points on the genome. This matrix is similar to an adjacency matrix and can be easily plotted with MATLAB or R. Examples of DNA interaction matrices can be seen in Fig. 5.

15 To compute the pairwise contact frequency between genomic bins *i* and *j*, we count the number of SPRITE clusters that contain reads overlapping both bins. Specifically, we count the number of unique SPRITE clusters overlapping the two bins and not the number of reads contained within them. In this way, if a SPRITE cluster contains multiple reads that mapped to the same genomic bin, we only count the SPRITE cluster once to eliminate possible PCR duplicates.

16 **Cluster weighting**: because the number of pairwise contacts scales quadratically with the number of reads (n) contained within a SPRITE cluster, larger clusters will contribute a disproportionately large number of the contacts observed between any two bins. To account for this, we reasoned that a minimally connected graph containing n reads would contain $n-1$ contacts. Therefore, we downweight each of the (n choose 2) = ($n(n-1)/2$) pairwise contacts in a SPRITE cluster such that each pairwise contact is weighted by $2/n$. In this way, the total contribution of pairwise contacts from a cluster is proportional to the minimally connected edges in the graph and will have $n-1$ contacts. This also ensures that the number of pairwise contacts contributed by a cluster is linearly proportional to the number of reads within a cluster (Fig. 4b).

17 We recommend making heatmaps using '*n* over 2' weighting to generate heatmaps capturing both long-range (from larger clusters) and short-range (from smaller clusters) interactions.

Step-by-step workflow

18 Divide the genome or chromosome of interest into *N* bins of a given resolution. Finer resolutions will take longer to generate and require much more memory. Initialize a contact matrix with dimensions *N* by *N*.

19 **Raw contacts**: for each cluster in the input clusters file, generate all pairs of reads and record the implied contacts by incrementing the corresponding matrix cells. A value of one per contact will be added to the cells if no downweighting is applied; otherwise, the value will be some value less than one that depends on the downweighting strategy.

20 **ICE normalization**: generate a vector of *N* bias factors corresponding to each row in the matrix. (The bias factor for a given row will be identical to the bias factor for a given column since the matrix is symmetric.) For each cell in the matrix, divide by (bias_factor(row_num) × bias_factor (col_num)). These bias factors are generated with Hi-Corrector[45].

21 **Final heatmaps**: this will transform all values in the matrix to a value between zero and one to make the color scale easier to work with and to compare between samples. Currently, this is done by calculating the median value of all cells one-off from the diagonal and dividing all values by it. Any values that were originally greater than this median value (and so would be greater than one after division) are simply set to one.

22 Run python get_sprite_contacts.py as shown for the following example:

```
python get_sprite_contacts.py --clusters [sample_name_here].clusters --raw_contacts [sample_name_here]
_raw_contacts.txt --biases [sample_name_here]_biases.txt --iced [sample_name_here]_iced.txt --output
[sample_name_here]_final.txt --assembly mm9 --chromosome genome --max_cluster_size 1000000
--min_cluster_size 2 --resolution 1000000 --iterations 100 --downweighting n_over_two --hicorrector
hicorrector/1.2/bin/ic
```

Description of the flags required for heatmap generation:

--clusters: the input clusters file from the previous cluster generation steps Steps 12–13.

--raw_contacts: the downweighted output filename.

--biases: Hi-Corrector outputs a text file of bias factors. Save them to this filename.

--iced: the ICE normalized matrix output filename.

--output: the final, transformed output filename.

--assembly: genome assembly; either 'mm9', 'mm10', 'hg19', 'hg38'. Used to initialize the contact matrix with the appropriate size.

--chromosome: for intrachromosomal heatmaps, one of 'chr1', 'chr2',..., 'chrX'. For interchromosomal heatmaps, 'genome'.

--min_cluster_size and --max_cluster_size: ignore clusters that fall outside these parameters, i.e., clusters that are too big or too small. Default: 2–1,000.

--resolution: the binning resolution in nt. Default: 1,000,000 (1 Mb).

--downweighting: the downweighting strategy, one of 'none', 'n_minus_one', and 'n_over_two'.

none: each contact contributes a value of 1 to the contact matrix. For example, a contact from a five-cluster will contribute 1

n_minus_one: each contact contributes a value of $1/(N-1)$, where *N* is the size of the corresponding cluster. For example, a contact from a five-cluster will contribute 1/4

n_over_two: each contact contributes a value of $2/N$, where *N* is the size of the corresponding cluster. For example, a contact from a five-cluster will contribute 2/5. We recommend n_over_two weighting

--hicorrector: ensure that the path to Hi-Corrector/bin/ic is executable. This is found in the path for Hi-Corrector within the scripts subfolder of the sprite-pipeline downloaded to the user's local workstation.

--iterations: the number of iterations to perform when running Hi-Corrector. Default: 100.

Visualize heatmap in R

23 Run plot_heatmap.R. This takes a 'final.txt' heatmap file and max value to define the cutoff of a heatmap in R.

129 Check the alignment statistics, trimming logs and all other intermediate steps in the pipeline by opening the multiqc file.

130 Open the DNA interaction matrices to visualize the heatmaps. The pipeline outputs a single heatmap, but users can easily modify the script to generate heatmaps at various resolutions. Examples of DNA interaction heatmaps can be seen in Fig. 5.

Troubleshooting

Troubleshooting advice can be found in Table 1.

Table 1 | Troubleshooting table

Step	Problem	Possible reason	Solution
6	5 min have passed and cells have not started to lift from the plate	Forgot to wash with PBS; residual serum will inhibit trypsinization	Quench the trypsin immediately with wash solution to prevent over-trypsinization and premature lysis of cells. Dislodge cells from the plate by washing vigorously with wash solution
27	Pellet is not resuspending in lysis buffer and remains clumpy even after vigorous pipetting	Over-cross-linking	Either thaw a new pellet and repeat lysis resuspending pellet fully by pipetting repeatedly and vortexing lightly or repeat cross-linking steps
	Pellet is not visible due to cell loss during cross-linking	Cells cross-linked in suspension are more prone to loss due to adherence to the plastic surfaces	Adherent cells may be cross-linked on a plate to reduce cell loss, rather than trypsinizing and cross-linking in solution. Additional coating of cell culture dishes is recommended to improve adherence. Alternatively, repeat cross-linking in suspension using low-retention plastics
42	Fragment sizes are far smaller than 200 bp or far greater than 1,000 bp	Short fragments (<200 bp): overdigested with DNase	Repeat DNase digestion step with more titration conditions to achieve optimal size range (50–1,000 bp) and average size (200–300 bp)
		Long fragments (>1,000 bp): underdigested with DNase, reduced activity of enzyme	If underdigestion is occurring, reduce the amount of lysate input into the digestion reaction
46	DNA yield is lower than expected	Incomplete reverse cross-linking	We typically reverse cross-links for 55 °C for 1 h, then increase the temperature to 65 °C for overnight incubation (>12 h)
117	Library yield is lower than expected, despite a successful library obtained during DPM QC	Unsuccessful SPRITE ligations	Make sure the SPRITE barcodes have been ligated in the correct order (e.g., DPM, Odd, Even, Odd, Terminal). Confirm all barcodes have been ligated by PCR using primers that anneal to the DPM tag and terminal tag and ensure that the correct size product is obtained corresponding to all five tags
127	Ligation efficiency is <50%	Residual RLT buffer not completely washed away prior to subsequent round of ligation	Ensure at least three washes are performed to completely buffer exchange into SPRITE wash buffer prior to proceeding with the next round of split-and-pool
		Barcoded tags not annealed	Confirm tags are annealed in an agarose gel by running a few annealed tags alongside the corresponding unannealed DNA tags to confirm a visible size shift upon annealing
128	SPRITE library is devoid of interactions (e.g., >50% reads unpaired)	Not sequencing library to saturation (undersampling reads)	Resequence library at higher sequencing depth (see Box 5)
		Overfragmentation	Redo cell lysis and reduce DNA fragmentation (perform QC by microscopy, see Box 2)
		Pelleting of lysate, resulting in loss of multiway contacts	Avoid high-speed centrifugation after sonication or DNase digestion. (Do not pellet and take supernatant)
	SPRITE library skewed to primarily large SPRITE clusters (>70% reads within 1,000+ reads per cluster)	Too many molecules coupled per bead	Recouple lysate at more dilute molecule-to-bead ratio (see Box 3)
		Insufficient sonication or DNAse digestion	Redo cell lysis and increase sonication and/or DNA fragmentation (perform QC by microscopy, see Box 2)

Timing

Steps 1–23, cell culture and cross-linking: 3 h
Steps 24–40, cell lysis: 1.5 h
Steps 41–53, DNA fragmentation: 4 h

Steps 54–67, coupling: overnight or 3 h
Steps 68–77, phosphorylation and end repair: 1 h
Steps 78–98, DPM, Odd, Even, Odd and Terminal adaptor ligations: ~1–2 d; 8 h for SPRITE barcoding and 4 h for optional DPM QC
Steps 99–120, final library preparation: ~12–16 h overnight + 3 h
Step 121, deep sequencing: ~1 d
Steps 122–130, computational pipeline and data analysis: ~5–8 h
Box 2: QC to visualize cross-linked complexes: 4 h
Box 5: check to determine ligation efficiency of the DPM adaptor: 4 h

Anticipated results

SPRITE libraries typically range in size between 300 bp and 1.3 kb, corresponding to the sum of all ligated SPRITE barcodes (128 bp), the Illumina library amplification primer sequences (124 bp) and the genomic DNA insert (50 bp to 1 kb) (Fig. 3a). We typically achieve >75% total reads tagged with all five barcodes identified, which corresponds to a ligation efficiency of ~95% at each round ($0.95^{5\ \text{rounds}} = 0.75$) (Fig. 3b). To QC the ligation efficiency, a 'ligation efficiency' file is generated during the SPRITE pipeline. A successful SPRITE library will achieve ~65% or more clusters with at least two or more interactions. Furthermore, SPRITE clusters will capture a range of structure sizes: namely, ~30% of reads correspond to clusters with 2–10 reads, and a distribution of interactions correspond to larger cluster sizes containing 11–100 reads, 101–1,000 reads and >1,000 reads per cluster (Fig. 3c).

One critical point for SPRITE is that libraries must be sequenced to saturation to detect as many interactions as possible occurring in each SPRITE cluster (Box 5). For example, if a SPRITE cluster contains 100 molecules and libraries are sequenced with 0.1× coverage, only approximately ten reads within the SPRITE cluster will be sequenced and the remaining ~90 molecules will be missed. Instead, sequencing libraries to >1.0× coverage ensures that the majority of molecules within a SPRITE cluster are sampled. To ensure libraries are sequenced to saturation and that most interactions are detected, we measure the library molarity post-PCR, and provide a calculator to estimate the number of reads to sequence a given SPRITE library at 1.5× and 2.0× coverage (Supplementary Table 1). For first-time users of the protocol, we recommend eluting, reverse cross-linking and PCR amplifying SPRITE aliquots of different sizes (e.g., 0.5%, 1%, 2%, 5%, 10%) to identify a library with 4–10 million unique molecules pre-PCR, and sequencing the library to saturation (e.g., using iSeq or MiSeq). In a successful SPRITE experiment, libraries sequenced to >1.0× coverage will reveal interactions within a range of SPRITE cluster sizes (Fig. 3c) and known DNA structures such as chromosome territories with as few as ~4–10 million reads per sample. After sequencing, we also QC library complexity using the Preseq software[42] to ensure each library was sequenced at enough depth (Box 5).

We have applied SPRITE to several cell lines, including mES cells, human lymphoblasts, human H1 ES cells and human foreskin fibroblast cells. In all of these cases, smaller SPRITE clusters containing 2–10 or 11–100 reads per cluster capture primarily close-range interactions occurring on the same chromosome and corresponding to TADs and A/B compartments (Fig. 3d). Larger-range structures such as chromosome territories and interactions between chromosomes occurring around nuclear bodies are enriched in larger SPRITE clusters containing 101–1,000 and >1,000 reads per cluster (Fig. 3e). Notably, some of the most common interchromosomal contacts are those between gene-poor or transcriptionally inactive regions on chromosomes containing nucleolar organizing regions (chromosomes 12, 15, 16, 18 and 19 in mouse ES cells) that make contacts around the 'inactive nucleolar hub'. Another common set of interchromosomal contacts is between highly active, gene-dense regions on chromatin (as defined by RNA Pol II density by ChIP-seq) that tend to co-localize in the same 3D space that we describe as the 'active speckle hub'.

SPRITE and Hi-C data look quite comparable based on matched sequencing depth and bin resolution. We note that SPRITE has recently been used to specifically map local enhancer–promoter contacts[23], highlighting that this approach works well for mapping close-range contacts. Furthermore, as we described previously[14], we do not observe any SPRITE-specific biases for capturing interactions in open (A compartment regions) or closed (B compartment regions) chromatin.

Data availability

Example DNA SPRITE datasets have been deposited on the 4DN Data Portal[24] under accession numbers 4DNFI8ZROQ87 and 4DNFIY9HL35V.

Code availability

The DNA SPRITE software is available for download on the Guttman Laboratory github page at https://github.com/GuttmanLab/sprite-pipeline/[43]. Version v0.2 is explained in detail within this paper.

References

1. Martin, C. et al. Genome restructuring in mouse embryos during reprogramming and early development. *Dev. Biol.* https://doi.org/10.1016/j.ydbio.2006.01.009 (2006).
2. Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
3. Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 772–772 (2016).
4. Pombo, A. & Dillon, N. Three-dimensional genome architecture: players and mechanisms. *Nat. Rev. Mol. Cell Biol.* **16**, 245–257 (2015).
5. Lieberman-aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–294 (2009).
6. de Laat, W. & Dekker, J. 3C-based technologies to study the shape of the genome. *Methods* **58**, 189–191 (2012).
7. Dekker, J. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
8. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
9. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
10. Dekker, J. Mapping the 3D genome: aiming for consilience. *Nat. Rev. Mol. Cell Biol.* **17**, 741–742 (2016).
11. Lawrence, J. B. & Clemson, C. M. Gene associations: true romance or chance meeting in a nuclear neighborhood? *J. Cell Biol.* **182**, 1035–1038 (2008).
12. Bintu, B. et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, eaau1783 (2018).
13. Takei, Y. et al. Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344–350 (2021).
14. Quinodoz, S. A. et al. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* https://doi.org/10.1016/j.cell.2018.05.024 (2018).
15. Beagrie, R. A. et al. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).
16. Zheng, M. et al. Multiplex chromatin interactions with single-molecule precision. *Nature* **566**, 558–562 (2019).
17. Olivares-Chauvet, P. et al. Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. *Nature* **540**, 296–300 (2016).
18. Tavares-Cadete, F., Norouzi, D., Dekker, B., Liu, Y. & Dekker, J. Multi-contact 3C reveals that the human genome during interphase is largely not entangled. *Nat. Struct. Mol. Biol.* **27**, 1105–1114 (2020).
19. Vermeulen, C. et al. Multi-contact 4C: long-molecule sequencing of complex proximity ligation products to uncover local cooperative and competitive chromatin topologies. *Nat. Protoc.* **15**, 364–397 (2020).
20. Vogel, M. J., Peric-Hupkes, D. & van Steensel, B. Detection of in vivo protein–DNA interactions using DamID in mammalian cells. *Nat. Protoc.* **2**, 1467–1478 (2007).
21. Chen, Y. et al. Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J. Cell Biol.* https://doi.org/10.1083/jcb.201807108 (2018).
22. Yáñez-Cuna, J. O. & van Steensel, B. Genome–nuclear lamina interactions: from cell populations to single cells. *Curr. Opin. Genet. Dev.* **43**, 67–72 (2017).
23. Vangala, P. et al. High-resolution mapping of multiway enhancer–promoter interactions regulating pathogen detection. *Mol. Cell* **80**, 359–373.e8 (2020).
24. Dekker, J. et al. The 4D nucleome project. *Nature* https://doi.org/10.1038/nature23884 (2017).
25. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* https://doi.org/10.1126/science.aal3327 (2017).
26. Dudchenko, O. et al. The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. Preprint at *bioRxiv* https://doi.org/10.1101/254797 (2018).
27. Quinodoz, S. A. et al. RNA promotes the formation of spatial compartments in the nucleus. *Cell* https://doi.org/10.1016/j.cell.2021.10.014 (2021).
28. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* https://doi.org/10.1038/nmeth.4380 (2017).
29. Gaublomme, J. T. et al. Nuclei multiplexing with barcoded antibodies for single-nucleus genomics. *Nat. Commun.* https://doi.org/10.1038/s41467-019-10756-2 (2019).
30. Arrastia, M. V. et al. Single-cell measurement of higher-order 3D genome organization with scSPRITE. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-021-00998-1 (2021).
31. Blecher-Gonen, R. et al. High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein–DNA interactions and epigenomic states. *Nat. Protoc.* **8**, 539–554 (2013).

32. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* https://doi.org/10.14806/ej.17.1.200 (2011).
33. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: identification of problematic regions of the genome. *Sci. Rep*. https://doi.org/10.1038/s41598-019-45839-z (2019).
34. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
35. Zhang, J. et al. ChIA-PET analysis of transcriptional chromatin interactions. *Methods* **58**, 289–299 (2012).
36. Mumbach, M. R. et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, 919–922 (2016).
37. Jäger, R. et al. Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nat. Commun.* **6**, 6178 (2015).
38. 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*. https://doi.org/10.3791/57320 (2018).
39. Fang, R. et al. Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res*. https://doi.org/10.1038/cr.2016.137 (2016).
40. Engreitz, J. M. et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973–1237973 (2013).
41. Terranova, C. et al. An integrated platform for genome-wide mapping of chromatin states using high-throughput chip-sequencing in tumor tissues. *J. Vis. Exp*. https://doi.org/10.3791/56972 (2018).
42. Daley, T. & Smith, A. D. Predicting the molecular complexity of sequencing libraries. *Nat. Methods* https://doi.org/10.1038/nmeth.2375 (2013).
43. Quinodoz, S. A. et al. SPRITE: a genome-wide method to map higher-order 3D interactions in the nucleus using combinatorial split-and-pool barcoding. *GitHub* https://doi.org/10.5281/zenodo.5142570 (2021).
44. Daley, T. & Smith, A. D. Modeling genome coverage in single-cell sequencing. *Bioinformatics* **30**, 3159–3165 (2014).
45. Li, W., Gong, K., Li, Q., Alber, F. & Zhou, X. J. Hi-Corrector: a fast, scalable and memory-efficient package for normalizing large-scale Hi-C data. *Bioinformatics* **31**, 960–962 (2015).

Acknowledgements

We thank all members of the Guttman Laboratory for input on the SPRITE protocol; M. Blanco, I. Goronzy and P. Vangala for helpful comments; B. Tabak for insights on cluster size analysis and weighting; C. Lai and P. Russell for initial support with the SPRITE pipeline; I.-M. Strazhnik and S. Knemeyer for illustrations; and G. Napolitan-Witz for producing and editing the SPRITE video. S.Q. is supported by the HHMI Gilliam Fellowship and NSF GRFP Fellowship. P.B. is supported by the UCLA-Caltech MSTP, NIH F30CA247447, the Josephine de Kármán Fellowship Trust and a Caltech Chen Graduate Innovator Grant. N.O. is supported by the American Cancer Society Postdoctoral Fellowship grant number PF-17-240-01. This work was funded by the NIH 4DN (U01 DA040612 and U01 HL130007), the NYSCF, NIH Director's Early Independence Award (DP5OD012190), CZI Ben Barres Early Career Acceleration Award, Sontag Foundation, Searle Scholars Program, Pew-Steward Scholars program, and funds from Caltech. M.G. is a NYSCF-Robertson Investigator.

Author contributions

S.A.Q. and M.G. conceptualized the experimental method; S.A.Q., P.B., J.J., E.S. and E.D. developed the experimental method and generated data; P.C., N.O. and M.G. developed the computational method and data analysis tools; S.A.Q, P.B., P.C. and M.G. wrote the manuscript.

Competing interests

S.A.Q. and M.G. are inventors of a patent on the SPRITE method.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41596-021-00633-y.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41596-021-00633-y.

Correspondence and requests for materials should be addressed to Mitchell Guttman.

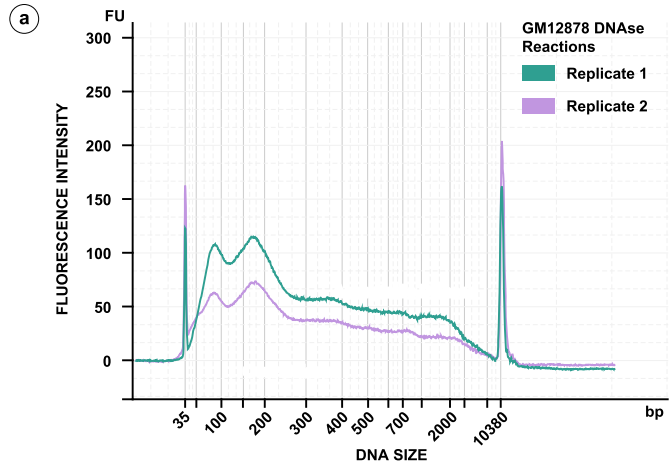
Peer review information *Nature Protocols* thanks Xiaohua Shen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 9 December 2019; Accepted: 13 September 2021;
Published online: 10 January 2022

Related links
Key reference using this protocol
Quinodoz, S. et al. *Cell* **174**, 744–757.e24 (2018): https://doi.org/10.1016/j.cell.2018.05.024



Extended Data Fig. 1 | DNA sizes postfragmentation by DNase for human GM12878 cells. As with mouse ES cells (or any other cell type we have tested), for human GM12878 cells, we optimize DNAse digestion to obtain DNA sized with a range of 50–1,000 base pairs with an average size between 200 and 300 base pairs.