

Ready for production

EDITORIAL SUMMARY:

Micro Capture-C (MCC) is a chromatin conformation capture method for visualising reproducible three-dimensional contacts of regulatory regions in the genome at base pair resolution.

Determining chromatin architecture with Micro Capture-C

Joseph C. Hamley^{1*}, Hangpeng Li^{1*}, Nicholas Denny¹, Damien Downes¹ and James O. J Davies^{1,2,3}

Affiliations

¹MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK.

²Oxford Biomedical Research Centre, Genomic Medicine and Cell and Gene Therapy Themes, Oxford, UK.

³National Institute of Health Research Blood and Transplant Research Unit, Oxford, UK.

*These authors contributed equally to this work

Abstract

Micro Capture-C (MCC) is a chromatin conformation capture (3C) method for visualising reproducible three-dimensional (3D) contacts of specified regions of the genome at base pair resolution. 3C methods are an established family of techniques that use proximity ligation to assay the topology of chromatin. MCC can generate data at substantially higher resolution than previous techniques through multiple refinements of the 3C method. Using a sequence agnostic nuclease, the maintenance of cellular integrity, and full sequencing of the ligation junctions, MCC achieves sub-nucleosomal levels of resolution, which can be used to reveal transcription factor binding sites analogous to DNase I footprinting. Gene dense regions, close range enhancer-promoter contacts, individual enhancers within super-enhancers, and multiple other types of loci or regulatory regions which were previously challenging to assay with conventional 3C techniques, are readily observed using MCC. MCC requires training in common molecular biology techniques and bioinformatics to perform the experiment and analyse the data. The protocol can be expected to be completed in a three week timeframe for experienced molecular biologists.

Introduction

Within the nucleus, chromosomal DNA combines with protein and RNA to form highly complex, dynamic structures that play a key role in determining gene expression^{1,2}. Chromatin structure also plays vital roles in other key cellular processes including the DNA damage response³, DNA replication⁴, VDJ recombination⁵ and cell division⁶ and it can become dysregulated in the context of disease⁷. Assaying the 3D contacts of chromatin is thus an invaluable tool for multiple research questions.

Chromatin architecture is organised on a number of different scales⁸. The chromosomes are largely segregated into chromosomal territories⁹. Within these, chromatin forms large scale structures called topologically associated domains (TADs)¹⁰. TADs tend to colocalize to form 'A' and 'B' compartments¹¹. The A compartments generally contain more active euchromatin and are preferentially localized to the centre of the nucleus in nearly all cell types. Whereas the 'B' compartments colocalize with heterochromatin, which preferentially localizes towards the nuclear periphery in lamina associated domains¹². Within TADs, there are further layers of complexity, with sub-compartments containing genes and enhancers¹³. Inside these compartments enhancers, promoters and CTCF sites make highly specific contacts¹⁴, which are likely mediated in part by transcription factors. Furthermore we are now able to resolve contacts between individual transcription factor binding sites and we have recently identified that complex ordered structures occur on an even smaller scale, within individual enhancers and promoters¹⁵.

Our understanding of DNA structure has been informed by several orthogonal techniques including fluorescence microscopy¹⁶ and genome architecture mapping (GAM)¹⁷ but many of the key discoveries have been made using Chromosome Conformation Capture (3C) based methods, which remain the best way of determining sequence specific contacts at high resolution¹⁸. Although the 3C field was initiated by the seminal paper by Dekker et al¹⁹, the first restriction enzyme based proximity ligation assays were undertaken to detect DNA-DNA interactions in plasmids in the late 1980s²⁰ and subsequently in minichromosomal DNA²¹. However, it was Dekker et al. who described the principle of cutting fixed eukaryotic chromatin with restriction enzymes followed by religation to detect physical proximity.

Over the last two decades improvements in these techniques have resulted in marked increases in throughput, sensitivity and resolution. Initially individual ligation junctions were painstakingly detected using PCR and gel electrophoresis or quantitative PCR (qPCR)²². The adoption of high-throughput sequencing to analyse these libraries dramatically improved the sensitivity and scale of these assays. Although possible in smaller genomes, such as bacteria²³ and fly²⁴, the full human 3C library cannot be sequenced due to the cost of sequencing such large amounts of DNA. Instead, methods have been developed to sample all fragments containing ligation junctions ("all-versus-all" methods e.g. Hi-C²⁵, DNase Hi-C^{26,27} and Micro-C²⁸) or focus on specific regions of

the genome ("many-versus-all" methods e.g. Capture-C²⁹, 4C-seq³⁰) reducing sequencing requirements and markedly improving data quality.

One limitation of these 3C techniques is that their resolution is inadequate below ~500 bp (Fig. 1a,b). Recently we have developed a new method called Micro Capture-C (MCC)¹⁴, which allows DNA structures to be determined down to base pair resolution, thus representing a significant advancement in the field of 3D genome architecture (Fig. 1b,d). Given that the key proteins which drive the formation of chromatin topology such as insulator elements (e.g. CTCF) and transcription factors bind short sequences of DNA (7-20 bp) this technique is providing unique insights into gene regulation because it allows the sites of binding of the proteins involved in mediating contacts to be determined precisely (Fig. 1d).

This protocols paper will discuss the experimental design and protocol of an MCC experiment in detail and place the technique in context of other 3C and non-3C methods to analyze chromatin topology. We will discuss expected results and new experimental opportunities arising from the development of MCC.

Development of Micro Capture-C

We were able to substantially increase the resolution of 3C assays by applying five advances to the Capture-C method^{29,31–33}, which previously provided the greatest sensitivity and resolution 3C data from individual viewpoints. Capture-C itself combines 3C library generation with targeted oligonucleotide capture, which allows very deep sequencing from hundreds to thousands of sites in the genome simultaneously from multiple samples. The improvements to Capture-C include: replacement of strong detergents to maintain cellular integrity; use of micrococcal nuclease (MNase) in place of conventional restriction enzymes; extremely deep targeted sequencing; direct sequencing of ligation junctions and new bioinformatic approaches for precisely locating ligation junctions and visualizing data.

Firstly, one of the limitations of 3C methods is that the resolution is dictated by restriction enzymes which cut in a sequence specific manner. This caps the resolution because data are only generated at the restriction enzyme cut sites; interactions will always be between restriction enzyme cut sites at the end of fragments, rather than between the precise regions making the contact *in vivo*. Initially 3C techniques used 6-cutter enzymes which generate datapoints on average every 5 kb, but this has now been superseded by approaches that use 4-cutter enzymes which generate a mean fragment size of 256 bp. However, restriction enzyme based approaches struggle to generate data below 500 bp because the fragment sizes follow a geometric distribution, which means that the bin size needs to be substantially higher than the mean fragment size to prevent large numbers of bins without data (for a 4-cutter restriction enzyme the bin size has to be >580 bp for over 90% of bins to contain at least one fragment). Recently several groups have increased the resolution using two restriction enzymes³⁴, which increases the potential resolution with over 90% of fragments being < 295 bp in

size. We adopted the use of MNase because this fragments the genome largely independently of DNA sequence and it has a propensity to cut in between nucleosomes³⁵, which are the basic building blocks in the chromatin fiber.

Micro-C was the first 3C method to utilize MNase in place of restriction enzymes, with initial data produced in yeast²⁸ before being applied to human cell types³⁶. This all-versus-all method generates high resolution contact data across the genome. Nuclease based approaches are more technically demanding than restriction enzyme based Hi-C methods and the challenges with obtaining sufficiently deep data with Micro-C mean that for most applications Hi-C generally offers higher resolution. However, the signal from MNase based approaches (Table 1) delineates more precise contacts than restriction enzyme-based approaches. When combined with targeted capture we find that even sub-nucleosomal detail can be resolved when MNase digests are carefully titrated to maintain inter-nucleosomal linkers

Second, we find that the resolution can be substantially improved through minimizing the disruption of nuclear architecture by the avoiding detergents required to make a nuclear preparation from cells. Previous 3C methods have largely used chromatin in solution or purified nuclei, which are usually extracted using detergents such as NP-40 to remove the cytoplasmic membrane³³. This is generally required to enable restriction enzymes to cut chromatin. It is possible to permeabilize cells with digitonin and digest the chromatin with MNase, which substantially improves the signal to noise ratio by avoiding spurious trans ligations of chromatin between cells. Restriction enzymes such as *DpnII* are larger proteins and these do not digest chromatin adequately in cells which have been treated with digitonin.

Third, we generate extremely deep data from individual viewpoints (on average 120,000 up to 500,000 unique contacts per 120 bp viewpoint) equating to over 1,000-fold the depth of data obtained with all-versus-all approaches such as Hi-C and Micro-C. Over 3 trillion ligation junctions would be required for this depth of coverage genome-wide. This is achieved by performing sequencing adaptor ligations in parallel generating highly complex sequencing libraries.

Fourth, we directly sequence the ligation junctions. This allows us to locate precisely two regions of interacting DNA and generate contact maps with base pair accuracy. We reconstruct single reads from paired end sequencing by sonicating MNase 3C libraries to 200 bp fragments and sequencing with 300 bp reads. Most conventional 3C approaches such as Hi-C and Micro-C sequence the libraries with 50 bp paired end of the reads, which is sensible to reduce the cost per read and allow deeper sequencing. However, this means that the position of the ligation junctions is inferred, which limits the resolution for Micro-C.

Finally, we have developed a new analysis pipeline based on using non stringent aligners to identify ligation junctions precisely within reads. In addition, our novel bioinformatic approach allows us to footprint the contacts in a manner analogous to DNase I footprinting³⁷ and reconstruct detailed chromatin interactions.

In addition, by designing adjacently binding oligo probes over a much larger region, MCC can also generate Hi-C (all-versus-all) like data (Tiled-MCC)³⁸. These datasets are higher resolution maps than those that can be achieved by Micro-C and Hi-C. Due to the sub-nucleosomal levels of detail afforded by the MCC methods, it can be used to visualize intra-regulatory region contacts, showing the complex topological landscape within promoters, enhancers, and insulator elements.

Applications of Micro Capture-C

MCC can be used to target any region of the genome so long as a unique oligonucleotide probe can be designed. MCC has already been used to study promoters, enhancers, super-enhancers, and insulator and boundary elements^{14,38}. The increased resolution of MCC allows previously impenetrable loci to be investigated, particularly since many genes are clustered in gene-dense loci in the genome³⁹. MCC can produce interpretable tracks at these loci and start to untangle the underlying biology. For example, in mouse erythroid cells the *Klf1* locus contacts 15 other promoters and enhancers within the TAD (Fig. 1).

Many enhancers are located within 5 kb of a gene promoter⁴⁰. Most 3C techniques are unable to differentiate close contacts but MCC is able to define specific contacts extremely close to the viewpoint, for example it shows a specific contact with an enhancer 2kb upstream of the *Pou5F1* promoter¹⁴. The high signal to noise ratio also allows the absence of 3D contacts to be observed. For example, the silenced embryonic genes at the alpha and beta globin loci are clearly seen not to contact the active enhancers with MCC⁴¹.

The largest impact from MCC is likely to be in interpreting the effects of disease associated variants in the non-coding genome. This is exemplified by the demonstration that the major genetic determinants of susceptibility to COVID-19⁴² is caused by a gain-of-function variants in an enhancer which specifically contacts the promoter of the *LZTFL1* gene⁴³.

Comparison with other methods

All-versus-all

Hi-C couples the 3C principle with massively parallel sequencing, in order to visualize chromatin contacts genome-wide^{25,44}. Hi-C creates contact matrices which sample all interactions from all regions of the genome (all-versus-all). Hi-C has been pivotal in elucidating key principles in the organization of the genome, such as chromatin territories⁴⁵, TADs¹⁰, X-inactivation^{46,47}, and even interchromosomal *trans*-interactions⁴⁸. The resolution has markedly improved with optimization of the technique and increased sequencing depth^{44,49}. For example, in order to circumvent the limitation in resolution inherent with sequence specific restriction endonucleases, MNase (Micro-C²⁸) and DNaseI (DNase Hi-C²⁶) can be used in 3C approaches. However, it is challenging to

sequence the libraries to sufficient depth to generate higher resolution data compared to restriction enzyme based approaches.

The major limitation of this family of techniques remains the cost of sequencing to the depth required for high-resolution maps. At any given region MCC will produce much greater resolution and using tiled MCC one can generate all-versus-all data at Mb-regions. However, if one is interested in global chromatin topology, Hi-C and Micro-C remain the techniques of choice.

Many-versus-all

Many 3C methods improve data quality by targeting specific regions of the genome. This improves the data quality and reduces the cost per sample allowing different cell types and experimental conditions to be analysed. The downside of these approaches is that these protocols are more complex and great care needs to be taken to minimize loss of library complexity, which can result in large numbers of PCR duplicates. This is particularly problematic with protocols with more than one enrichment step and can even result in worse data quality than non-targeted approaches.

Many-versus-all methods either enrich for specific regions of the genome based on sequence, or they combine 3C with chromatin immunoprecipitation to enrich contacts between from specific histone markers or proteins (such as polymerase or CTCF). Techniques such as ChiA-PET⁵⁰, PLAQ-seq⁵¹, HiChIP⁵² offer the possibility of defining genome structure genome wide from all sites of interest. However, the immunoprecipitation step in these techniques leads to loss of library complexity, which limits the depth of data at individual loci. There is also potential for co-enrichment bias to erroneously identify contact between regions that are enriched by the antibody. We therefore, favor methods which enrich specific sites in the genome.

4C was the first such method developed^{30,53}, which circularizes small ligated fragments and uses inverse PCR to generate high resolution interaction profiles. A number of seminal studies have leveraged 4C, including the complex regulation of the Hox locus⁵⁴, linking disease associated SNPs in the FTO locus to the *IRX3* gene⁵⁵, and monitoring enhancer-promoter interactions during Drosophila development⁵⁶.

Targeted oligonucleotide capture approaches such as Capture-C³³, Tiled Capture-C⁵⁷ and Capture Hi-C⁵⁸ have a significant advantage over 4C because they allow very flexible experimental design, allowing anything from a single viewpoint up to tens of thousands of sites to be assayed in a single experiment. In addition, they have intrinsic PCR duplicate filtering analogous to UMI-4C⁵³.

MCC was initially developed from Capture-C, which is a well-established technique with years of optimisation^{29,31–33}. Although it is unable to generate data with the same resolution of MCC, Capture-C is more straightforward to undertake and the profiles are interpretable with lower quality data than MCC. In addition, both Capture-C and Tiled Capture-C can give good quality 3D data from as few as 2,000 cells^{57,59}, making them the technique of choice when working with rare primary cell types or patient samples. Since the Capture-C protocol is very similar to MCC but significantly more forgiving we

would recommend that researchers without extensive 3C experience consider gaining experience with Capture-C prior to undertaking MCC.

Non-3C methods

There are also non-3C methods for the study of chromatin topology. Genome architecture mapping¹⁷ takes ultra-thin slices of nuclei before sequencing in order to assay the position the chromatin in a single cell manner. In large part GAM and Hi-C contact matrices largely agree, although GAM seems to better detect contacts in active euchromatin whereas Hi-C has more detail in inactive heterochromatin, when comparing the contacts directly⁶⁰. Recently, GAM has been applied to specific mouse brain cell types (immunoGAM)⁶¹ and discovered “melting” of long active genes in those cell types, showing that this technique can offer insight not seen by traditional 3C methods. Other non-3C methods (which work at the single cell level) include split-pool recognition of interactions by tag extension (SPRITE)⁶², DNA-seqFISH+⁶³ and DIP-C⁶⁴. Although the resolution of these methods is improving, at present these techniques are largely restricted to the study large scale chromosome formations.

Expertise needed to implement the protocol

The molecular biology techniques utilized in MCC, such as DNA ligation, library preparation and streptavidin bead pull-down are standard practice in many other protocols. Titration of MNase concentrations improves with experience of the technique and requires careful laboratory technique to minimize sources of noise. This protocol requires a large number of steps and therefore can be challenging for those with limited laboratory experience. Data analysis requires basic understanding of Linux command line and we provide an in-depth protocol below.

Limitations

The major limitation of MCC currently is the large number of input cells required to generate useable libraries (currently ~3 million per replicate, with at least 6 replicates being required for footprinting quality data). This will make study of rare primary cell populations or patient samples challenging and conventional Capture-C³³ would be a more appropriate method.

MCC has a small but quantifiable bias towards detecting contacts between regions of open chromatin. Extensive work to quantify this bias through sequencing both the digestion controls (which are non-ligated) and MNase libraries has shown that this bias mainly results from the ligation step rather than incomplete digestion. We have found that there are around 40% more ligation junctions in regions of open chromatin compared to the inactive chromatin background. Importantly, at hypersensitive sites there was no correlation with the degree of DNase I hypersensitivity. This is partially caused by a biological effect. In nucleosome depleted regions there are greater numbers of potential sites for MNase to cut because transcription factors bind shorter

DNA sequences than histones and this results in a shorter fragment size and larger numbers of potential ligation junctions. It is possible to correct for this effect by sequencing the raw 3C library but this probably is not necessary unless it is important to detect very small differences in contact frequencies between a hypersensitive site and one that is not hypersensitive.

Like all 3C techniques MCC is not dynamic, it takes a static snapshot of chromatin conformation. A time-course can be used to increase dimensionality⁶⁵, but 3C methods can be complemented by microscopy studies of chromatin conformation such as live cell imaging¹⁶ to gain a better understanding of spatio-temporal and dynamic elements of chromatin topology.

Experimental design

Samples

The high-throughput and multiplex nature of MCC, through the use of readily available 120 bp biotinylated oligos and captures performed on pooled libraries, makes designing experiments relatively straight forward. Different cell types, conditions, genotypes etc. can be compared and batch effect minimized by the pooling of uniquely indexed libraries into one tube before capture.

Interpretable read pileup data is generated from every replicate and data of this resolution suffices for most applications. The analysis pipeline produces these tracks by default. These data are plotted without windowing but have an effective window size of ~100 bp, compared to plots of base pair resolution, due to the size of the collapsed reads. It is more challenging to generate footprinting quality data and this normally requires merging data from at six least 6 replicates; we usually use a minimum of two technical replicates for each of 3 biological replicates (Fig. 2c).

Viewpoint Selection and Design

One of the major advantages of MCC is that viewpoints are no longer dictated by restriction enzyme cut sites. Therefore, capture oligonucleotides can be precisely positioned at any non-repetitive site in the genome. We have exclusively used 120 mer capture oligonucleotides because these are relatively cost effective to synthesise and they generate good quality data but any size oligonucleotides over 50 bp are likely to work with this protocol.

One major limitation in probe design is that it is difficult to disentangle the contact frequency between two regions which are both targeted in the same capture reaction because this leads to co-targeting bias³². However, it is possible to use tiling approaches for both Capture-C and MCC to visualize cell type specific domains because it is possible to compensate for differences in capture efficiency within the region of interest using ICE normalization⁶⁶.

Library Generation

For an overview of the library generation see Fig. 2. The most challenging stage of MCC is to produce sufficient quantities of ligated 3C DNA libraries for extremely deep

targeted capture. The quantity of MNase has to be precisely controlled relative to the cell number. Differences in the order of 20-30% MNase will alter the digestion profile. This probably requires greater accuracy than it is possible to quantify cell numbers even with modern cell counters. We also find that the digestion profiles vary depending on cell type. However, digestion reactions are usually consistent between different vials of a particular cell type, which have been carefully fixed simultaneously with the same number of cells. This, therefore, allows multiple technical replicates to be performed in order to titrate the amount of MNase required.

A second challenge is to generate highly complex 3C libraries for targeted capture. High quality MCC data sets require 10^5 - 10^6 fold read depth at the capture site particularly to generate footprinting quality data. This has been achieved by performing Next Generation Sequencing (NGS) library preparation on multiple biological and technical replicates. Our general strategy is to use 3 biological replicates each with a technical replicate at a different concentration of MNase and to undertake NGS library preparation for each of these samples in duplicate (i.e. 12 library preparations in total) (Fig. 2c). This provides sufficient material to undertake oligonucleotide capture with at least 10 times the recommended amount of DNA.

Data Analysis

There are large number a 3C analysis tools, a reflection on the ever expanding number of techniques, which subtly differ in the way they process and visualize the data^{20,67-69}, including a recently developed integrated pipeline for all Capture-C based approaches^{33,70}. However, MCC data required development of a novel analysis strategy and set of scripts (Fig. 3). Initially the reads are fully reconstructed by Fast Length Adjustment of Short reads (FLASH)⁷¹ with a central area of overlap (>90% of reads should align if the material has been sonicated adequately). An 800 bp region centered on the oligonucleotide capture is used as a custom reference genome on which to align the reads using the non-stringent aligner BLAT⁷². 800 bp was chosen because the natural decline signal around the capture site means that large numbers of reads do not tend to fall at the edge of the site. However, the region is sufficiently small to allow efficient non-stringent alignment to the region. This enables the junction between the read at the capture site and its interacting partner to be identified. The reads are digested in silico at the identified junction and split into separate fastq files based on which oligonucleotide sequence they best map to. These split reads are then stringently mapped to the genome using Bowtie2⁷³. These data are subsequently analysed by a script that removes PCR duplicates. The precise identification of ligation junctions is somewhat challenging because it requires the relative position of fragments in the initial read to be tracked as well as use of the genome position and the strand to which the read is mapped. However, this allows base pair resolution data to be generated, as well as footprinting to identify the precise location of DNA binding proteins (e.g. transcription factors), which protect the DNA from MNase digestion. The sites of protein binding can be located with even greater precision by subcategorising the ligation junctions

depending on whether the ligation junction is upstream or downstream of the protected DNA sequencing an analogous fashion to ChIP nexus (Fig. 3b).

MATERIALS

Biological materials

Cells

- MCC has been performed on a variety of murine and human primary cells (e.g. erythroid cells (RRID:CVCL_8432), HUVEC) and cell lines (e.g. HUDEP2 (RRID:CVCL_VI06), SEM (RRID:CVCL_0095), THP-1 (RRID:CVCL_0006), Jurkat (RRID:CVCL_0065), HCT-116 (RRID:CVCL_0291)).

CRITICAL Generally at least 10-20 million cells are required to successfully generate an MCC library. This allows for technical replicates of 3-5 million cells with different concentrations of MNase. We generally advise to manufacture 3-5 identical aliquots of 10-20 million cells. This allows for repeatable titration experiments to optimize MNase units and maximises the chances of well digested and ligated libraries. The amount of MNase needs to be carefully controlled against cell number and we believe that fixing identical batches of cells reduces variability compared to using cell counters to generate aliquots of similar numbers of cells at different times. Fixed cells can be stored in -80°C for several years, so if working with a cell type where number is not an issue (e.g. an immortalized cell line) we recommend fixing upwards of 100 million cells.

Caution! Any experiments using human material must conform to relevant institutional and national regulations, and informed consent must be obtained.

Caution! The cell lines used should be regularly checked to ensure that they are authentic, possess a normal karyotype and are not infected with mycoplasma.

Reagents

Common Reagents

- PCR-grade water (Ambion, cat. no. AM9932)
- Absolute ethanol (VWR, cat. no. 20821.330)
- Agencourt AMPure XP SPRI beads (Beckman Coulter, cat. no. A63881)
- M-270 Streptavidin Dynabeads (Invitrogen, cat. no. 65305)
- Qubit dsDNA BR Assay Kit (Invitrogen, cat. no. Q32850)
- D1000 reagents (Agilent, cat. no. 50675583)

Fixation and MNase 3C library generation

- Formaldehyde, 37% (vol/vol) (Sigma, cat. no. 47608-250ML) **Caution!** Formaldehyde is toxic, wear appropriate PPE and avoid contact with skin.
- Glycine (MP, cat. no. 194825)
Digitonin (Sigma D141) **Caution!** Digitonin is toxic, wear appropriate PPE and avoid contact with skin.

- MNase (NEB M0247S)
- Dimethyl sulfoxide (Sigma D8418)
- T4 PNK (NEB M0201L)
- Klenow DNA Pol I Large Fragment (NEB M0210L)
- Ethylene glycol-bis(2-aminoethylether)-N,N,N',N'-tetraacetic acid (EGTA) (VWR 0732-50G)
- Ethylenediaminetetraacetic acid (EDTA) (Fisher Scientific BP2473-100)
- T4 DNA HC ligase (Life Technologies EL0013)
- Qiagen DNeasy DNA extraction kit (Qiagen 69506)
- Proteinase K (Thermo Fisher EO0491)
- CaCl₂ 1M (Sigma 21115)
- NaCl 5M (Thermo Fischer AM9760G)
- Glycerol (Sigma G5516)
- Trypan blue stain 0.4% (Invitrogen T10282)

Library indexing

- NEBNext Ultra II (New England: 7645S/L)
- NEBNext Multiplex Oligos for Illumina Primer set 1 (New England: E7335S/L)
- NEBNext Multiplex Oligos for Illumina Primer set 2 (New England: E7500S/L)
- Herculanase II Fusion Polymerase kit (Agilent: 600677)

Oligonucleotide pull-down:

- Biotinylated probes (e.g., Sigma, IDT xGen Custom Lockdown Probes). For details on oligonucleotide design see Box 1.
- HyperCapture Target Enrichment Kit includes Human Cot-1DNA (Roche, cat. no. 9075828001) **Caution!** Items in this kit contain Formamide and/or Tetramethylammounium cholride. Handle with care and wear appropriate PPE.
- (Optional) Mouse Cot-1 DNA (invitrogen 18440-016) **Critical!** Use Cot-1 DNA specific to the organism. Use COT-1 included in the Hypercap target enrichment kit for human cells. When no species-specific Cot-1 DNA is available, the KAPA Hybrid Enhancer Reagent may be used

Equipment

- Temperature controlled mixer (microcentrifuge ThermoMixer)
- Thermocycler
- Low input DNA analyzer e.g. Agilent 4200 TapeStation (with D1000 Reagents (Agilent, cat. no. 50675582/ G2991AA)) or Agilent 5300 Fragment Analyser [N.B. Agilent Bioanalyser does not work because large fragments from one sample are carried over into the next].
- TapeStation Analysis Software 4.1.1
- Automated cell counter
- Covaris ME220 sonicator, or equivalent

- Covaris microtube AFA Fiber pre-Slit Snap-Cap (Covaris 520045)
- SonoLab Software (Covaris sonication software)
- Speedy-Vac vacuum centrifuge, or equivalent
- Magnetic Separation Rack (DynaMag-2 Thermo Fischer 12321D)
- High-quality DNA LoBind 1.5 mL microcentrifuge tubes and PCR tubes (Eppendorf 0030108051 or Sorenson Bioscience 11500). **Critical!** M270 beads stick to Eppendorf tubes resulting in loss of beads and potentially loss of library complexity. This can be reduced with Sorenson Bioscience 11500 tubes during post capture pull down and washes (Step 55-78, Streptavidin Bead Binding). For all other stages of the protocol low bind DNA tubes such as Eppendorf 0030108051 can be used.

Reagent setup

MNase Storage Buffer

Combine the reagents listed in the table below. Make up in 1 mL aliquots. These can be stored -20°C for 1-2 months.

Reagent	Stock concentration	Volume	Final concentration
Tris-HCl pH 7.5	1 M	10 μL	10 mM
NaCl	5 M	10 μL	50 mM
Glycerol	100 %	500 μL	50 % (vol/vol)
EDTA	500 mM	2 μL	1 mM
Ultra-pure H_2O	-	478 μL	-

MNase

- Stock MNase: Make 10 μL aliquots and store at -80°C (long-term storage).
- 1:100 MNase dilution: Dilute stock MNase aliquots 1:100 with MNase storage buffer (for example add 10 μL aliquot of stock MNase to 900 μL MNase storage buffer to the). This can both be stored at -20°C but it will need to be replaced every 2-3 months or if any concerns about reduced MNase activity.

MNase Reaction Buffer

Combine the reagents listed in the table below. MNase reaction buffer can be stored at room temperature for several months.

Reagent	Stock concentration	Volume	Target concentration (x10)
Tris-HCL pH 8	1 M	500 μ L	500 mM
CaCl ₂	1 M	10 μ L	1 mM
Ultra-pure H ₂ O	-	490 μ L	

1% (w/v) Digitonin

Make 1% (w/v) Digitonin by weighing 2 - 5 mg in an empty microcentrifuge and adding the appropriate amount of DMSO (200 – 500 μ L). Use within 1 week and store at 4 °C.

CRITICAL It is reported that different batches of digitonin can have different activity⁷⁴. Therefore, we test that new batches of digitonin with permeabilized live cells at the concentration used in the protocol (5 μ L 1% digitonin in 1 mL (0.005% overall) using trypan blue to show that digitonin permeabilizes nearly all cells compared to unpermeabilized cells).

EGTA

EGTA requires adjustment of pH in order to go into solution. To make 6 mL of 500 mM solution dissolve 1.14 g of EGTA with 290 mg NaOH in 4.5 ml PCR-grade water. Then adjust the pH to 8 and make volume up to 6 mL then filter sterilize with a 0.2 μ m filter. Store at –20 °C for long term storage (replace every 5-6 months) as 1 mL aliquots.

80% (v/v) ethanol

Add 40 mL of 100% ethanol to 10 mL of PCR-grade water. Store at RT (20–22 °C) (remake every few weeks).

1 M Glycine solution

Add 3.75 g glycine to 50 mL of ultra-pure water. Make up fresh. Store on ice until use.

PROCEDURE

MCC library generation

Formaldehyde Fixation

Timing: 3h

CRITICAL

MCC library digestion and ligation requires precise matching of cell number to concentration of MNase. We therefore recommend making several identical aliquots of cells, if possible, to reduce the variability of cell numbers between aliquots. This allows the concentration of MNase to be adjusted without altering cell number in successive experiments.

1. If required, carefully disaggregate the cells into a single cell suspension in an appropriate cell culture medium. Count the cells with an automated cell counter and adjust the concentration to 1-2 million cells / mL. This can be scaled up accordingly to 50 ml Falcon tubes. Separate into multiple tubes if needed.

CRITICAL STEP

Ideally around 50-100 million cells should be processed for each biological replicate to allow multiple digests to be performed and to obtain enough high-quality libraries to achieve foot printing quality data (Fig. 2c).

2. Based on the number of cells acquired, add 570 μ L 37% (vol/vol) formaldehyde (2% (vol/vol) final concentration) per 10 mL of cell suspension (2280 μ L of 37% formaldehyde for 40 million cells) and incubate for 10 min at RT while tumbling or rotating. In our experience it is easiest to fix up to 40 mL of cell suspension in a 50 mL centrifuge tube.

CAUTION

Formaldehyde is toxic; avoid skin contact and dispose of waste appropriately.

3. Quench formaldehyde by adding 1.5 mL 1 M glycine per 10 mL of cell suspension. Centrifuge 5 min at 300 g (RT unless otherwise stated). Remove supernatant and resuspend in 10 mL of PBS to wash.
4. Centrifuge 5 min at 300 g, remove supernatant without disturbing pellet and resuspend in PBS at a density of 10-20 million cells/mL. Resuspend the cells carefully to make identical 1 mL aliquots each with 10-20 million cells in separate microcentrifuge tubes.

CRITICAL STEP

Carefully ensure cell numbers remain similar between aliquots of fixed cells. Each aliquot of fixed cells (10-20 million) is sufficient for 3 separate digest reactions allowing titration of MNase concentration for optimization. Only 1-2 digested libraries are required per biological replicate (Fig. 2c).

5. Add 5 μ l 1% (v/v) digitonin, mix incubate for 10 minutes at RT.
6. Snap freeze with dry ice/ethanol (or liquid nitrogen).

PAUSE POINT

Store in -80°C freezer for 18+ months.

Digestion

Timing: 2h

7. Defrost one aliquot of fixed cells and centrifuge 5 min at 500 g (RT). Gently remove supernatant with P1000 and then P20 pipette without disturbing pellet and resuspend in 910 μ L PCR-grade water.
8. Count cells with automated cell counter or equivalent. Expect a concentration of between 8×10^6 to 1.6×10^7 cells / mL after losses during fixation.
9. Divide resuspended aliquot of fixed cells into 3 separate MNase digest reactions (300 μ L cells per tube, to which with varying MNase concentrations will be added for titration).
10. Add 80 μ L reduced Ca^{2+} content MNase reaction buffer to each reaction. Then add a variable quantity of PCR grade water and MNase to each reaction as per the table below into the microcentrifuge tubes.

Kunitz units	5	7.5	10	15	20	25	30	40	50
PCR-grade water Vol (μ L)	417.5	416.3	415	412.5	410	407.5	405	400	395
MNase (1in100) Vol (μ L)	2.5	3.75	5	7.5	10	12.5	15	20	25

CRITICAL STEP

The amount of MNase should differ by 20-30% between different digests (for example: 20, 25, 30 Kunitz units). Both under and over digestion of the material leads to poor ligation (Fig. 4a). For an initial experiment on a new cell type, you may want to do 6

digests using half quantities of the cells / reagents to identify the best range of MNase to use.

11. Shake horizontally 550 rpm for exactly 1 h at 37 °C (microcentrifuge Thermomixer or equivalent).

CRITICAL STEP

The cells can become very fragile during digestion and the thermomixer speed may need to be adjusted depending on the cell type and model of thermomixer.

12. Add 8 µL EGTA 500 mM (final concentration 5 mM).
13. Centrifuge 300 g 5 min (RT), gently remove supernatant with P1000 and then P20 pipette without disturbing pellet and resuspend in 1 mL PBS with 10 µL EGTA.

CRITICAL STEP

Carefully remove as much supernatant as possible to eliminate the MNase left with the cells but leave the cell pellet intact.

14. To prepare the digest control, take 200 µL of cell suspension from each sample for the digest control and extract the DNA using the Qiagen DNeasy kit with the following modifications. Add 20 µl Proteinase K mix and then add 200 µL buffer AL. Incubate for 4 h or more at 65 °C. After 4 hours add 200 µl 100% ethanol mix; add to the spin column and centrifuge 6000 g for 1 min. Then add 500 µL buffer AW1 and centrifuge at 6000 g for 1 min. Discard the flow through. Add 500 µL buffer AW2 to the column and centrifuge at 20000 g for 3 min. Finally place the column onto a new microcentrifuge tube and add 20 µL buffer AE/H₂O. Incubate for at least 1 min at RT then centrifuge at 6000 g for 1 min and retain the flow through. Digest controls should be stored at –20 °C until required.

Ligation

Timing: 15h

15. Centrifuge the remaining 800 µL each of the three samples at different MNase concentrations in PBS (from step 14) at 300 g for 5 min and discard supernatant very carefully with a pipette using P1000 and then P20 pipette.
16. Resuspend the cells by adding the reagents the table below:

Reagent	Volume
PCR-grade water	322.2 µL
DNA ligase buffer	40 µL
dNTPs (10 mM each)	16 µL

EGTA (500 mM)	2 μ L
PNK (10 U/ μ L)	8 μ L
Klenow (5 U/ μ L)	8 μ L
T4 DNA ligase (30 U/ μ L)	4 μ L

17. Shake horizontally 550 rpm (shake 15 s and pause 30 s) for 2 h at 37 °C.
18. Shake horizontally 550 rpm (shake 15 s and pause 30 s) for 8 h at 20 °C then program to hold still at 5 °C.
19. Centrifuge 300 g 5 min, discard the supernatant and resuspend in 200 μ L PBS.
20. Use the Qiagen DNeasy kit to extract the DNA with the following modifications. First add 20 μ L Proteinase K, mix and then add 200 μ L of buffer AL. Incubate at 65 °C for 4 h or longer. Allow to cool briefly and add 200 μ L 100% ethanol. Mix and add to a spin column; centrifuge 6000 g 1min. Then wash with 500 μ L buffer AW1; centrifuge 6000 g 1min. Add 500 μ L buffer AW2; centrifuge 20000 g 3 min. Elute in 135 μ L buffer AE or H₂O.
21. Assess 1 μ L of the ligated sample (from Step 20) and 1 μ L of control (from Step 14 per library using D1000 TapeStation (see Fig. 4 for expected results).

TROUBLESHOOTING

22. Quantify the DNA concentration with the Qbit BR reagents.

PAUSE POINT

Both controls and ligation reactions can be frozen and stored for a year or more at -20 °C

Sequencing library preparation

Sonication

Timing: 2h

CRITICAL

The library preparation and oligonucleotide capture process involve multiple steps, which can fail. It is therefore important to minimize the consequences of a failed reaction. We therefore recommend not to use all the material from the MCC ligation reactions for sonication and library preparation.

23. Select 1 or 2 (Fig. 2c) libraries with the most optimal digest and re-ligation profiles as determined in step 21 (Fig. 4) for downstream processing. Aliquot 2-4 μ g of the selected libraries for library preparation (up to 10 μ g can be sonicated if desired).
24. Make up to 130 μ L with PCR-grade water and transfer 130 μ L to a Covaris microtube.

25. Since the starting size of the libraries is significantly smaller than genomic DNA, start by using 300 bp settings for genomic DNA. Sonicate one sample initially and check sonication by removing 1 μ L from the Covaris tube for analysis on the tapestation. This allows the sample to be easily sonicated further if necessary. Once the optimal time of sonication is determined, it can be used to sonicate the other samples.

CRITICAL STEP

Pipette carefully to avoid getting bubbles in the Covaris tube because these will reduce transmission of ultrasonic waves and lead to inadequate sonication. Check for bubbles in the tube and centrifuge if necessary.

CRITICAL STEP

Aim for a peak of 180-200 bp and at least 80% of DNA <300 bp. The precise identification of ligation junctions requires that the read is fully reconstructed from 150 bp paired end sequencing. Reads over ~280 bp will be discarded during analysis.

TROUBLESHOOTING

26. Proceed with sonicating the remaining samples.
27. Bring 234 μ L (1.8 x vol) AMPure XP SPRI beads per sample to RT in DNA Lo bind microcentrifuge tubes (one tube per sample) and set aside.
28. Transfer 130 μ L sonicated sample from the Covaris tube to 234 μ L beads (1.8 x) (from Step 27).
29. Incubate sample with beads at RT for 5 mins. Open tube caps and apply them to magnetic column. Allow the liquid to clear (2-3 mins) and then remove carefully with a pipette.
30. Wash with 500 μ L 80 % (v/v) ethanol. Pipette up and down 10 times. Leave for 30 s and discard.
31. Repeat wash with 600 μ L 80 % (v/v) ethanol as Step 30. Remove as much ethanol as possible. Briefly spin down and then reapply to magnetic column. Use P20 tip to remove the rest of the ethanol if necessary. Leave magnetic beads to dry until matt in appearance.
32. Elute in 100 μ L of PCR-grade water. Incubate at RT for 5 min then reapply to magnetic column. Allow to clear (1-2 min) and then recover in two lots of 50 μ L. (Fig. 2c, parallel NGS libraries) Transfer each to separate PCR tubes.

PAUSE POINT

Samples can be stored at -20°C if required (probably long term).

CRITICAL STEP

50 µL sonicated DNA is required per adaptor ligation reaction. Library complexity is key for MCC high-resolution. To maximize DNA input and minimize DNA losses during library prep, each technical replicate will be parallelized into 2 library preps with separate sequencing indices (overall 1-2 technical replicates with 2 PCR duplicates each per biological replicate) (Fig. 2c)

End Prep and Adaptor Ligation

Timing: 3h

33. Add 7 µL End Prep buffer and 3 µL End Prep enzyme to 50 µL of sonicated DNA (defrost if samples were frozen at Step 32). Mix and incubate in a thermocycler for 30 min at 20 °C, followed by 30 min at 65 °C (lid 75 °C).
34. Add 5-7.5 µL NEBNext Adaptor depending on whether 1 or 2 µg of starting material is used. If adapter dimers are a problem, consider calculating the ratio of the concentrations of sonicated DNA to adaptors and aiming for a ratio of 1: 5-7.
35. Then add 30 µL Ultra II Ligation Master Mix and 1 µL Ligation Enhancer. Mix and incubate in a thermocycler for 30 min at 20 °C.

CRITICAL STEP

Mix the NEBNext Adaptor with the DNA prior to adding the Ligation Master Mix. Do not mix the Adaptors with the ligation reagents prior to the DNA because this can lead to adaptor dimers. Size selection is not necessary to remove adapter dimers because these are removed by the oligonucleotide capture steps.

36. Add 3 µL of USER enzyme (from the NEBNext Multiplex Oligos for Illumina Primer set 1), mix and incubate in a thermocycler for 30 min at 37 °C (lid 47 °C).
37. Perform an SPRI bead cleanup as described at Steps 27-31 with 180 µL (1.8 x vol) of AMPure XP SPRI beads. Elute in 56 µL of PCR-grade water and recover 28 µL into two PCR tubes.

PAUSE POINT

Samples can be stored overnight at –20 °C if required.

PCR addition of indices

Timing: 2h

CRITICAL

As samples are split into 2 PCR tubes, run the PCR reactions separately in case the samples are under amplified or over amplified or the reaction fails. Perform one first and check on the D1000 TapeStation before moving on.

CRITICAL

We recommend using the Agilent Herculase PCR reagents as these allow a higher yield than standard reagents.

38. To one of the two PCR tubes for each sample with 28 μL of adaptor ligated DNA (from Step 37), add indexing reagents with index-specific primers. The same indices can be used for the two aliquots from Step 37 (Fig. 2c).

Reagent	Volume
Adaptor Ligated library	28 μL
NEB Universal primer	5 μL
NEB Index primer	5 μL
Herculase II 5 \times buffer	10 μL
dNTPs	0.5 μL
Herculase II polymerase	1 μL

39. Mix and amplify DNA using the settings below for a total of 6 cycles of amplification.

Step 1	98 $^{\circ}\text{C}$	30 s
Step 2	98 $^{\circ}\text{C}$	10 s
Step 3	65 $^{\circ}\text{C}$	30 s
Step 4	72 $^{\circ}\text{C}$	30 s
Step 5	Go to Step 2	
Step 6	72 $^{\circ}\text{C}$	5 min
Step 7	4 $^{\circ}\text{C}$	Hold

40. Check the DNA profiles using D1000 TapeStation. This can be done on 1 μL of the neat PCR reaction prior to the bead clean up as this allows further PCR cycles if necessary. If the profile looks as expected (Fig. 4) repeat Steps 38-39 to add the indices to the remainder of the material (the second tube for each sample from step 39) by PCR.
41. Combine batched PCR reactions and perform an AMPure XP SPRI bead cleanup as described at Steps 27–31 using 1.8 x the volume of AMPure XP SPRI beads. Elute in 30 μL of PCR-grade water for each reaction.
42. Quantify 1 μL of indexed library using Qubit dsDNA BR assay kit.

CRITICAL STEP

If the yield is inadequate (<1 μg per indexed library), repeat the PCR for a further 3-4 cycles. Repeat Steps 40-41 to clean up the PCR reaction.

PAUSE POINT

Indexed MCC DNA can be stored at $-20\text{ }^{\circ}\text{C}$ for one year or more

TROUBLESHOOTING

Capture Enrichment

CRITICAL

The oligonucleotide capture outlined below is optimized for pools of single stranded biotinylated oligonucleotides and uses the Roche capture reagents. For large scale capture applications it can be more cost effective to use double stranded oligonucleotides, such as those synthesized by Twist biosciences and in this instance we would recommend using the manufacturers protocol and reagents.

Oligonucleotide stock preparation

Timing: 1h

43. If oligonucleotides were ordered individually, reconstitute them at the same concentration (e.g. 1 μ M) and generate pools of oligonucleotides by mixing in an exact 1:1 stoichiometric ratio.

CRITICAL STEP

Oligonucleotides will capture effectively at very low concentrations. Take great care not to contaminate oligonucleotides with one another. In addition, take care not to contaminate the reagents used for the oligonucleotide capture with the biotinylated oligonucleotides.

44. Adjust the concentration of the oligonucleotide pool depending on the number oligonucleotides used as in the table below. However, the oligonucleotides are effective over a large range of concentrations and both lower and higher concentrations are effective^{31,32}.

Number of targets	Total concentration of oligonucleotides
1-10	29 nM
10-100	290 nM
>100	2.9 μ M

PAUSEPOINT

Oligonucleotides can be stored at -20°C for several years.

Hybridisation

Timing: 2 days

45. In a PCR tube, combine 0.5-2 µg from each of uniquely indexed samples 1:1 by mass. The addition of unique sequencing indices added in the previous steps allows samples to be multiplexed in the capture process. This allows all biological replicates to be combined and captured with the same set of probes in one tube, reducing batch effect. We recommend not to exceed 12 µg in total mass of input DNA in a capture. The maximum amount of input DNA is used for single replicates of experiments aiming for footprinting level data quality because these require at least 12 uniquely indexed libraries (with 1 µg of DNA per library (Fig. 2c). We find that going above 5X for the capture reagents makes volume handling difficult and it is possible to capture 12 µg with the reagents scaled up 5 fold. However, multiple samples can be combined in a single capture experiment if lower resolution data are acceptable. It is advisable not to use all the material in a single capture in case one of the steps fails. In addition, this allows for the possibility of capturing individual libraries with several different panels of oligonucleotides. Finally, it is possible to amplify this material using the PCR with primers for the P5 and P7 sequences in the adaptors.

Total mass of multiplexed capture (ug)	Capture 1 reagent scaling
1-2	1X
3-4	2X
5-6	3X
7-8	4X
9-12	5X

46. Add 10 µL COT DNA of the correct species per library (50 µL for 5 X).
47. Vacuum centrifuge the mix at 55 °C with the PCR tube lid open. Check the samples regularly to avoid excessive drying. It is possible to dry the sample in a 1.5 mL microcentrifuge tube if PCR tube adaptors are not available.
48. Add 13.4 µL of Universal Enhancing Oligonucleotides per library (67 µL for 5 X) and mix.
49. Add 28 µL of hybridization buffer per library (140 µL for 5 X) and 12 µL of Hybridization Component H per library (60 µL for 5 X); mix carefully and briefly centrifuge, then incubate at RT for 2 min.
50. Replace all buffers and blocking reagents in the freezer to avoid contamination with hybridization oligonucleotides.

CRITICAL STEP

Keep capture probes for different experiments apart. They are in excess and prone to tiny amounts of contamination. Avoid having Universal Enhancing Oligonucleotides out at the same time as capture oligos to minimize contamination of the blocking oligo set.

51. Add 4.5 μL of diluted capture oligonucleotides per library to a new PCR tube and place in a thermocycler set to hold at 47 °C indefinitely.
52. Denature the reconstituted sample including the Universal Enhancing Oligonucleotides using a second thermocycler to incubate at 95 °C for 10 min then hold at 47 °C indefinitely (lid 105 °C).
53. Add the capture oligonucleotides to the denatured DNA and incubate in a PCR machine at 47°C (lid 57 °C) for 24-36 hours (up to 72 h may increase yield).

Bead washes for capture 1

Timing: 4 hours

54. Make up wash buffers as per table below per library. Preheat the Stringent wash buffer and 100 μL of Wash buffer 1 per library on a thermomixer.

Buffer	Buffer to add for 1x	Water to add for 1x	Temperature
10x Stringent wash buffer	40 μL	360 μL	47 °C
10x Wash buffer I (Hot)	10 μL	90 μL	47 °C
10x Wash buffer I (Cold)	20 μL	180 μL	RT
10x Wash buffer II	20 μL	180 μL	RT
10x Wash buffer III	20 μL	180 μL	RT
2.5× Bead Wash buffer	200 μL	300 μL	RT

55. Aliquot 100 μL of M-270 Streptavidin Dynabeads per library into high-quality, microcentrifuge tubes and allow to equilibrate to RT. Note that the beads bind to some brands of low bind tubes and pipette tips, which may cause loss of material (we use tubes from Sorenson BioScience).

TROUBLESHOOTING

56. Place beads on a magnetic stand; remove liquid once clear.
57. Remove from column and add 200 μL of 1 X Bead Wash buffer per library (1 mL for 5 X) to resuspend beads. Vortex to mix. Reapply to magnet and allow to clear. Remove liquid once clear
58. Repeat Step 57 for a total of two washes.
59. Re-apply to the magnet and remove 2nd wash/supernatant. Remove from magnet and resuspend the beads in 100 μL of 1 X Bead Wash buffer per library (500 μL for

5 X). Apply them to the magnet and leave to clear. Do not discard liquid until ready to add sample.

60. Prepare pipettes to transfer sample from the hybridization reaction so that this can be rapidly added to the beads without the temperature falling. Then remove the wash buffer from the beads and spin briefly to move to the bottom of the microcentrifuge tube.

CRITICAL STEP

Hybridisation and washing Steps 61-70 need to be done QUICKLY to maintain temperature (aim to avoid non-specific annealing of capture oligos). It is sensible to move the thermomixer and minifuge onto the bench where the work is being done.

61. Remove sample from long incubation (at 47 °C) on PCR machine. Briefly spin to remove condensate from lid. Add to the M-270 Streptavidin Dynabeads and place on the thermomixer to keep warm. Remove transiently from the thermomixer and mix thoroughly by vortexing gently and pipetting if necessary to resuspend (try to avoid pipetting too much because the beads often stick to the inside of the pipette tips). Pulse spin briefly to ensure all the liquid is mixed with the beads.
62. Incubate at 47 °C on Thermomixer for 45 min at 600 rpm. Check every 10-15 min that the beads have not settled at the bottom of the tube and vortex or pipette up and down to correct this if necessary.
63. Add 100 µl per library of heated (47 °C) Wash buffer 1 to the beads and bound DNA (500 µl for 5 X) on the thermomixer.
64. Vortex for 10 s to mix and spin briefly.
65. Place in magnetic stand and, when clear, carefully discard supernatant.
66. Remove from magnet and place back on the thermomixer. Then quickly add 200 µL of heated 1× Stringent Wash buffer per library (1000 µL for 5 X) to the beads and bound DNA. Vortex to mix.
67. Incubate at 47 °C for 5 min at 600 rpm.
68. Spin briefly on a minifuge and place in magnetic stand. After a few seconds when clear, carefully discard supernatant (first lot of 1× Stringent Wash buffer).
69. Perform a second stringent wash with 200 µL of heated 1× Stringent Wash buffer per library (1000 µL for 5 X).
70. Vortex to mix. Incubate on a thermomixer for 5 min at 47 °C (600 rpm), then briefly centrifuge to remove any liquid from lid.
71. Place back on magnet and discard supernatant when clear. Remove from magnet and add 200 µL RT 1x wash buffer I per library (1000 µL for 5 X).
72. Mix by vortexing for 10 s, and briefly spin in benchtop microcentrifuge to remove any liquid from lid, then move all liquid to a fresh microcentrifuge tube and incubate at RT for 1 min.

73. Place in magnetic stand and discard all the liquid when clear. Remove from magnet and add 200 μ L of RT 1 \times Wash Buffer II per library (1000 μ L for 5 X).
74. Mix by vortexing for 10 s, and briefly spin in benchtop microcentrifuge to remove any liquid from lid, then move all liquid to a fresh microcentrifuge tube and incubate at RT for 1 min.
75. Place in magnetic stand and discard all the liquid when clear. Remove from magnet and add 200 μ L of RT 1 \times Wash Buffer III per library (1000 μ L for 5 X).
76. Mix by vortexing for 10 s, and briefly spin in benchtop microcentrifuge to remove any liquid from lid, then incubate at RT for 1 min.
77. Place in magnetic stand and discard all the liquid when clear.
78. Remove from magnet and elute in 40 μ l per library (200ul for 5 X) of UPW. Vortex to mix.

PAUSE POINT

Beads suspended in PCR-grade water can be frozen at -20°C for a few days if necessary but it is best to avoid freezing.

79. Make up master mix as per the table below per library. This will allow half of the beads to be amplified. It is safer to amplify the material in two batches in case the PCR reaction fails.

Hybridized beads	20 μ L
KAPA HiFi HotStart ReadyMix	25 μ L
POST-LM_PCR Oligo 1&2	5 μ L
	Total 50 μ L

80. Mix and amplify the captured material using the settings below for a total of 14 cycles of amplification.

Step 1	98 $^{\circ}\text{C}$	45 s
Step 2	98 $^{\circ}\text{C}$	15 s
Step 3	60 $^{\circ}\text{C}$	30 s
Step 4	72 $^{\circ}\text{C}$	30 s
Step 5	Go to Step 2	
Step 6	72 $^{\circ}\text{C}$	60 s
Step 7	4 $^{\circ}\text{C}$	Hold

81. Mix contents of PCR reaction including the streptavidin beads with 1.8 x volume of AMPure SPRI beads and perform a bead clean up as described in Steps 27-32. Elute in 20 μ L of PCR-grade water.

82. Analyse with D1000 TapeStation and repeat the PCR amplification (Step 80) with the second half of the streptavidin beads from capture reaction.

TROUBLESHOOTING

83. Repeat Step 81 with the contents of PCR reaction from second half of the streptavidin beads.

84. Pool the contents of both PCR reactions into a new high quality PCR tube ready for second capture.

Double capture

Timing: 1 day

85. Take 75% of the DNA from the first capture and perform hybridization as described using volumes for a single library (Steps 45-54). Retain 25% of the library in case there are problems with the second capture.

86. Place in PCR machine at 47 °C (lid 57 °C) for 20-24 hours.

87. Perform streptavidin bead binding (Steps 55-78) as described using volumes for a single library.

88. Perform PCR amplification (Steps 79-84) as described using volumes for a single library.

89. Perform an AMPure XP SPRI bead cleanup as described at Steps 27-31 using 180 µL of AMPure XP SPRI beads. Elute in 30 µL of PCR-grade water.

90. Analyse with D1000 TapeStation and quantify DNA for sequencing with BR Qbit

TROUBLESHOOTING

91. Store at –20 °C until ready for sequencing

Sequencing and Analysis

Sequencing

Timing: 2 days

92. Sequence material with 300 bp reads (150 bp paired end) with the Illumina platform. We generally recommend 0.5 - 1 x 10⁶ reads per capture oligo per library.

Data analysis

Timing: 24 hours

93. Generate the oligo sequence and oligo colour file using the MCC_BLATfa.pl script. \t Prepare a bed file of the oligonucleotide coordinates (chr \t start \t stop \t name_of_target). The script requires the path to a fasta file of the reference genome (such as the one used by bowtie2).

```
perl MCC_BLATfa.pl -f file_name.bed -g genome.fa
```

This will output two files:

- A FASTA file of the oligos with the sequence of the surrounding 800bp

>chr:start-stop_name (The coordinates of the 120bp of the capture oligo)
 NNNNNNNNNNNNNNNNNNNNNNNNNNN... (The 800bp of the sequence surrounding the oligo)

- A BED file with the coordinates of the target oligo and the 800bp around the oligo. This file also specifies the colour in RGB format, which will be used in the track hub.

94. Prepare fastq files for analysis by merging replicates. If necessary, first rename with the extensions _R1.fastq and _R2.fastq

95. Use trim_galore to remove adaptor sequences from the reads

```
trim_galore --paired filename_R1.fastq filename_R2.fastq
```

96. Merge overlapping reads using Flash

```
flash filename_R1.fastq filename_R2.fastq
```

97. Convert the fastq file to fasta format. This is most easily done using awk

```
sed -n '1~4s/^@/>/p;2~4p' filename_ext.fastq > filename_ext.fa
```

98. Map the fasta file to the sequences around the oligos (file generated by MCC_BLATfa.pl)

```
blat -minScore=20 -minIdentity=5 -maxIntron=10000 -tileSize=11  
oligo_file.fa filename_ext.fa filename_ext.psl
```

99. Split the reads in the merged fastq file using the sequences from blat using MCC_splitter.pl

```
perl MCC_splitter.pl -f filename_ext.fastq -p filename_ext.psl  
-r reference_for_the_experiment
```

100. Align each fastq file generated to the genome using bowtie2 and sort with samtools

```
bowtie2 -p 4 -X 1000 -x bowtie_genome_path target_filename.fastq  
-S target_filename.sam
```

```
samtools      sort      -n      -o      target_filename_sorted.sam
target_filename.sam
```

101. Analyse the resulting sam files using MCC_analyser.pl

```
perl      MCC_analyser.pl      -f      target_filename_sorted.sam      -pf
public_folder      -pu      public_url      -bf
path_to_bigwig_genome_sizes_file -genome genome -o oligo_file.fa
```

102. If configured correctly this script will output the following:

- Read density of the interaction profile (in BigWig format)
- Footprinting from the interaction profile (in BigWig format)
- Separated upstream and downstream footprinting tracks (in BigWig format)
- A file of all ligation junctions
- It should copy all BigWig files to the public folder and create a track hub in the UCSC genome browser

103. Peak calling of MCC data is best achieved with the neural network based peak caller Lanceotron⁷⁵ (<https://lanceotron.molbiol.ox.ac.uk/>).

Troubleshooting

Troubleshooting advice can be found in Table 2.

TIMING

3C library generation

Day 1-2.

Steps 1-6, Formaldehyde Fixation: 3 h

Steps 7-14, Digestion: 2 h

Steps 15-22, Ligation: 15 h

Sequencing library preparation

Day 3.

Steps 23-32, Sonication: 2 h

Steps 33-37, End Prep and Adaptor Ligation: 3 h

Steps 38-42, PCR Addition of Indices: 2 h

Capture Enrichment

Day 4-8.

Steps 43-44, Oligonucleotide Stock Preparation: 1 h Steps 45-53, Hybridisation: 2 d

Steps 54-78, Streptavidin Bead Binding: 2 h

Steps 79-84, PCR Amplification: 2 h

Steps 85-91, Double Capture: 1 d

Sequencing and Analysis

Day 9-10

Step 92, Sequencing: 2 d

Steps 93-103, Data processing: 24 h

Anticipated Results

Ideally MNase digested libraries should have a characteristic digestion pattern with a high intensity at the mononucleosomal peak and decreasing intensities of di- and trinucleosomal peaks (Fig. 4a). Successfully ligated samples will show a shift to higher molecular weights and the large mononucleosomal peak should reduce substantially relative to the di-nucleosomal peak.

The analysis pipeline requires full reconstruction of reads from 150 bp paired end sequencing, therefore samples require careful sonication because reads over ~290 bp are discarded from the analysis (Fig. 4b). Successfully sonicated samples should have a peak between 190-220 bp with >80% of fragments <300 bp size. Duration of sonication will vary with the size of the ligation product and requires titration.

Tapestation profiles are shown for successful adapter ligation. Some degree of adapter dimers are not a problem since these are removed by the oligonucleotide capture process but very high levels of adapter dimers will reduce library complexity and the quality of data produced. As long as a peak is detectable, low levels of DNA are not problematic after the first capture. Overamplification of libraries results in 'daisy chaining' (due to annealing of adapter sequences) (Fig. 4d). We do not find that this is generally a major problem because the DNA is denatured prior to sequencing but it can be corrected by amplifying 50% of the sample with 2 PCR cycles with fresh reagents using the P5 and P7 primers.

A successful MCC experiment should have 10-30,000 unique ligation junctions per probe per replicate. The capture efficiency should be in excess of 80% and the cis – trans ratio should be over 0.7. For interpretable footprinting (base-pair) resolution, data from multiple replicates needs to be combined so that there are > 100,000 unique ligations per probe although it may be possible to footprint close contacts with significantly less data.

MCC data experiments often report high numbers of PCR duplicates. This is because the MCC analysis pipeline has very stringent duplicate filtering based on the end positions of ligated fragments. Reads which do not contain ligation junctions are removed by the PCR duplicate filter because they have a more limited repertoire of end positions. In contrast to Hi-C, Micro C, and capture based Hi-C methods, MCC does not enrich for successful ligation junctions and unlike Capture-C the position of cut sites is not predictable. Therefore, the efficiency of data extraction is lower for comparable sequencing. However, for designs of up to a thousand viewpoints the cost of the sequencing is small compared to the cost of the reagents and other methods are unable to achieve equivalent library complexity for individual viewpoints.

With tiled designs where Mb sized regions of interests are being captured the library complexity falls with the current oligonucleotide capture protocol down from ~1000 ligation junctions per bp captured to ~10 junctions per bp and the sequencing requirements increase. Above this scale of design it becomes more sensible to enrich for ligation junctions with a Micro-C based approach.

Acknowledgements

J.O.J.D. was supported by grants from Wellcome (098931/Z/12/Z and 225220/Z/22/Z); the MRC (MR/R008108/1 and MC_UU_00029/04); the NIHR Blood and Transplant Research Unit and the Oxford Biomedical Research Centre Genomic Medicine and Cell and Gene Therapy Themes. J.C.H. is supported by an MRC studentship.

Author Contributions

J.O.J.D. developed the method and wrote the first draft of the manuscript. J.C.H., H.L. and N.D. optimized the method and wrote the manuscript. D.J.D. optimized the capture methodology and contributed to the manuscript.

Conflicts of interest

J.O.J.D. is a co-founder of Nucleome Therapeutics and he provides consultancy to the company. He also holds a patent for the MCC method, which is licensed to the company.

Data availability

Sequencing data have been submitted to the NCBI Gene Expression Omnibus ([GSE144336](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144336)).

Code availability

The code required for analysis of MCC data are available for academic use through the Oxford University Innovation software store (<https://process.innovation.ox.ac.uk/software/p/16529a/micro-capture-c-academic/1>). A pipeline to run this code is available on GitHub (<https://github.com/jojdavies/Micro-Capture-C>). Instructions for setting up and running the pipeline are available on github.

References

1. Cavalli, G. & Misteli, T. Functional implications of genome topology. *Nat. Struct. Mol. Biol.* **20**, 290–299 (2013).
2. Lam, M. T. Y., Li, W., Rosenfeld, M. G. & Glass, C. K. Enhancer RNAs and regulated transcriptional programs. (2014) doi:10.1016/j.tibs.2014.02.007.
3. Arnould, C. *et al.* Loop extrusion as a mechanism for formation of DNA damage repair foci. *Nat. 2021 5907847* **590**, 660–665 (2021).
4. Marchal, C., Sima, J. & Gilbert, D. M. Control of DNA replication timing in the 3D genome. *Nat. Rev. Mol. Cell Biol.* **20**, 721–737 (2019).
5. Jain, S., Ba, Z., Zhang, Y., Dai, H. Q. & Alt, F. W. CTCF-Binding Elements Mediate Accessibility of RAG Substrates During Chromatin Scanning. *Cell* **174**, 102–116.e14 (2018).
6. Hirota, T., Gerlich, D., Koch, B., Ellenberg, J. & Peters, J. M. Distinct functions of condensin I and II in mitotic chromosome assembly. *J. Cell Sci.* **117**, 6435–6445 (2004).
7. Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–1251 (2013).
8. Krumm, A. & Duan, Z. Understanding the 3D genome: Emerging impacts on human disease. *Semin. Cell Dev. Biol.* **90**, 62–77 (2019).
9. Cremer, T. & Cremer, M. Chromosome Territories. *Cold Spring Harb. Perspect. Biol.* **2**, (2010).
10. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nat. 2012 4857398* **485**, 376–380 (2012).
11. Fortin, J. P. & Hansen, K. D. Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. *Genome Biol.* **16**, 1–23 (2015).
12. Falk, M. *et al.* Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **570**, 395 (2019).
13. S, S. & P, F. Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).
14. Hua, P. *et al.* Defining genome architecture at base-pair resolution. *Nature* <https://doi.org/10.1038/s41586-021-03639-4>. (2021).
15. Aljahani, A. *et al.* Analysis of sub-kilobase chromatin topology reveals nano-scale regulatory interactions with variable dependence on cohesin and CTCF. *Nat. Commun.* **13**, 1–13 (2022).
16. Lakadamyali, M. & Cosma, M. P. Visualizing the genome in high resolution challenges our textbook understanding. *Nat. Methods* **17**, (2020).
17. Beagrie, R. A. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, (2017).
18. Davies, J. O. J., Oudelaar, A. M., Higgs, D. R. & Hughes, J. R. How best to identify chromosomal interactions: a comparison of approaches. *Nat. Methods* **14**, (2017).
19. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Sci.* **295**, (2002).
20. Cairns, J. *et al.* CHiCAGO: robust detection of DNA looping interactions in

- Capture Hi-C data. *Genome Biol.* **17**, 127. (2016).
21. Cullen, K. E., Kladde, M. P. & Seyfred, M. A. Interaction between transcription regulatory regions of prolactin chromatin. *Science* **261**, 203–206 (1993).
 22. Hagege, H. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat. Protoc.* **2**, (2007).
 23. Marbouty, M. *et al.* Condensin- and Replication-Mediated Bacterial Chromosome Folding and Origin Condensation Revealed by Hi-C and Super-resolution Imaging. *Mol. Cell* **59**, 588–602 (2015).
 24. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458–472 (2012).
 25. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
 26. Ma, W. Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. *Nat. Methods* **12**, (2014).
 27. Ramani, V. *et al.* Mapping 3D genome architecture through in situ DNase Hi-C. *Nat. Protoc.* **2016 1111 11**, 2104–2121 (2016).
 28. Hsieh, T. H. S. Mapping nucleosome resolution chromosome folding in yeast by Micro-C. *Cell* **162**, (2015).
 29. Hughes, J. R. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.* **46**, (2014).
 30. Zhao, Z. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* **38**, (2006).
 31. Davies, J. O. J. Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat. Methods* **13**, (2016).
 32. Downes, D. J. High-resolution targeted 3C interrogation of cis-regulatory element organization at genome-wide scale. *Nat. Commun.* **12**, (2021).
 33. Downes, D. J. *et al.* Capture-C: a modular and flexible approach for high-resolution chromosome conformation capture. *Nat. Protoc.* **2022 172 17**, 445–475 (2022).
 34. Akgol Oksuz, B. *et al.* Systematic evaluation of chromosome conformation capture assays. *Nat. Methods* **2021 189 18**, 1046–1055 (2021).
 35. Mieczkowski, J. *et al.* MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. *Nat. Commun.* **2016 71 7**, 1–11 (2016).
 36. Krietenstein, N. *et al.* Ultrastructural Details of Mammalian Chromosome Architecture. *Mol. Cell* **78**, 554-565.e7 (2020).
 37. Galas, D. J. & Schmitz, A. DNase footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.* **5**, 3157 (1978).
 38. Aljahani, A. *et al.* Analysis of sub-kilobase chromatin topology reveals nano-scale regulatory interactions with variable dependence on cohesin and CTCF. *bioRxiv* **2021.08.10.455796** (2021) doi:10.1101/2021.08.10.455796.
 39. Xie, T. *et al.* Analysis of the Gene-Dense Major Histocompatibility Complex Class III Region and Its Comparison to Mouse. *Genome Res.* **13**, 2621 (2003).
 40. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).
 41. King, A. J. *et al.* Reactivation of a developmentally silenced embryonic globin

- gene. *Nat. Commun.* <https://doi.org/10.1038/s41467-021-24402-3> . (2021).
42. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N. Engl. J. Med.* **383**, 1522–1534 (2020).
 43. Downes, D. J. *et al.* Identification of LZTFL1 as a candidate effector gene at a COVID-19 risk locus. *Nat. Genet.* **2021 5311 53**, 1606–1615 (2021).
 44. Akgol Oksuz, B. *et al.* Systematic evaluation of chromosome conformation capture assays. *Nat. Methods* **2021 189 18**, 1046–1055 (2021).
 45. Schmitt, A. D. *et al.* A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep.* **17**, 2042–2059 (2016).
 46. Giorgetti, L. *et al.* Structural organization of the inactive X chromosome in the mouse. *Nature* **535**, 575–579 (2016).
 47. van Bemmelen, J. G. *et al.* The bipartite TAD organization of the X-inactivation center ensures opposing developmental regulation of Tsix and Xist. *Nat. Genet.* **51**, 1024–1034. (2019).
 48. Monahan, K., Horta, A. & Lomvardas, S. Lhx2/Ldb1-mediated trans interactions regulate olfactory receptor choice. *Nature* **565**, 448 (2019).
 49. Rao, S. S. P. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, (2014).
 50. Li, G. Chromatin interaction analysis with paired-end Tag (ChIA-PET) sequencing technology and application. *BMC Genomics* **15**, (2014).
 51. Fang, R. *et al.* Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res.* **26**, 1345–1348. (2016).
 52. Mumbach, M. R. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, (2016).
 53. Schwartzman, O. UMI-4C for quantitative and targeted chromosomal contact profiling. *Nat. Methods* **13**, (2016).
 54. Andrey, G. *et al.* A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* **340**, (2013).
 55. Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371 (2014).
 56. Ghavi-Helm, Y. *et al.* Enhancer loops appear stable during development and are associated with paused polymerase. *Nat.* **2014 5127512 512**, 96–100 (2014).
 57. Oudelaar, A. M. *et al.* Dynamics of the 4D genome during in vivo lineage specification and differentiation. *Nat. Commun.* **11**, 2722. (2020).
 58. Mifsud, B. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, (2015).
 59. Oudelaar, A. M. . D. D. . D. J. & H. J. Low-input capture-C: a chromosome conformation capture assay to analyze chromatin architecture in small numbers of cells. *Bio Protoc.* **7**, e2645. (2017).
 60. Beagrie, R. A. *et al.* Multiplex-GAM: genome-wide identification of chromatin contacts yields insights not captured by Hi-C. Preprint at bioRxiv <https://doi.org/10.1101/2020.07.31.230284>. (2020).
 61. Winick-Ng, W. *et al.* Cell-type specialization is encoded by specific chromatin topologies. *Nat.* **2021 5997886 599**, 684–691 (2021).
 62. Quinodoz, S. A. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* **174**, (2018).

63. Takei, Y. et al. Integrated spatial genomics reveals global architecture of single nuclei. *Nature* 590, 344–350. (2021).
64. Tan, L., Xing, D., Chang, C. H., Li, H. & Xie, X. S. Three-dimensional genome structures of single diploid human cells. *Science* (80-.). **361**, (2018).
65. Owens, D. D. G. et al. Dynamic Runx1 chromatin boundaries affect gene expression in hematopoietic development. *Nat. Commun.* 2022 131 **13**, 1–15 (2022).
66. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* 2012 910 **9**, 999–1003 (2012).
67. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome* <https://doi.org/10.1186/s13059-015-0831-x>. (2015).
68. Buckle, A. . G. N. . M. D. & B. C. A. capC-MAP: software for analysis of Capture-C data. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/btz480>. (2019).
69. Thongjuea, S. . S. R. . G. F. G. . S. E. & L. B. R3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. *Nucleic Acids Res.* 41,.
70. Smith, A. L. . R.-A. K. & S. D. CapCruncher. *Zenodo* <https://doi.org/10.5281/zenodo.5113088> . (2021).
71. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
72. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* 12, 656–664. (2002).
73. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 2012 94 **9**, 357–359 (2012).
74. Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* **13**, (2018).
75. Hentges, L. D., Sergeant, M. J., Downes, D. J., Hughes, J. R. & Taylor, S. LanceOtron: a deep learning peak caller for ATAC-seq, ChIP-seq, and DNase-seq. *bioRxiv* 2021.01.25.428108 (2021) doi:10.1101/2021.01.25.428108.
76. Oudelaar, A.M., Davies, J.O.J., Hanssen, L.L.P. et al. Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nat Genet* 50, 1744–1751 (2018). <https://doi.org/10.1038/s41588-018-0253-2>
77. Schoenfelder, S. et al. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* **25**, (2015).
78. Cheng, L. et al. Single-nucleotide-level mapping of DNA regulatory elements that control fetal hemoglobin expression. *Nat. Genet.* **53**, (2021).

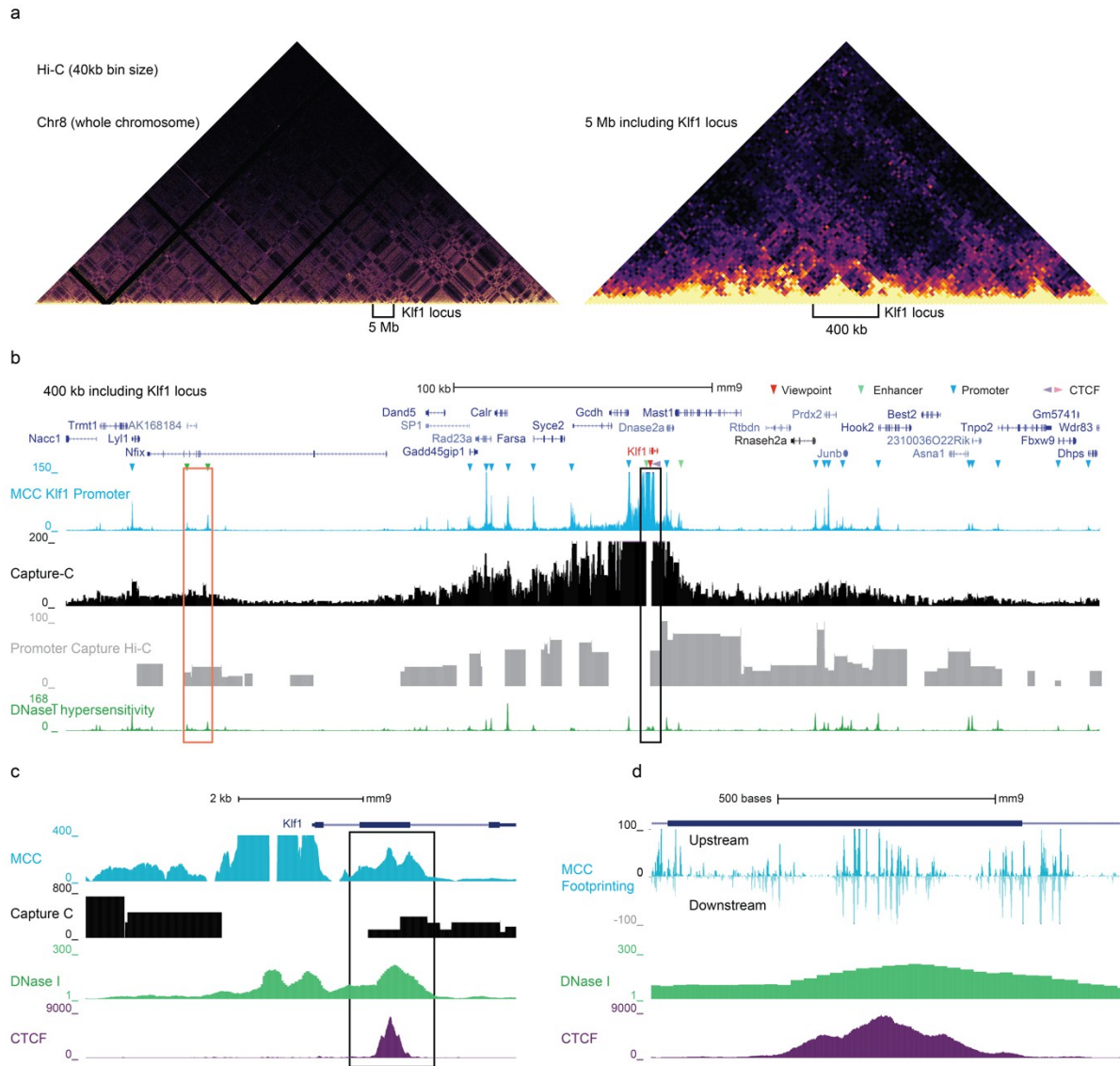


Figure 1 | Comparison of MCC with other techniques at the *Klf1* locus a, Hi-C at 40 kb resolution in murine erythroid cells. The whole of chromosome 8 is shown on the left, and 5 Mb region, which encompasses the *Klf1* locus on the right⁷⁶. b, 400kb region encompassing the gene dense *Klf1* locus showing a comparison of MCC with Capture-C³¹ and Promoter Capture Hi-C⁷⁷ profiles from the promoter of *Klf1*³¹. Interestingly, genome editing of the enhancers in the red box has been shown to alter *Klf1* expression⁷⁸. c, shows the region highlighted with the black box in (b), demonstrating contacts with a potential regulatory element very close to the gene promoter (< 2kb), which are not easily visualised with other techniques. d, boxed region in (c) showing base pair resolution data of ligation junctions separated by whether the junction is upstream or downstream of the read.

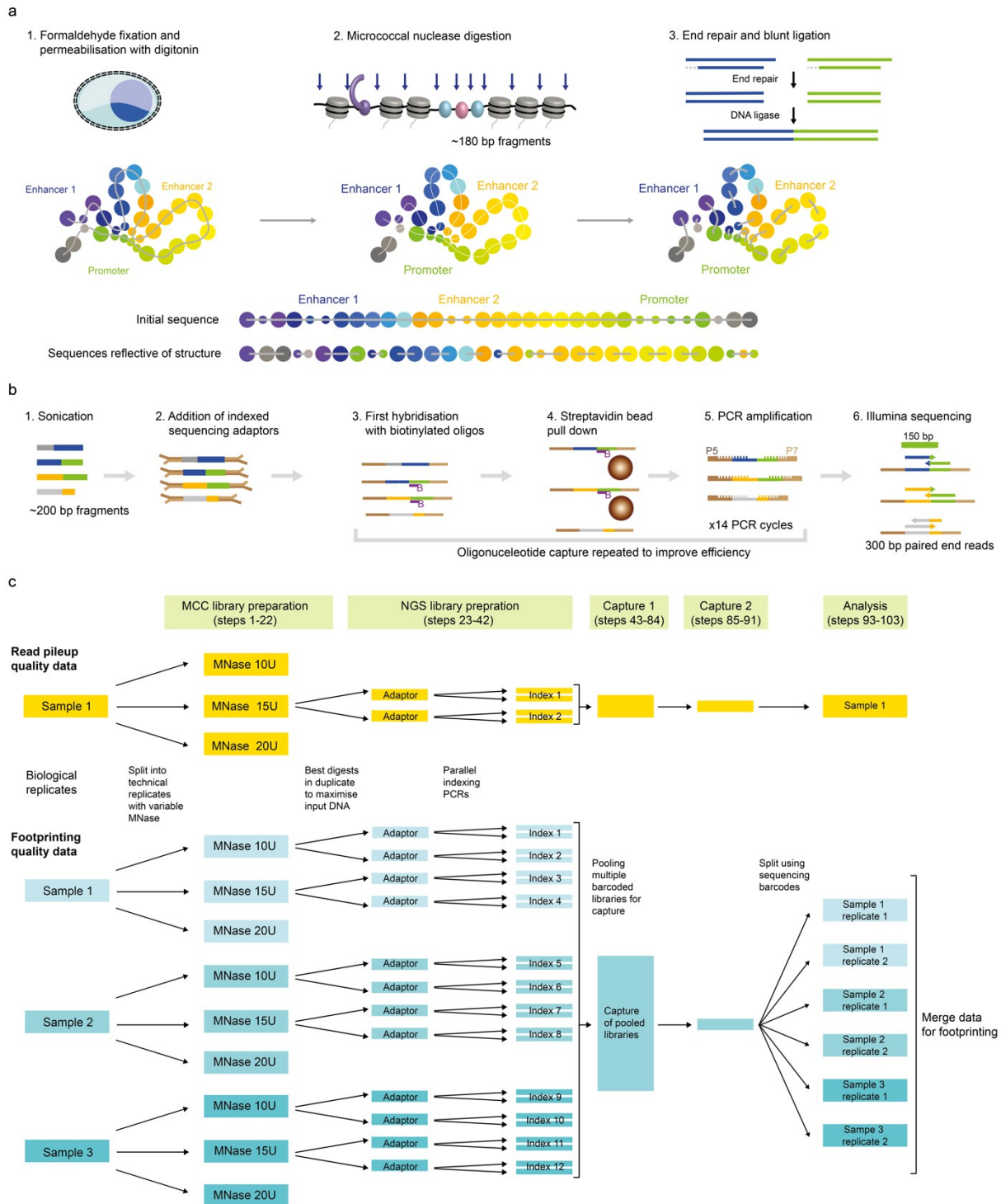


Figure 2 | Overview of experimental workflow

a, 3C principal. Involves initially fixing chromatin with formaldehyde, which causes covalent crosslinks between proteins and nucleic acids (Steps 1-6). In MCC the cells

are permeabilized with digitonin in contrast to other methods which generally use stronger detergents to extract nuclei or chromatin. The chromatin is then digested using an endonuclease. In MCC we have adopted the use of MNase, which cuts between nucleosomes and between transcription factor binding sites in nucleosome deplete regions (Steps 7-14). A blunt ligation reaction is then performed, which requires prior end repair for MNase digested chromatin (Steps 15-22). This results in ligated fragments which can be used to define which sequences are in proximity in the nucleus.

b, Overview of NGS library preparation. and oligonucleotide capture process. DNA is first sonicated to 200 bp (Steps 23-32). Sequencing adaptors are added (Steps 33-42). The material is hybridized to a pool of oligonucleotides (Steps 43-54) and pulled down with a streptavidin bead clean up (Steps 55-78). The material is PCR amplified from the beads (Steps 79-84) and the hybridization reaction is repeated to improve purity (Steps 85-91). Following this the material is analysed using paired end sequencing (Step 92). Note that the central region of the fragments is sequenced from both sides, which subsequently allows reconstruction of the entire fragment.

c, Overview of strategy for generation of lower resolution data for read pileup tracks compared to footprinting quality data.

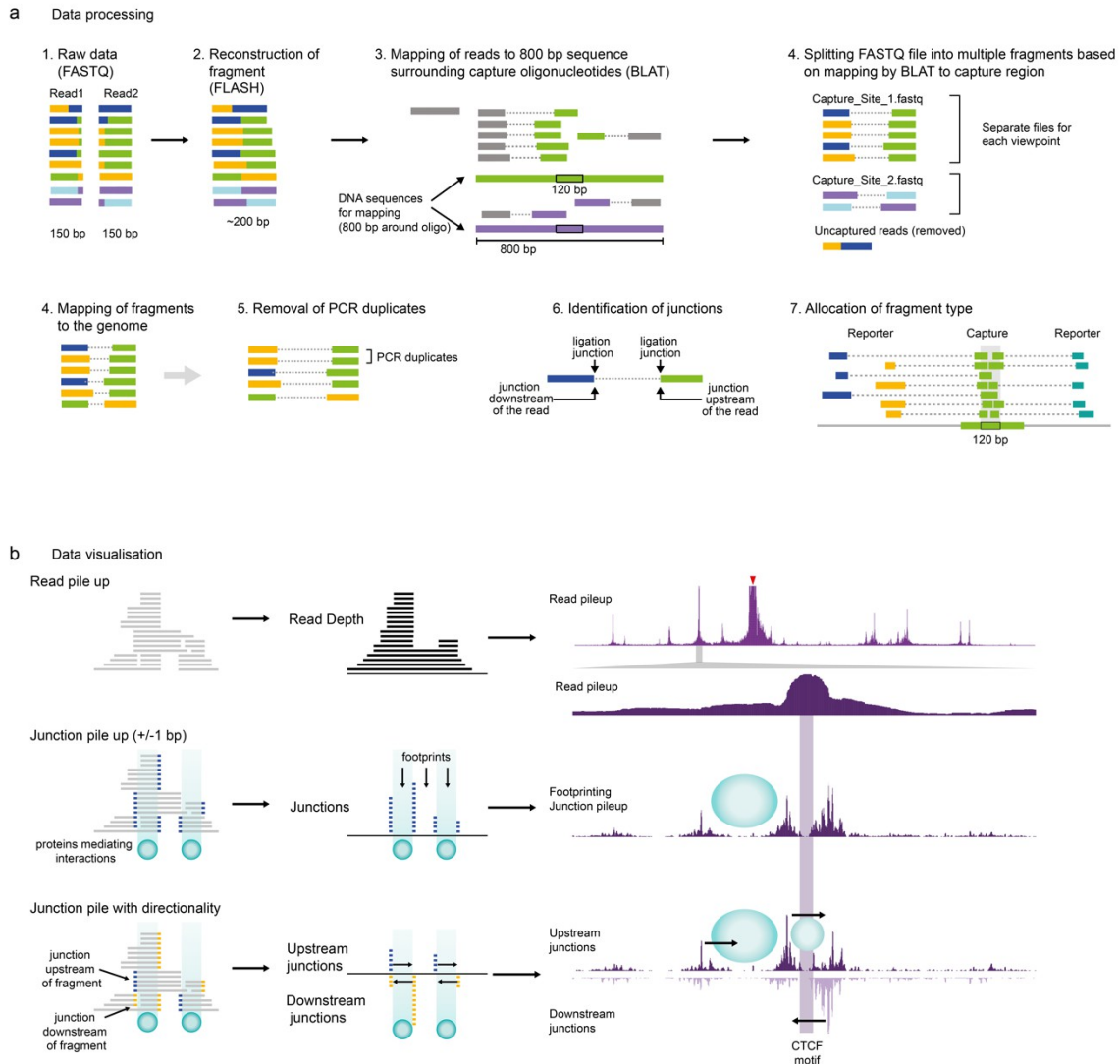


Figure 3 | Data analysis a, Data processing involves initially reconstructing paired end reads into a single sequence based on the central area of overlap (1-2). Reads are then

mapped to the 800 bp sequence surrounding the oligonucleotide (3), which allows each read to be assigned to a viewpoint and critically it identifies which part of the read maps to the oligonucleotide capture region. This allows the reads to be separated into a fastq file for each viewpoint and each read to be separated into two or more fragments depending on whether they map to a single sequence in the 800 bp around the oligonucleotide (4). These individual fragments are then mapped to the genome with a conventional aligner (we recommend bowtie2) (5). The resulting SAM file is filtered stringently to remove PCR duplicates and the location of ligation junctions is identified by tracking where the fragments lie relative to each other in the read (6). Fragments are then labelled depending on whether they map to the 120 bp sequence of the oligonucleotide or not (6-7). b, Data visualisation involves conventional read pileups as in ChIP or RNA-seq analysis. However, more precise information about the proteins mediating the contacts can be defined by generating tracks which show the precise location of ligation junctions (± 1 bp), which allows footprinting of transcription factor binding sites. This can further be improved by considering the orientation of the read relative to the ligation junction. Here the ligation junction tracks are separated depending on whether the ligation junction is upstream of the read or downstream of the read (analogous to ChIP nexus). (Adapted from Hua et al.¹⁴)

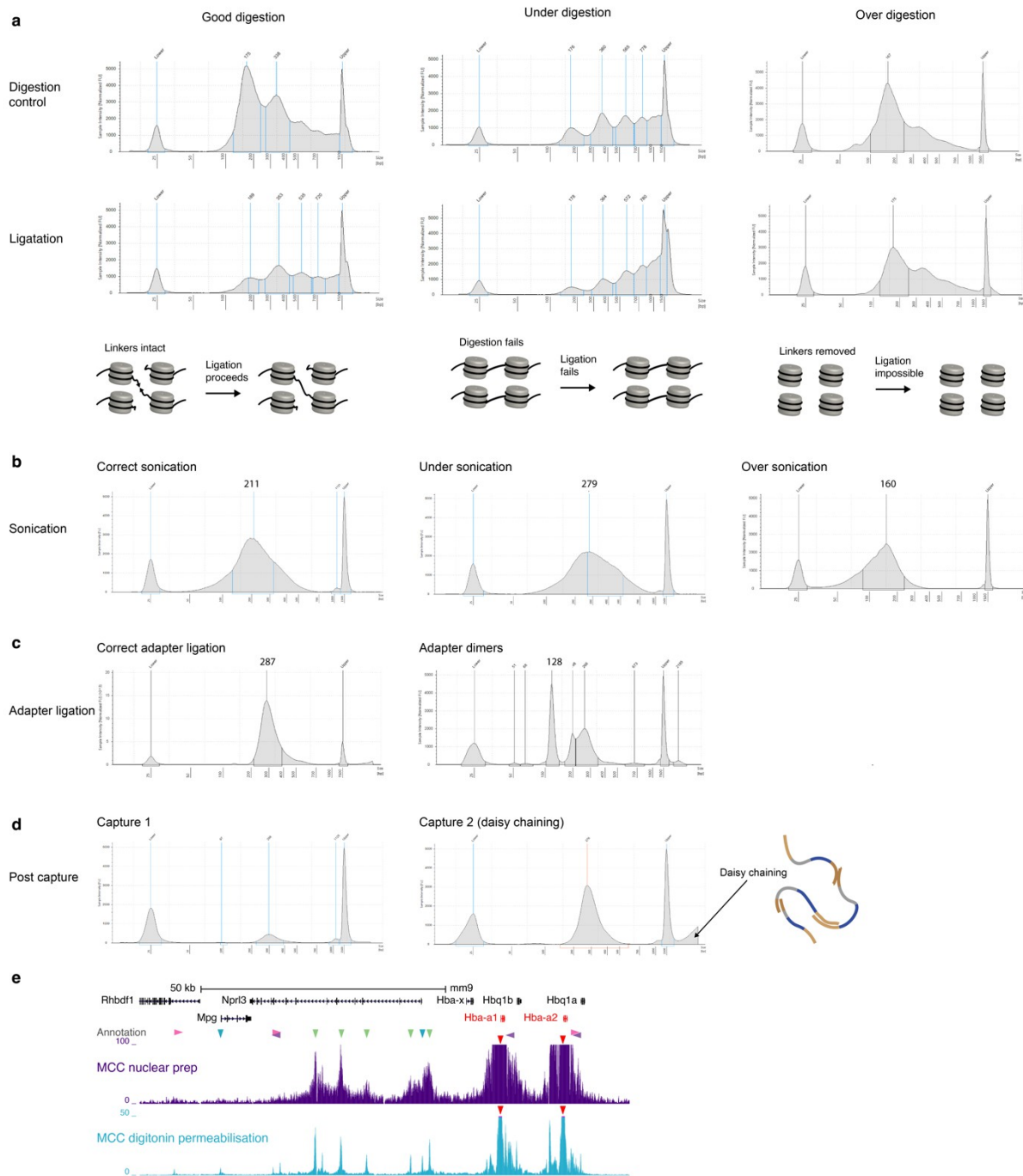


Figure 4 | Expected results. a-d, show D1000 TapeStation traces a, the results of the DNA fragment size profiles from MCC digestion reactions with examples of under and

over digested chromatin, which both fail to ligate properly. Poor quality libraries occur if either the internucleosomal linkers are not cut by MNase or if they are over digested and the linkers are completely removed, which results in failure of ligation (see Step 21 and associated troubleshooting advice). b, Shows examples of profiles for correct sonication as well as over and under sonication (see Step 25 and associated troubleshooting advice). c, Shows correct adapter ligation and an extreme example of adapter dimers. d, shows an example of traces following the first and second oligonucleotide capture (example of daisy chaining is included in the capture 2) (see Step 90 and associated troubleshooting advice). e, Shows a comparison between the profiles from an MCC experiment performed with a conventional 3C nuclear preparation and our intact cell approach, which uses digitonin to permeabilise cells. Poorly fixed and permeabilised libraries will show blurring of the profiles similar to the nuclear library preparation (adapted from Hua et al.¹⁴)

Table 1: Summary of the major differences between Micro C and Micro Capture C (MCC)

	Micro-C	MCC
Scale	Genome-wide data	Individual regions or tiled regions
Fixation	Double fixation	Single fixation
Cell permeabilisation	Nuclear prep – NP40 lysis	Intact cells – permeabilization with digitonin
Enrichment strategy	Enrichment of ligation junctions through biotin fill in and selection	Targeted oligonucleotide capture
Fragment size	Fragment size ~350 bp (gel extraction of dinucleosomal band)	Sonication to 200 bp fragment size
Sequencing strategy	50 bp PE sequencing	150 bp PE sequencing
Accuracy of ligation junction identification	+/- 200 bp (location inferred)	Reconstruction of full read and direct identification of junction to base pair resolution
Ligation junctions per bp	0.8-1.4 ³⁶	1000-1500

Table 2: Troubleshooting table

Step	Problem	Possible reason	Solution
Box 1: MCC probe design	Regions have high density scores	Trying to design an experiment to repetitive sequences	Design multiple oligonucleotides over the region and chose the sequences with less repeats. If this does not work consider reducing the length of the oligonucleotides.
21	Under digested	Low MNase concentration	Use more units of MNase per digest.
		MNase activity has reduced	Remake 1:100 diluted MNase from stock

		Poor permeabilisation with digitonin	Permeabilize subsequent reactions after defrosting with fresh digitonin
	Over digested	MNase units per digest too high	Use fewer MNase units per digest
		EGTA has degraded	Remake EGTA
		Incomplete washing after digestion	Carefully remove all the supernatant post digest spin
		Over treatment of cells with detergent	Refix cells
25	Over/under sonication	Sonication time varies based on the DNA amount.	Test sonication conditions with 1 library, checking on the tapestation, before proceeding with the rest of the libraries in order to avoid over/under sonicated libraries. Optimal fragment size should be mean size 200 bp with >80% of DNA between 150 – 300bp
42	Loss of DNA during AMPure bead cleanup	Overdried AMPure beads	Make sure to elute the beads with water before they become over dried (crack). Carefully remove all the ethanol so that droplets do not remain on the inside of the tube, if necessary using a small pipette tip.
55	Dynabeads beads sticking to plastic	High affinity of streptavidin beads for plastic tubes.	Streptavidin Dynabeads tend to stick to plastics. Consider changing microcentrifuge tubes. The effect is caused by static and is worse in some buffers than others. We

			recommend performing capture in high quality microcentrifuge tubes to minimize this effect.
82	Loss of DNA after capture	Failed PCR reaction	Perform PCR reactions in two batches to protect against error or misfortune and to determine if adequate amplification has occurred.
		User error during DNA bead cleanup	Captured material is amplified off the beads in four PCR reactions (two per hybridization reaction). Reactions should be done in two batches to protect against failure of the PCR reaction and to determine if adequate amplification has occurred.
90	Large fragment size following PCR (Daisy chaining)	Too many PCR cycles or insufficient dNTPs or primers	This is not a major problem the material will denature in the capture reaction. Consider reducing PCR cycles post capture in future to avoid this.

Box 1: Oligonucleotide Design

By altering the oligonucleotide design strategy, MCC can be used to determine contacts either for many individual viewpoints to the whole genome ('many vs all') for capture of all contacts to generate Hi-C like heatmaps for individual loci ('many vs many').

We generally recommend using individual viewpoints rather than tiled approaches because this generates higher resolution data and is significantly less costly. Any site in the genome can potentially be used as a viewpoint in MCC experiments. We have generally tended to use from 10s to 100s of viewpoints simultaneously.

In our opinion it is not possible to compare contact frequencies accurately between two different capture sites and it is therefore sensible to undertake two separate experiments if it is important to delineate the contact profiles from more than one viewpoint in a locus (for example the enhancers and promoters of the same gene).

It is important to use oligonucleotide design tools to design probes for oligonucleotide capture to prevent capture of repetitive sequences. We have extensively used Capsequm (<https://github.com/jbkerry/capsequm>) to design MCC probes.

This requires use of the command line version of the scripts with the 'Tiled' flag:

```
python design.py Tiled -f hg38_genome.fa -g hg38 -c 3 -r 94641258-94641378 -o 120 -t 120 --contig --blat
```

We recommend using 120mer probes for single viewpoints, but it is possible to use either smaller or larger probes if desired.

The recommended parameters for filtering oligonucleotides with Capsequm are: Blat Density <30%; GC content < 70%; Repeats FALSE.

When designing probes, it is important not to design oligonucleotides that target repetitive sequences. Capture from a single repetitive sequence can compromise a whole experiment with because the capture oligonucleotides are in excess and able to capture large amounts of DNA from repeats relative to well-designed oligonucleotides, which will only capture the target site. By contrast, it is safe to include oligonucleotides with a high GC content because although these may not capture efficiently, they will not affect other targets.

Oligonucleotides are in huge excess in the capture process and will capture effectively over a large range of concentrations. Take care not to order probes that should not be included in the same capture (e.g. an enhancer and promoter of the same gene) simultaneously as a small degree of contamination during manufacture will compromise the experiment.

TROUBLESHOOTING

[Production: end of Box 1]

Key Papers:

1. Hua, P., Badat, M., Hanssen, L.L.P. *et al.* Defining genome architecture at base-pair resolution. *Nature* **595**, 125–129 (2021). <https://doi.org/10.1038/s41586-021-03639-4>
2. Aljahani, A., Hua, P., Karpinska, M.A. *et al.* Analysis of sub-kilobase chromatin topology reveals nano-scale regulatory interactions with variable dependence on cohesin and CTCF. *Nat Commun* **13**, 2139 (2022). <https://doi.org/10.1038/s41467-022-29696-5>
3. Downes, D.J., Cross, A.R., Hua, P. *et al.* Identification of *LZTFL1* as a candidate effector gene at a COVID-19 risk locus. *Nat Genet* **53**, 1606–1615 (2021). <https://doi.org/10.1038/s41588-021-00955-3>