

Multi-contact 4C: long-molecule sequencing of complex proximity ligation products to uncover local cooperative and competitive chromatin topologies

Carlo Vermeulen^{1,3}, Amin Allahyar^{1,3}, Britta A. M. Bouwman¹, Peter H. L. Krijger¹, Marjon J. A. M. Verstegen¹, Geert Geeven¹, Christian Valdes-Quezada¹, Ivo Renkens², Roy Straver², Wigard P. Kloosterman², Jeroen de Ridder^{2*} and Wouter de Laat^{1*}

We present the experimental protocol and data analysis toolbox for multi-contact 4C (MC-4C), a new proximity ligation method tailored to study the higher-order chromatin contact patterns of selected genomic sites. Conventional chromatin conformation capture (3C) methods fragment proximity ligation products for efficient analysis of pairwise DNA contacts. By contrast, MC-4C is designed to preserve and collect large concatemers of proximity ligated fragments for long-molecule sequencing on an Oxford Nanopore or Pacific Biosciences platform. Each concatemer of proximity ligation products represents a snapshot topology of a different individual allele, revealing its multi-way chromatin interactions. By inverse PCR with primers specific for a fragment of interest (the viewpoint) and DNA size selection, sequencing is selectively targeted to thousands of different complex interactions containing this viewpoint. A tailored statistical analysis toolbox is able to generate background models and three-way interaction profiles from the same dataset. These profiles can be used to distinguish whether contacts between more than two regulatory sequences are mutually exclusive or, conversely, simultaneously occurring at chromatin hubs. The entire procedure can be completed in 2 w, and requires standard molecular biology and data analysis skills and equipment, plus access to a third-generation sequencing platform.

Introduction

Development of MC-4C

Chromatin conformation capture (3C) methods^{1,2} use in situ DNA digestion and proximity ligation to detect contacting DNA fragments and to investigate chromosome topologies inside cells. Most protocols also include a formaldehyde crosslinking step before digestion to preserve spatial proximity, although fixation does not seem to be strictly necessary³. Current high-throughput 3C variant methods include Hi-C (high-throughput sequencing with 3C, for unbiased and untargeted genome-wide contact analysis)⁴; 4C (3C-on-chip) and capture-C technologies (for unbiased analysis of contacts made by one or more selected genomic sites of interest)^{5–7}; 5C (3C-carbon copy) technology (for analysis of contacts between multiple selected genomic sites of interest)⁸; and ChIA-PET⁹ (chromatin interaction analysis by paired-end tag sequencing) and HiChIP¹⁰ (for contact analysis between sites occupied by a protein of interest)^{10,11}. All of these methods use Illumina short-read sequencing for massively parallel analysis of pairwise DNA contacts, which requires trimming the ligation products to sizes far below 1,000 bp for optimal bridge amplification. A limitation of all methods that measure pairwise contacts is that when three or more elements are identified to interact in a cell population, it remains unclear whether this occurs simultaneously in the same cell or in a pairwise manner in different cells or at different time points. Pairwise methods also fail to identify whether interactions occur (i) in a cooperative manner (forming preferred multi-way contacts, or hubs, on the same allele), (ii) in an uncoordinated (random) manner, or (iii) in a competitive manner,

¹Oncode Institute, Hubrecht Institute–KNAW and University Medical Center Utrecht, Utrecht, the Netherlands. ²Oncode Institute, Center for Molecular Medicine, University Medical Center, Utrecht University, Utrecht, the Netherlands. ³These authors contributed equally: Carlo Vermeulen, Amin Allahyar. *e-mail: j.deridder-4@umcutrecht.nl; w.delaat@hubrecht.eu

forming contacts that are mutually exclusive on single alleles. Increasingly, enhancers, gene promoters and CTCF-bound sequences are found engaged in contact networks with multiple other local regulatory sequences. Such contacts between individual regulatory sequences and genes, which can only be discovered by targeted high-resolution 3C approaches and in very deeply sequenced Hi-C datasets, are functionally the most meaningful chromatin interactions with the potential to directly affect gene expression. They usually take place within the context of topologically associated domains (TADs), which are chromosomal regions of (on average) 1 Mb in size that self-interact and thereby form the structural and functional units of chromosomes^{12–14}. MC-4C was developed to reach resolutions that allow the study of such local functionally important contact networks and enable investigation of whether regulatory sequences and genes spatially aggregate in chromatin hubs or, conversely, compete with each other for contacts. In recent work, we showed that MC-4C can identify enhancer hubs, formed between the individual elements of a super enhancer, and we demonstrated using MC-4C that some genes compete, whereas others cooperate with each other to interact with this enhancer hub. MC-4C further demonstrated that CTCF-bound boundaries of neighboring domains aggregate to form rosette-like chromatin structures in interphase cells. This phenomenon became even more apparent upon depletion of the cohesin-unloading factor WAPL, which led us to propose that the boundary bouquets that we observed are the consequence of ‘traffic jams’ of extruding cohesin molecules that collide into each other if not adequately removed by WAPL¹⁵.

Applications of MC-4C

MC-4C can potentially be applied to any genomic locus in any cell type or species, and can uncover how local individual regulatory sequences and genes within a 120-kb to maximally 2-Mb genomic interval (explained in the ‘Limitations’ section) spatially coordinate their action at the single-allele level. Uncovering these higher-order topological frameworks is expected to facilitate interpretation of genotype–phenotype relationships. It can help, for example, to understand why certain regulatory sequences appear redundant, whereas others seem crucial for proper gene expression. Application of MC-4C is particularly relevant when conventional pairwise chromatin contact methods (4C, 5C or Hi-C) show that, in a cell population, a given genomic site (e.g., gene promoter, enhancer or CTCF binding site) is engaged in multiple specific long-range interactions with other local regulatory DNA elements. MC-4C can then reveal whether such interactions are mutually exclusive or, conversely, cooperative at the level of single alleles. In addition to regulatory interactions and hubs, we expect MC-4C to be instrumental in uncovering repeated folding patterns, if they exist in given genomic intervals and at given points of the cell cycle. Finally, data generated by MC-4C should be useful for the validation and optimization of current 3D genome modeling algorithms^{16,17} that all rely on pairwise proximity ligation measurements to infer the higher-order topologies that MC-4C directly measures.

Comparison with other methods

Recently, several breakthrough methodologies were developed to study multi-way chromatin contacts. One example is chromosomal walks (C-walks), which uses serial dilution, fragmentation and barcoding of large proximity-ligation concatemers, followed by sequencing and in silico reconstruction of the ligation products¹⁸. C-walks is an untargeted, genome-wide method that randomly collects and identifies multi-way contacts from across the genome. It was used to investigate the spatial aggregation of intra- and interchromosomal regions inside so-called nuclear compartments (nuclear volumes in which genomic regions with similar epigenetic profiles aggregate) and therefore essentially analyzes how TADs position themselves relative to each other in the cell nucleus. It was suggested that, at this level of spatial organization, specific multi-way chromatin clustering is rare but may occur, for example, at polycomb bodies, where regions of the genome bound by polycomb factors spatially aggregate¹⁸. Three-way Hi-C contact analysis methods¹⁹ split-pool recognition of interactions by tag extension (SPRITE)²⁰ and genome architecture mapping (GAM)²¹ are other novel untargeted, genome-wide methods for investigating chromatin clustering at the level of nuclear compartments. SPRITE uses a sophisticated barcoding system based on repeated pooling and splitting of crosslinked DNA ‘hairballs’ to uncover clustered DNA fragments derived from the genomes of many cells and has enabled hubs of interchromosomal contacts at nucleoli and nuclear speckles to be revealed²⁰. GAM is an orthogonal untargeted method that is not dependent on DNA digestion and ligation but that isolates and sequences the DNA content of many ultra-thin cryo-fixed nuclear slices for in silico analysis of favored spatial chromosome configurations. GAM also showed preferential

multi-way clustering of intra- and interchromosomal regions dense in active genes and super enhancers (exceptionally large and active enhancers)²¹.

Although elegant and insightful, C-walks, GAM, SPRITE and other untargeted methods¹⁹ currently provide genome-wide data. They do not uncover thousands of multi-way contacts per individual genomic site and therefore do not achieve the resolution needed to enable analysis of higher-order topologies (hubs) formed by individual genes or enhancers with other regulatory DNA elements inside TADs. A recently developed targeted multi-contact strategy, Tri-C (ref. ²²), was designed to study such higher-order topologies of individual genomic sites and was shown to generate data and results similar to those obtained by MC-4C. Tri-C uses biotinylated capture probes instead of inverse PCR methods to enrich for multi-way contacts centered on a viewpoint fragment of interest, in a protocol that requires minimal PCR amplification. For efficient pull-down, the long proximity ligation products are first sheared to ~450-bp products before pull-down, and Illumina paired-end sequencing is used to analyze the products (~50% of products contain multiple ligation junctions). The unique ends that are created by sonication serve as unique molecule identifiers (UMIs) for filtering out PCR duplicates. Tri-C can even be scaled to analyze millions of multi-way contacts with sites across the genome, of which the great majority will be three-way. In comparison, with MC-4C we analyze roughly 10,000 untrimmed proximity ligated concatemers per single Nanopore MinION run. Most of these concatemers represent three- to five-way interactions, and some span up to ten contacts within the region of interest (ROI). Those concatemers carrying more than two local captured fragments intrinsically harbor multiple three-way interactions within the ROI. Adding up all captured fragments within the ROI across all analyzed individual concatemers shows that a typical MC-4C experiment considers nearly 100,000 different local contacts made with a single restriction fragment. Tri-C and MC-4C are orthogonal methods for targeted analysis of multi-way chromatin contacts made with a genomic site of interest (SOI).

Most researchers now have access to Illumina sequencing, which, to an extent, makes Tri-C easier to implement. However, MC-4C pioneers the application of long-read single-molecule sequencing on novel third-generation sequencing platforms (e.g., from Oxford Nanopore Technologies (ONT) and Pacific Biosciences). These new platforms are quickly becoming available to the wider scientific community and their performance (in terms of sequencing throughput, accuracy and pricing) is rapidly improving, making long-read single-molecule sequencing attractive for future multi-way chromatin interaction studies. Currently, a MinION run for the analysis of a single MC-4C experiment costs ~US\$600. The latest ONT device, the Flongle, should allow sequencing for an MC-4C experiment in a few hours, for less than US\$100. As long-read sequencing-based MC-4C analyzes large concatemers, it provides the unique opportunity to analyze more than three-way interactions, but a meaningful analysis requires exponentially deeper sequencing. With our current datasets, we are limited to the analysis of three-way interactions.

A distinct feature of MC-4C is the analysis package that comes with the protocol, which we developed to enable robust interpretation of three-way interactions formed by a sequence of interest. It evaluates each dataset independently and calculates *z*-scores for co-occurring contact frequencies based on a statistical comparison of observed three-way contact frequencies versus predicted background three-way contact frequencies in the same dataset. This enables robust distinction between cooperative, random and competitive multi-way interactions between regulatory sequences and genes in a locus of interest, as we showed recently¹⁵. Although not explored yet, we expect that it should be straightforward to apply our analysis strategy to data generated by orthogonal-targeted multi-contact methods such as Tri-C.

Overview of the protocol

MC-4C is based on the 4C technology: a ‘one-versus-all’ contact analysis method that we developed previously to search the genome in an unbiased manner for sequences contacting a selected DNA fragment of interest^{5,23}. Both methods are based on 3C (ref. ¹) and involve circularization of proximity ligation products followed by inverse PCR with primers specific for the target fragment of interest, to amplify and analyze the DNA fragments that were ligated to the fragment of interest in the cell population. The original 4C protocol aims to minimize the sizes of circles that need to be amplified by PCR, whereas MC-4C is modified such that it aims to best preserve the long concatemers with multiple DNA fragments ligated to the fragment of interest. To generate and analyze such long PCR-amplified concatemers, the procedure was tailored as described in the ‘Experimental design’ section and combined with long-read, single-molecule sequencing on third-generation

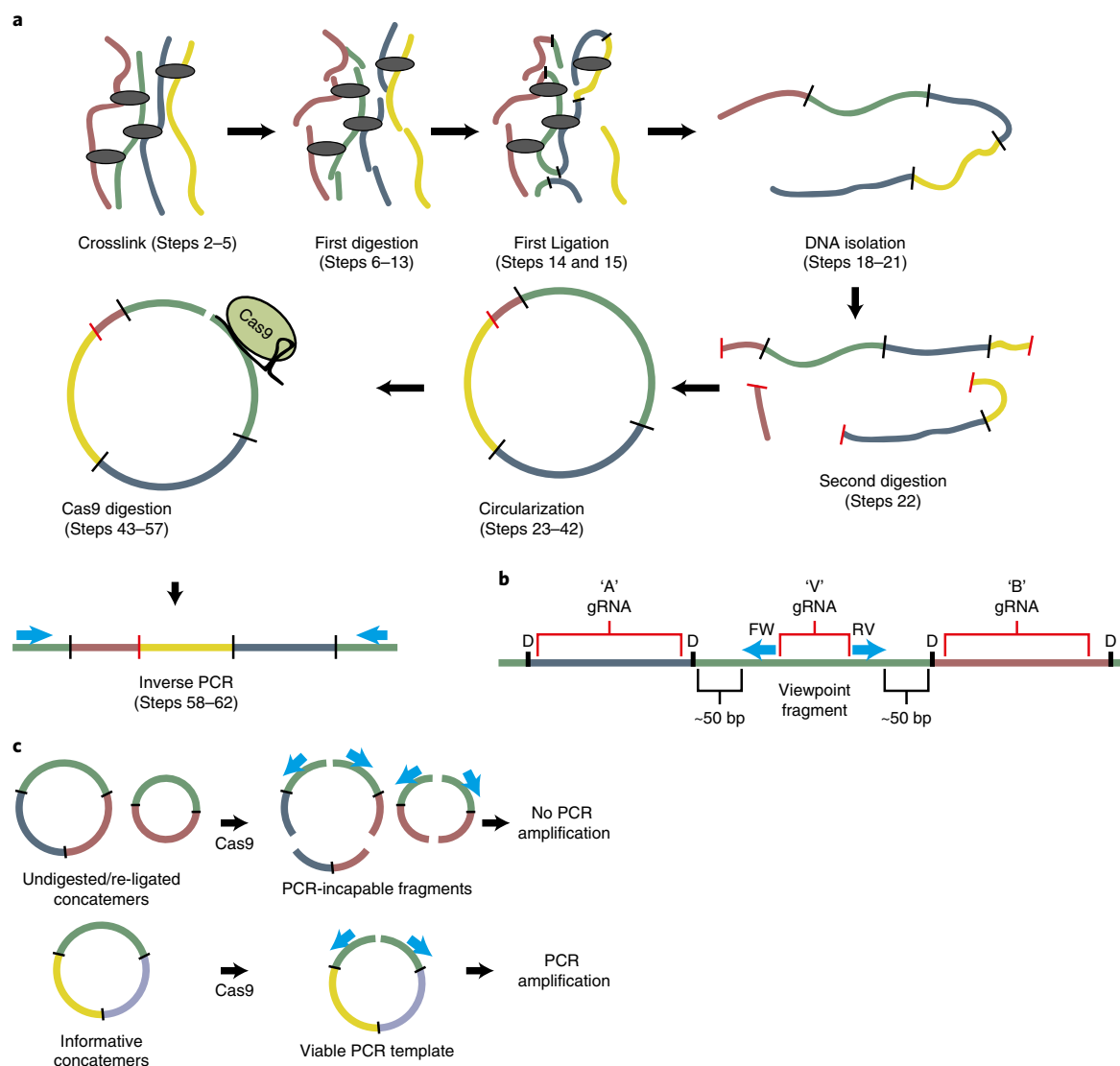


Fig. 1 | Schematic overview. **a**, Summary of the MC-4C protocol: chromatin is first crosslinked using formaldehyde and then digested using a restriction enzyme (DpnII). Subsequently, fragments in close physical proximity are ligated in situ. The DNA concatemers are then isolated and re-digested with a 6-bp cutter restriction enzyme. Subsequent ligation results in circularization of the template concatemers. A Cas9 digestion using three gRNAs (specific for the viewpoint of interest and the fragments flanking it in the linear genome) is then applied to linearize the template while digesting frequent background products and rolling-circle PCR products, thus avoiding their subsequent amplification. Next, a viewpoint-specific inverse PCR is performed. **b**, Primer and gRNA design overview. For the selected viewpoint fragment, a set of inverse primers is designed with ~50 bp between the primer binding site and the recognition sites of the first cutter (DpnII, recognition sites indicated with 'D'). Three gRNAs are then designed to cut in-between the inverse primers (gRNA 'V', and in each of the 'neighboring' fragments directly adjacent to the viewpoint fragment on the linear genome (gRNA 'A' and 'B')). **c**, Rationale of cutting the 'neighboring' fragments. Viewpoint-fragment self-ligation and undigested or re-ligated concatemers are frequent by-products of proximity ligations. These take up a large proportion of the PCR product and sequencing capacity. When the three gRNAs are successfully used to target Cas9 to the intended sites, such by-products are fragmented and can no longer yield PCR product, whereas informative ligations between the viewpoint and non-neighboring fragments are left intact.

sequencing platforms such as minION or PromethION platforms offered by Oxford Nanopore Sequencing or Pacific Biosciences Sequel platforms.

The protocol aims to isolate single-cell suspension from a tissue or cell type of interest and to fixate the chromatin conformation using formaldehyde. Subsequently, the chromatin is fragmented using a restriction enzyme, and in situ ligation is performed to concatenate fragments in close proximity. Next, the concatemers are isolated and purified from the fixated nuclei. To further circularize the template, an additional round of enzymatic cleavage is performed, followed by a self-ligation. To avoid rolling-circle PCR amplification and to increase the fraction of informative PCR products, an in vitro Cas9 cleavage is performed, targeted at the region in between the inverse primers and at the fragments directly adjacent to the viewpoint fragment (see Fig. 1a for an overview). A PCR is

performed on the template and size selected for large products. These are subsequently sequenced on an ONT platform. The processing and analysis pipeline consists of a tailored mapping and duplicate removal algorithm. The processed data can then be analyzed with the tools provided for the analysis of three-way contact profiles.

Experimental design

Proximity ligation

As in most 4C and Hi-C protocols, four-cutter restriction enzymes are used for in situ digestion of formaldehyde crosslinked DNA. This results in small DNA fragments (average size, 250 bp) that, as a consequence of formaldehyde fixation, often stick together in ‘hairballs’ with multiple other cross-linked DNA fragments. Although we have successfully used MboI and DpnII for the first digestion, other four-cutter enzymes are likely to yield similar results. We note, however, that the restriction enzymes must be able to reach and digest chromatin in crosslinked nuclei, which may not be the case for all restriction enzymes. Subsequent in situ ligation fuses these ‘hairball’ DNA fragments together, yielding large concatemers of ligated fragments.

Template circularization

We further process these large DNA molecules with a secondary six-cutter restriction enzyme (HindIII) to obtain concatemers with sticky ends (for efficient ligation-mediated circularization) and of a size (average, 4 kb) that will harbor many ligation partners but can still be amplified by PCR with inversely oriented primers that bind to the restriction fragment of interest (the viewpoint). Although we have exclusively used HindIII for this purpose, many other six-cutter enzymes are available and are likely to yield similar results as the secondary digestion is performed on clean DNA molecules. We do note that circularizing blunt rather than sticky-ended fragments is less efficient.

Viewpoint fragment

It is essential that the viewpoint fragment (i) is of a sufficient size to accommodate two primer binding sites and a guide RNA (gRNA) binding site in between the primers (the viewpoint fragment is therefore usually between 200 and 500 bp); (ii) overlaps or is at least in close (<1 kb) linear proximity to the genomic element for which multi-way interactions are to be studied (i.e., overlapping with a promoter, enhancer or CTCF binding site); and (iii) does not contain any internal secondary restriction (i.e., HindIII) sites.

PCR primer design

For the viewpoint fragment, a set of inverse primers must be designed that bind uniquely in the genome and are positioned more than 50 bp from each side of the DpnII restriction sites. Allowing some spacing between the restriction sites and the primer binding sites facilitates viewpoint detection after long-read sequencing, as the high-sequence error rate may prevent mapping of short (i.e., primer) sequences. We recommend testing the PCR primers on a small batch of MC-4C template, to ensure that they yield products in the 2–5-kb range. If this is not the case, their efficiency can be optimized with an alternative annealing temperature and the addition of DMSO and/or MgCl₂ to the PCR reaction, although satisfying results are often faster obtained by simply designing a new set of primers elsewhere on the same viewpoint.

PCR amplification of concatemers

PCR amplification is performed using a highly processive polymerase (PhireII hotstart polymerase) in a reaction with an extension time long enough to enable amplification of the entire length of the long concatemers.

In vitro Cas9 digestion

In PCR reaction conditions optimized for amplifying long template molecules, rolling-circle PCRs can occur when shorter, circular template DNA is present, as is the case with an MC-4C library. We noticed that this causes an overrepresentation of small products with only the viewpoint fragment of interest with or without its natural neighbor fragments on the linear chromosome. These small products account for most reads but are not informative for genome topology studies. To eliminate the uninformative small fragments and target sequencing to informative concatemers, we included an in vitro Cas9 digestion step before PCR that simultaneously cuts the target fragment of interest

(in between the inversely oriented PCR primer binding sites) and its two neighboring fragments (the fragments directly flanking the viewpoint fragment in the linear genome) (Fig. 1b,c). Note that Cas9 digestions are used to exclude unwanted products, not to improve PCR specificity, which is determined by the PCR primers and not affected by possible off-target Cas9 activity. gRNAs can be designed using several available online tools (described in the 'Guide RNA design' section). We routinely first test their effectiveness on genomic DNA or a small amount of MC-4C template before using them on bulk MC-4C template. For this test, we designed PCR primers to amplify across the Cas9 cleavage target site and measure the *in vitro* digestion efficiency of Cas9 at this site by qPCR and comparison of the C_T values between cleaved template and uncut genomic DNA.

gRNA design

For each MC-4C viewpoint, we recommend designing three different gRNAs, one in between the inverse PCR primers and one in each of the two DpnII restriction fragments directly flanking the viewpoint fragment in the linear genome. We use the ATUM gRNA design tool (<https://www.atum.bio/eCommerce/cas9/input?multipleContacts=false>). Set the design to wild-type Cas9 and enter the restriction fragment sequence into the 'Target a sequence' dialog box. In some cases, extremely small DpnII fragments flank a viewpoint; in such cases, we target the closest fragment for which gRNA design is possible. Select the gRNA with the highest score and click 'Continue'. This will show the gRNA targeting sequence (excluding the universal scaffold sequence) to be used. It is advisable to perform a BLAT search (<https://genome-euro.ucsc.edu/cgi-bin/hgBlat>) for potential off-targets: when unintended hits with high (>90%) similarity to the gRNA sequence are found elsewhere in the genome, we recommend choosing a different gRNA.

gRNAs can be obtained commercially or can be synthesized through *in vitro* transcription; in the latter case, a T7 promoter must be included in the template DNA as well as the CRISPR sequence. Therefore, we order DNA oligonucleotide consisting of the following sequences for each gRNA sequence:

5'-AAGCTAATACGACTCACTATAGG-gRNA-GTTT'TAGAGCTAGAAATAGCAAG-3'

See Box 1 for examples of gRNA template primers. Note that *in vitro* transcription initiates at the second G-nucleotide of the GG dinucleotide at the 3' end of the T7 promoter sequence (marked in bold). If the gRNA starts with a G, this last G of the T7 promoter can be omitted. For more detailed instructions on the design and synthesis of gRNAs, see Box 1.

Long-read sequencing

Before sequencing, products are size selected and sequenced on a long-read, single-molecule sequencing platform, following the standard library preparation procedures recommended by ONT or Pacific Biosciences.

Data analysis

To facilitate data processing, analysis and interpretation of sequenced MC-4C experiments, we developed a computational tool called MC-4C pipeline that not only processes the sequenced MC-4C data but also stringently removes PCR duplicates. A schematic overview of the MC-4C pipeline is demonstrated in Fig. 2. The resulting processed dataset can be readily used to identify and quantify three-way interaction profiles. In addition, this pipeline includes an analysis module that compares observed three-way interaction frequencies to dataset-specific background contact frequencies, allowing data-intrinsic statistical analysis of complex interaction profiles. The MC-4C pipeline is available as Supplementary Data 1, and a small test dataset is available as Supplementary Data 2. It should be noted that this pipeline is under continuous development that adds new functionalities or enhances the existing ones. Therefore, we recommend that the user refers to the GitHub repository of MC-4C pipeline (https://github.com/deLaatLab/mc4c_py) to access the latest version of the MC-4C pipeline as well as the corresponding manual (the most recent manual at the time of publication is included in Supplementary Manual), which describes the newly added features or enhancements in the pipeline. Below, we describe the general requirements in the MC-4C pipeline.

Configuration files

Experiment-specific parameters for each MC-4C dataset are organized in a configuration file. Parameters in this file are defined line by line and stored in a two-column, tab-delimited format (i.e., parameter name and value). Multiple values for a parameter are separated by a ',' sign. For example,

Box 1 | gRNA preparation and validation ● **Timing 2 d**

Procedure

The protocol for synthesis of gRNAs is adapted from Nakayama et al.³⁰

- 1 Perform a gRNA template PCR using Q5 High Fidelity Polymerase and a universal reverse primer together with the gRNA-specific primer to make double-stranded template for the in vitro transcription in Step 3.

For example, the gRNA primers used to synthesize the gRNAs used in the β -globin HS2 experiments in our previous publication¹⁵ are listed below. See Fig. 1 for further explanation of the binding location relative to the viewpoint and intended purpose of each gRNA.

HS2 viewpoint gRNA FW	5'-AAGCTAATACGACTCACTATAGGGTGGAGAACTGACTCCGCCTGTTTTAGAGCTAGAAATAGCAAG-3'
HS2 neighbor A gRNA FW	5'-AAGCTAATACGACTCACTATAGGTTTTCTTTTCACCTTCCCTGGTTTTAGAGCTAGAAATAGCAAG-3'
HS2 neighbor B gRNA FW	5'-AAGCTAATACGACTCACTATAGGGCAGAGGCTCGTTAATGTGTGTTTTAGAGCTAGAAATAGCAAG-3'
Universal reverse (RV) primer	5'-AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTATTTCTAGCTCTAAAC-3'

The target-specific FW and universal RV primer can be pooled into a pre-mix containing 5 μ M of each primer.

Reagent	Amount (1 reaction, μ l)	Amount (4 reactions, μ l)	Final concentration
5 \times Q5 reaction buffer	10	40	1 \times
5 μ M primer mix	2.5	10	0.25 μ M
10 mM dNTPs	1	4	0.2 mM
Q5 polymerase (2,000 U/ml)	0.5	2	20 U/ml
MilliQ water	36.0	144	—
Total	50	200	—

Run the following program:

Cycle number	Denature (98 $^{\circ}$ C)	Anneal (57 $^{\circ}$ C)	Extend (72 $^{\circ}$ C)
1	30 s	—	—
2–36	10 s	10 s	10 s
37	—	—	5 min

The resulting gRNA template product should be 125 bp. Optionally, this can be verified on a 2% (wt/vol) agarose gel stained with ethidium bromide (0.5 μ g/ml) (Supplementary Fig. 1).

- 2 Purify the PCR product using Qiagen's PCR Purification Kit, elute in 25 μ l of RNase-free water, and quantify using the Qubit spectrophotometer, and the DNA Broad Range Kit usually concentration is between 125 and 180 ng/ μ l.
- 3 Perform in vitro transcription using the T7 Megashortscript Kit while taking care to avoid RNase contamination. Assemble the reaction in the following order:

Reagent	Amount	Final concentration
DEPC water	X (up to 10 μ l final reaction volume)	—
Template DNA	(500 ng) up to 4 μ l	50 ng/ μ l
10 \times Transcription buffer	1 μ l	1 \times
NTP mix (18.75 mM each)	4 μ l	7.5 mM each
T7 enzyme mix	1 μ l	—
Total	10 μ l	—

Mix well and incubate at 37 $^{\circ}$ C for 4 h.

- 4 Add 0.5 μ l of TURBO DNase to remove the DNA template and incubate at 37 $^{\circ}$ C for 10 min.
- 5 Perform a 4 \times AMPure XP purification as described in Steps 47–56 of the main procedure but adding 40 μ l (4 \times sample volume) of AMPure XP beads to the sample. Elute in 100 μ l of RNase-free water and quantify using the Qubit RNA BR Kit. gRNAs can be stored at -80° C for at least several months.

■ PAUSE POINT (Optional) gRNAs can be frozen in 2- μ g aliquots if they are to be used repetitively. This avoids degradation due to repeated freeze-thawing.

Box 1 | gRNA preparation and validation ● **Timing 2 d (Continued)****Testing gRNA efficiency** ● **Timing 8 h**

6 gRNA efficiency can be tested on 5-μg genomic DNA (prepared, for example, using a phenol-chloroform extraction³¹ or a commercial DNA Extraction Kit) or, if available, on 5-μg MC-4C template (as obtained at Step 42 of the main procedure). Prepare a Cas9 pre-incubation mixture as follows and incubate for 30 min at room temperature:

Reagent	Amount (μl)	Final concentration
10× NEB 3.1	1.25	1×
gRNA (100 ng/μl)	1.25	10 ng/μl
Cas9 enzyme (1 μM)	0.6	48 nM
DEPC water	9.4	—
Total	12.5	—

- 7 Add an appropriate volume of NEB 3.1 buffer to the gDNA solution so that the final buffer concentration would be ~1× (e.g., add 11 μl of 10× NEB3.1 to 100 μl of gDNA solution).
- 8 Add the Cas9 pre-incubation mixture(s) and incubate for 3 h at 37 °C.
- 9 Add 1/25th volume of 10% (vol/vol) SDS and incubate for 10 min at 70 °C to inactivate the Cas9 enzyme.
- 10 Purify the template using a 0.6× AMPure XP purification (see Steps 47–56 of the main procedure).
- 11 Perform a qPCR using qPCR primers that hybridize on opposite sides of the Cas9 cleavage site, while using uncut gDNA or MC-4C template as a negative control. For normalization of the input quantity, we also recommend including an independent primer set. Digestion efficiency can then be calculated by comparison between the normalized C_T values of the uncut template. For example, one may obtain the following C_T values:

Template	Independent primer set	Cleavage site spanning primer set
Uncut	26	30
Cas9 treated	25	32

In this case, the difference in input quantity has to be normalized first, as there is a C_T value difference between the Cas9-treated sample and the control (uncut) template. If the Cas9 cleavage failed completely, the difference in C_T values between the control amplicon (amplified by independent primer set) and the test amplicon (amplified by cleavage sites spanning primer set) should be the same in the Cas9-treated template as in the control (uncut) template. Here, however, this difference is 4 C_T values (30 – 26) in the control (uncut) template, versus 7 C_T values (32 – 25) in the Cas9-treated sample. Therefore, there was 3 C_T values (7 – 4) less amount of test amplicon in the Cas9-treated samples, which corresponds to $-0.5^3 \times 100\% = 12.5\%$ of the original amount. Digestion efficiency was therefore ~87.5%. Efficiency of Cas9 cleavage is usually 80–95% (corresponding to 3 – 6 delta- C_T values).

? TROUBLESHOOTING

the following configuration file specifies that the viewpoint lies on chromosome 7, that fragments should be mapped to the mm9 mouse reference genome, and finally, that GATC and AAGCTT are used as first and second restriction enzymes to prepare the library:

```
vp_chr <tab> chr7
genome_build <tab> mm9
re_seq <tab> GATC,AAGCTT
```

Parameters that are recognized in the MC-4C pipeline and can be stored in a configuration file are listed in Table 1.

The user is strongly advised to consider the following suggestions when determining the coordinates for the ROI:

- The viewpoint should always be included in the ROI. This is important because exclusion of the viewpoint can change the expected distribution of fragments in the ROI, which in turn results in unreliable z-scores calculated in the three-way contact analysis (see the ‘Multi-contact analysis’ section for details).
- If an experiment is expected to investigate interactions within a given TAD, the ROI should encompass the TAD boundaries and a few dozen kbs of flanking sequences, as done previously¹⁵.
- The ROI size should not be <120 kb when the MC-4C library is prepared using the DpnII restriction enzyme as the first cutter. Smaller ROIs will result in a bin width of <600 bp (i.e., 120 kb/200), which is close to the expected minimum resolution of the MC-4C experiment (for human or mouse genomes).

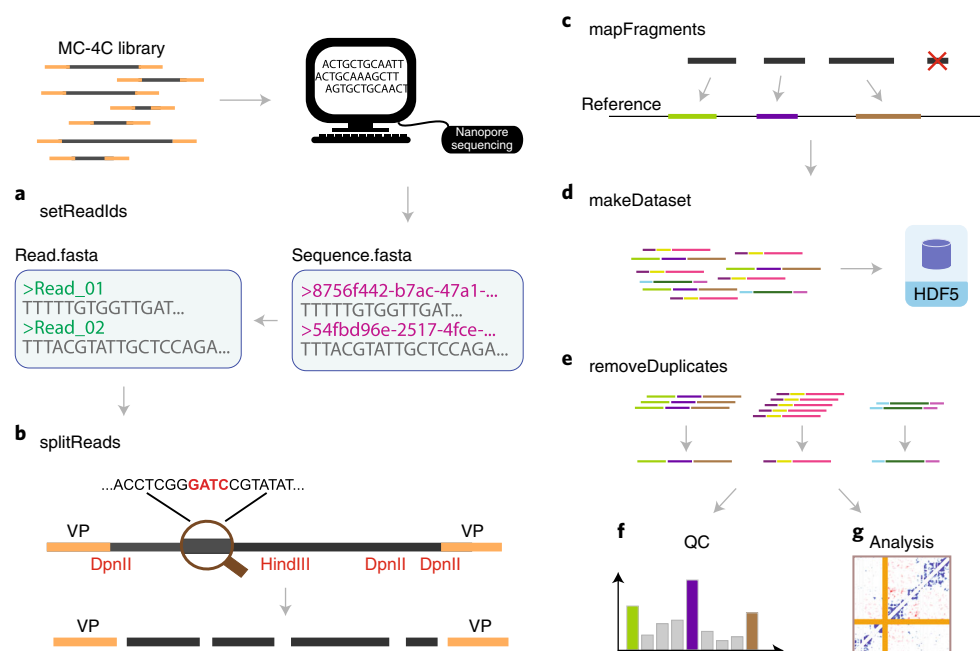


Fig. 2 | Schematic overview of the MC-4C pipeline and its modules. Each module is responsible for a specific task. **a**, Identifiers are assigned to each sequenced read to facilitate data tracking across the MC-4C pipeline. **b**, Reads are split into fragments according to restriction enzyme recognition sequences found in each sequenced read. **c**, Fragments are mapped to the reference genome (according to the configuration file). **d**, Mapped fragments are stored in a dataset (in HDF5 format). **e**, Duplicates are removed by discarding all but one reads with identical UMI. **f**, The user can use commands in **QC** module to assess the quality of performed MC-4C experiment. **g**, Commands in **analysis** module can be used to investigate the three-way contact associations captured in an MC-4C dataset.

- The ROI can be larger if interactions between TAD boundaries are investigated as done previously¹⁵. However, considering the expected coverage of an MC-4C experiment performed with a 4-bp cutter, the ROI should not be >2 Mb, because sufficient coverage is required for a reliable three-way contact analysis and further interpretation of results (further discussed in ‘Data quality control’ section).

Annotations

Several quality control functions or analyses in the MC-4C pipeline use or require annotations. These analyses include `overallProfile`, `atVpSoi`, `atSOISOI` and `atAcrossROI`. For example, in the `atSOISOI` analysis, each SOI should have coordinates assigned to it to be used in the association test and plotting. For each analysis, only annotations within the ROI are considered. These annotations can be stored in a tab-delimited file, which is named according to the genome build of interest (e.g., mm9 or h19). For example, SOIs in the mm9 genome should be stored in an annotation file named:

```
./annotations/ant_mm9.tsv
```

which follows the format below for coordinates:

```
Hbb-bh1 <tab> chr7 <tab> 110990974
HS1 <tab> chr7 <tab> 111007686
HS2 <tab> chr7 <tab> 111009600
```

In this case, we defined mm9 coordinates for the hemoglobin subunit beta-H1 (Hbb-bh1) gene and two hypersensitive sites (HS1 and HS2) using the example annotation file above.

Table 1 | Description of parameters that can be defined in a configuration file to be used for processing an MC-4C experiment

Name	Value example	Description
genome_build*	hg19	Reference genome of interest
vp_chr*	Chr7	Viewpoint chromosome; the chromosome for which the viewpoint primers are designed
prm_start*	110977147, 110977593	Start coordinate of primers used. Coordinates are separated by ','.
prm_end*	110977171, 110977623	End coordinate of primers used. Coordinates are separated by ','.
prm_seq*	GATTTGTGAGCTCAGGGTTTAC, GCAGTAGTGATTCTATTCAATTTTGGG	Sequence of primers used. Separated by ','.
re_name	DpnII, HindIII	Restriction enzyme name used to prepare the MC-4C library
re_seq*	GATC, AAGCTT	Restriction enzyme sequence used to prepare the MC-4C library
bwa*	/bin/bwa	Path to BWA aligner
bwa_index*	/bwa_indices/mm9	Path to corresponding bwa index of reference genome
reference_fasta*	/genomes/mm9.fa	Path to the corresponding reference genome (in fasta format)
roi_start	110933500	Start position of ROI. ROI, for example, will be used to define 'far-cis' fragments in PCR duplicate filter. This parameter will be set by default to 1 Mb upstream of the smallest primer coordinate, if a value is not given
roi_end	111066500	End position of ROI. This parameter will be set to 1 Mb downstream of the largest primer coordinate, if not specified

Selection of an ROI—the genomic interval in which three-way contacts will be analyzed for significantly co-occurring frequencies with the sequence of interest (viewpoint)—requires careful consideration. This is because ROI width has an important role in several parts of the MC-4C data processing pipeline. Initially, the ROI discerns far-cis fragments that can be used as UMIs for the duplication removal procedure. In addition, it identifies 'informative' reads (i.e., reads with more than one fragment in the ROI representing multi-way interactions with the view point) that can be used in three-way contact analysis. Finally, the three-way contact analysis is performed by splitting the ROI into 200 non-overlapping and equally spaced bins. Binning of the ROI can reduce coverage variability across adjacent restriction fragments in the linear chromosome (arising from factors such as sequence uniqueness and fragment size). Coverage stability in turn not only facilitates the visual exploration of contact profiles but also promotes a more robust association test in the downstream analysis. *Required parameters.

Modules

The MC-4C pipeline is partitioned into modules. Each module performs a specific task such as mapping fragments to the reference genome or removing PCR duplicates from a given MC-4C dataset. A list of available modules in the MC-4C pipeline is given in Table 2. For example, the user can map the fragments to the reference genome by calling the **mapFragments** module using the following command:

```
$ mc4c.py mapFragments <config_file> --input_file <input_file> --output_file <output_file>
```

In this protocol, the module names are shown in bold, and the <config_file>, <input_file>, <output_file> placeholders are used for the path to the configuration file, input file and output file, respectively. Arguments that are not provided will be obtained from the configuration file if possible (see MC-4C pipeline repository for further details).

Equipment needed to implement the protocol

The preparation of the MC-4C template is very similar to our published protocols for the preparation of the 4C template^{6,23} but has been optimized by the use of magnetic beads for DNA purification. Thus, the use of (commercially available or homemade) magnetic separators is required. The final

Table 2 | Modules defined in MC-4C

Module name	Function
setReadIds	Defines an identifier for each sequenced read
splitReads	Splits each read into fragments according to restriction enzyme recognition sequence
mapFragments	Maps the fragments to the reference genome
makeDataset	Creates a dataset (in HDF5 format) containing mapped fragments
removeDuplicates	Removes duplicate reads from an MC-4C dataset
QC	Generates various summary report plots for an MC-4C dataset
analysis	Performs three-way contact analysis including VP-SOI (viewpoint sequence of interest; a single SOI, that is, the viewpoint, versus all SOIs), SOI-SOI (every pair of SOIs) or cross-ROI (taking every three bins as an SOI) analysis

steps of the protocol, size selection and long-read single-molecule sequencing, require a specialized gel-separating device and an ONT or Pacific Biosciences sequencing device.

Limitations

A limitation of the MC-4C protocol is the high starting number of cells that are required. To accommodate the loss of material that occurs at various steps of the protocol (through inefficient circularization of large DNA fragments, among other reasons), and for the fact that most of the PCR products are smaller than the ideal size (>1.5 kb), we generally pool 96 individual PCR reactions for each viewpoint, with a total of 100 ng of template per PCR reaction. Thus, approximately 10 µg of Cas9-treated template is required per viewpoint (whereas conventional 4C requires only 800 ng); therefore, we aim to have 20 µg of 4C template before Cas9 digestion. Of 2×10^7 cells, we routinely generate 60 µg of 4C template, allowing for 3 MC-4C analyses. For most tissue culture cells, obtaining 2×10^7 cells is not difficult, but for certain primary tissues or FACS-sorted cell populations, this may be prohibitive amounts of material. Furthermore, MC-4C only provides meaningful insights into higher-order topologies for genomic intervals within a given size range. At the lower size limit, we recommend that regions <120 kb are not considered: our current analysis strategy pools restriction fragments into a standard number of 200 bins across the ROI, no matter the size of this region. Below 120 kb, the sizes of these bins become <600 bp (i.e., often erroneously implying a resolution below the restriction fragment size). At the upper size limit, relevant data interpretation requires a minimum amount of multi-way contacts with distal sites, and this relies on local sequence depth, which decays with increasing genomic distance. In practice, this limits analysis to contacts within TADs or between a few adjacent TADs within 1–2 Mb (further explained in the ‘Data analysis’ section).

We note that, compared to Illumina sequencing, the sequencing throughput provided by ONT and Pacific Biosciences platforms is still limited (but improving rapidly). Sequencing output fluctuates between experiments: we obtain between 150,000 and 3.65 million reads per experiment, but we expect that this range will stabilize or even increase subsequent to further experimenterunning MC-4C experiments on ONT sequencing devices (for example, with further optimizations to library preparation, sample quantification and loading protocols). Indexing strategies are not yet routine practice. As a result, we currently run only single MC-4C experiments per flow cell. However, when different viewpoints are used, the viewpoint-specific primer sequences should be traceable, and can serve as an inherent barcode, which should allow pooling of multiple samples in a single run. Finally, it is important to realize that our methodology uncovers highly complex multi-way contacts centered on a single SOI, but that the current throughput (i.e., the number of individual concatemers analyzed) is too limited for a meaningful analysis beyond that of three-way interactions. Our current analysis pipeline is therefore tailored to extract all different three-way interactions from the collection of concatemers and interpret the significance (z score) of their frequencies of co-occurrence with the SOI. Increased throughput in the future should enable use of our protocol to also study spatial DNA aggregation beyond three-way contacts.

Materials

Biological materials

▲ **CRITICAL** MC-4C is theoretically applicable to any tissue or cell type, irrespective of species, as long as a single-cell suspension can be obtained. Successful experiments have been performed previously in NIH-3T3 cells, Hap1 cells, K562 cells, embryonic day 14.5 (E14.5) liver cells and E14.5 brain cells.

- Cell lines for analysis: K562 human erythroleukemic cells were cultured in RPMI 1640 medium (Gibco) with 10% (vol/vol) FBS and 1% (vol/vol) penicillin–streptomycin. Cell identity was validated previously by translocation breakpoint mapping of the known K562 translocation breakpoints. Cells were tested for mycoplasma contamination every 2 months **!CAUTION** All cell lines should be checked regularly for authenticity and mycoplasma contamination.
- Mouse strains for liver isolation: C57/B6 mice (Charles River). The Animal Welfare Body of the Hubrecht Institute declares that the use of experimental animals is in compliance with the Dutch law on animal experimentation in line with the European Directive (2010/63/EU). The described studies are part of the project entitled ‘Single-cell quantification of the regulation of gene expression in the first stages of early embryonic development’ that was licensed by the Dutch Competent Authority (AVD 801002016728). The described studies and work protocols were subsequently positively reviewed by the Animal Welfare Body of the Hubrecht Institute, listed as HI 17.33.01-AVD80100 2016 728 **!CAUTION** Any experiments that involve live mice must conform to relevant institutional and national regulations.

Reagents

- 2-Propanol (Boom, cat. no. 76025379) **!CAUTION** 2-Propanol is flammable.
- AMPure-XP beads (Beckman Coulter, cat. no. A63881)
- Cas9 (New England Biolabs, cat. no. M0386) ▲ **CRITICAL** Take care to avoid RNase contamination.
- Cas9 buffer (NEB3.1; New England Biolabs, cat. no. B7203.5) ▲ **CRITICAL** Take care to avoid RNase contamination.
- dNTPs (Roche, cat. no. 3622614001)
- DpnII buffer (New England Biolabs, cat. no. B05435)
- DpnII enzyme (New England Biolabs, cat. no. R0543) ▲ **CRITICAL** For consistent digestion, we recommend using this particular brand and product number.
- Ethanol (Boom, cat. no. 84028185) **!CAUTION** Ethanol is flammable.
- EDTA (Invitrogen, cat. no. 15576028)
- Ethidium bromide (Sigma-Aldrich, cat. no. E8751-1G) ▲ **CRITICAL** Suspected to be carcinogenic; use gloves and avoid skin contact.
- FastStart Universal SYBR Green Mastermix Mix (Roche, cat. no. 4913914001)
- FBS (Sigma-Aldrich, cat. no. F7524)
- Formaldehyde (37% (wt/vol) solution; Sigma-Aldrich, cat. no. 1.03999.1000) **!CAUTION** Very hazardous; observe material safety data sheet, use in fume hood and dispose in appropriate waste container. ▲ **CRITICAL** For consistent crosslinking, we recommend using this particular brand and product number.
- Glycine (J.T. Baker, cat. no. 0582-05)
- HindIII (New England Biolabs, cat. no. R0104)
- T4 DNA Ligase (Roche, cat. no. 10799009001)
- Ligation buffer (Roche, cat. no. 11243292001)
- NaCl (Boom, cat. no. 76028327-5000)
- NEB buffer 2.1 (New England Biolabs, cat. no. B72025)
- NP-40 (Sigma-Aldrich, cat. no. 74385)
- Nucleomag 96 PCR Beads (P-Beads; Macherey Nagel, cat. no. 744100.34)
- Penicillin–streptomycin (Thermo Fisher Scientific, cat. no. 15070063)
- PhireII polymerase (Thermo Fisher Scientific, cat. no. F122) ▲ **CRITICAL** The type of polymerase used can substantially affect the resulting product; we recommend using this particular product exclusively.
- Proteinase K (Roche, cat. no. 3115801001)
- Q5 High Fidelity DNA Polymerase (New England Biolabs, cat. no. M0491)
- Qubit DNA Broad Range (Invitrogen, cat. no. Q32853)
- Qubit RNA Broad Range (Invitrogen, cat. no. Q10211) ▲ **CRITICAL** Take care to avoid RNase contamination.
- RPMI 1640 cell culture medium (Gibco, cat. no. 12633012) ▲ **CRITICAL** Use in a sterile environment, and take care to avoid cross-contamination.

- SDS (MP Biomedicals, cat. no. 811030) **▲CRITICAL** When SDS is dissolved in water, it precipitates when cooled to 4 °C.
- T7 Megashortscript Kit (Invitrogen, cat. no. AM1354) **▲CRITICAL** NTPs do not tolerate freeze-thawing well; we recommend separately storing pre-mixed NTPs at −20 °C in 8-μl aliquots, containing 2 μl of each NTP. Take care to avoid RNase contamination. Such aliquots can be stored safely for several months.
- Tris (pH 7.5; Invitrogen, cat. no. 15504-020)
- Triton X-100 (Sigma-Aldrich, cat. no. T8787)
- Universal gRNA reverse primer (AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTATTTCTAGCTCTAAAAC; Integrated DNA Technologies) **▲CRITICAL** A 100-μM long-term stock is stored at −20 °C in TE buffer (pH 7.5) and is stable for at least several years. A 5-μM working stock is prepared by diluting an aliquot of the long-term stock in 10 mM Tris pH 7.5 and can be stored at −20 °C for several weeks **▲CRITICAL** Take care to avoid RNase contamination.
- Liquid nitrogen
- TAE
- TBE
- TURBO DNase
- MilliQ water

Equipment

- Equipment for performing automated size selection (Pippin HT Sage Science, cat. no. HTP0001 with 1.5% agarose, PippinHT, 300–1,500 bp, 10 cassettes, Sage Science, cat. no. HTC1510)
- DNA quantifier (Invitrogen Qubit 4.0, cat. no. Q33226)
- Eppendorf centrifuge (Eppendorf 5424R, cat. no. 5404000010)
- Heat block (Eppendorf ThermoMixer Compact/F1.5, cat. no. 5384000012)
- Nanopore sequencing (Oxford Nanopore MinION or GridION in combination with R9.4.1 flowcells, can be used interchangeably)
- PCR machine (Bio-Rad S1000, cat. no. 1852196)
- qPCR machine (Bio-Rad CFX connect, cat. no. 1855200)
- Centrifuge with a swinging bucket rotor (Eppendorf, model no. 5810R)
- Water bath (Grand JBAqua 5)
- 40-μm cell strainer (Greiner Bio-One, cat. no. 542040)
- Hemocytometer or cell counting system such as a CASY cell counter (OLS OMNI Life Science, cat. no. 5651736)
- Roller
- Paper tissues
- Falcon tube
- Magnetic separator
- Magnet
- Magnet rack

Software

Computational tools

- A computer operating a Unix-like shell (e.g., Bash v3.2+) with at least a single core and about 20 GB of memory. Note that the actual memory requirement is related to the size of MC-4C dataset being analyzed.
- samtools v1.9+ (ref. ²⁴)
- bwa v0.7.17+ (ref. ²⁵)
- Python v2.7+
 - Python packages
 - h5py v2.7.1+ (ref. ²⁶)
 - numpy v1.13.3+ (ref. ²⁷)
 - pandas v0.23.4+ (ref. ²⁸)
 - pysam v0.15.1+ (<https://github.com/pysam-developers/pysam>)
 - matplotlib v2.1.2+ (ref. ²⁹) (only for producing quality check and analysis plots)

Procedure

Collect cells

- For the generation of high-quality template, it is essential to obtain a single-cell suspension, which will permit homogenous crosslinking. Isolate single cells from tissue (see option A for an example) or from cultured cells (see option B for an example). The quantity specified in this protocol is 2×10^7 cells.

? TROUBLESHOOTING

(A) Collecting mouse fetal liver cells ● Timing 1 h

▲ **CRITICAL** Any experiments involving live mice must conform to relevant institutional and national regulations.

- Collect embryos at E14.5, extract the dark red liver and remove any lighter-colored fatty tissue. Two livers usually yield enough material for a replicate of 20×10^7 cells, enough for the complete processing of two or three MC-4C viewpoints.
- Gather the livers from the entire litter in 10 ml of PBS and keep on ice.
- Use a P1000 pipette to dissociate livers by pipetting up and down. Place a 40- μ m cell strainer on a 50-ml Falcon tube and transfer the dissociated liver suspension onto the membrane. Pipette the livers up and down while the suspension is kept on the membrane. Single cells will pass through the membrane. If large chunks remain, use the plunger of a 5-ml syringe to press the chunks through the membrane.
- Spin down the single-cell suspension at 500g for 10 min at 4 °C in a swing-out centrifuge, discard the supernatant and resuspend the cell suspension in 5 ml of 10% (vol/vol) FBS–PBS. Count cells using a counting window or a cell-counter system and proceed with 2×10^7 cells (per biological replicate).

(B) Collecting K562 cells for MC-4C ● Timing 15 min

- Grow cells in suspension to a concentration of between 1×10^5 and 5×10^5 cells per ml of medium, count cells (using a Casycounter or similar cytometer or a counting window) and pool $\sim 2 \times 10^7$ cells into a 50-ml Falcon tube per biological replicate. 2×10^7 cells usually yield enough template for the analysis of two or three viewpoints.
- Centrifuge cells at 500g for 10 min at room temperature (20–23 °C) in a swing-out centrifuge, discard supernatant and resuspend cells in 5 ml of 10% (vol/vol) FBS–PBS.

Crosslinking ● Timing 2 h

- Crosslink the single-cell suspension by adding 5 ml of freshly prepared 4% (vol/vol) formaldehyde in 10% (vol/vol) FBS–PBS to the 5 ml of cell suspension from Step 1A(iv) or 1B(ii) (2% final formaldehyde concentration). Incubate for 10 min at room temperature on a roller.

▲ **CRITICAL STEP** Making a 4% (vol/vol) formaldehyde solution ensures an even distribution of formaldehyde upon mixing compared to adding 37% (vol/vol) formaldehyde directly to the cell suspension.

! **CAUTION** Formaldehyde is toxic; avoid skin contact, use in a fume hood and dispose of waste appropriately.

- Quench the crosslinking reaction by adding 1.425 ml of a 1 M glycine solution and cooling the samples to 4 °C by placing on ice. Spin down 10 min at 500g in a swing-out centrifuge at 4 °C.
- During the centrifugation, prepare lysis buffer according to the table below and place the tube on ice to chill it. Remove and discard the supernatant and resuspend the cells in 1 ml of cold and freshly prepared lysis buffer. Transfer cells to a 2-ml Eppendorf tube. Incubate for 15 min on ice.

▲ **CRITICAL STEP** For some cell types, longer incubation is required for digestion to be effective. This is likely to be due to a sturdier cell membrane. Incomplete lysis will become apparent at a later step (Step 13 or 16), if the digestion or ligation is not optimal. See ‘Troubleshooting’ section, if this is the case.

Component	Amount for 10 ml	Final concentration
1 M Tris-HCl pH 7.5	500 μ l	50 mM
5 M NaCl	300 μ l	150 mM
0.5 M EDTA	100 μ l	5 mM
Table continued		

(continued)

Component	Amount for 10 ml	Final concentration
20% (vol/vol) NP-40	250 μ l	0.5% (vol/vol)
10% (vol/vol) Triton X-100	100 μ l	0.1% (vol/vol)
MilliQ water	8.75 ml	—

- 5 Centrifuge for 10 min at 500g at 4 °C in a swing-out centrifuge, remove and discard the supernatant and resuspend cells in 500 μ l of 1.1 \times DpnII buffer.

▲ CRITICAL STEP It is possible to stuff a 15-ml Falcon tube with paper tissues to create a centrifuge adaptor for using 2-ml tubes in regular swing-out centrifuges. The stuffing should be sufficient to allow the bottom of the 2-ml tube to be supported by the stuffing and the tube to stand upright relative to the 15-ml tube, rather than hanging on the edge of the 15-ml tube when centrifugal force is applied. This allows the use of a regular swing-out centrifuge, which yields more convenient to resuspend pellets and limits cell loss as compared to fixed-rotor centrifuges.

■ PAUSE POINT Cells can be stored for several months after crosslinking and lysis. To do so, snap-freeze the cell pellet in liquid nitrogen and store at -80°C .

First digestion ● Timing 1 d

- 6 Spin down cells for 10 min at 500g at 4 °C in a swing-out centrifuge, remove and discard the supernatant and resuspend the pellet in 500 μ l of 1.1 \times DpnII buffer.
- 7 Warm up the sample to 37 °C in a heat block, and then add 15 μ l of 10% (vol/vol) SDS and incubate for 1 h at 37 °C while shaking at 600 r.p.m. (the shaking intensity required to prevent cell sedimentation varies by cell type).

▲ CRITICAL STEP Adding SDS to a cold sample will cause the SDS to precipitate.
- 8 Add 75 μ l of 20% (vol/vol) Triton X-100 and incubate for 1 h at 37 °C in a heat block while shaking at 600 r.p.m.
- 9 Take 10 μ l of the sample as undigested control (U) and store it at 4 °C. The sample can be stored safely for at least a week.
- 10 Add 200 U of DpnII restriction enzyme to the rest of the sample and incubate for ~3 h at 37 °C while shaking at 600 r.p.m.

▲ CRITICAL STEP We see improved performance if DpnII is added in two batches with a few hours in between. The indicated 3 h is intended as a guideline.
- 11 Add another 200 U of DpnII restriction enzyme and incubate overnight at 37 °C while shaking at 600 r.p.m.

Digestion control ● Timing 2 h

- 12 Take 10 μ l of the sample as digestion control (D) and store it at 4 °C after Step 13 for later reference. The sample can be stored safely for at least a week.
- 13 To both the undigested (Step 9) and digested (Step 12) controls, add 80 μ l of 10 mM Tris pH 7.5 and 10 μ l of 10 μ g/ μ l proteinase K, incubate for 1 h at 65 °C, then run 20 μ l from each sample on a 0.6% (wt/vol) TAE or TBE agarose gel stained with 0.5 μ g/ml ethidium bromide. We recommend also adding ethidium bromide (0.5 μ g/ml) to the running buffer to avoid destaining of the gel. In case of cells from species other than mouse, the digested sample should show a smear of DNA with very little or no material above the 10-kb range. In case of mouse cells, a distinct high-molecular-weight band always remains in the sample (owing to the absence of DpnII sites in the mouse pericentromeric repeats). See Fig. 3 for examples.

▲ CRITICAL STEP High-molecular-weight genomic DNA in the undigested control can be difficult to pipette and may rapidly diffuse out of the gel slot. In such cases, the digestion efficiency can be interpreted only by comparison to the marker: 10–12-kb marker bands run approximately at the same height as undigested genomic DNA.

? TROUBLESHOOTING

First ligation ● Timing 1 d

- 14 If digestion is as expected, inactivate the restriction enzyme by incubating the sample from Step 11 at 65 °C for 20 min while shaking at 600 r.p.m.

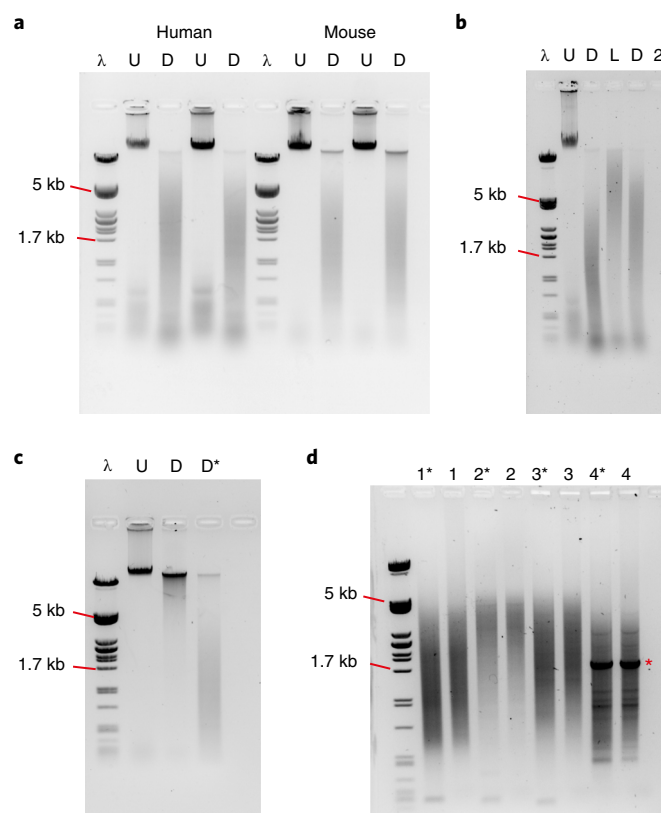


Fig. 3 | Intermediate controls in the MC-4C protocol. **a**, Examples of DpnII digestions (Step 13). λ marks λ -phage DNA digested with PstI restriction enzyme; U marks the lanes with undigested DNA, D marks DpnII digested DNA from K562 human cells on the left and mouse embryonic liver to the right. Note that DpnII-digested mouse DNA always contains a small amount of high-molecular-weight DNA, owing to the absence of a DpnII restriction site in the pericentromeric repeats. Samples were loaded on a 0.6% agarose gel. **b**, An example of all intermediate controls of an MC-4C template from K562 cells on a 0.6% agarose gel. λ indicates the ladder, U indicates undigested DNA (Step 13), D indicates the first digestion (DpnII) (Step 13), L indicates the ligation (Step 16) and D2 indicates the second digestion (HindIII) (Step 22). **c**, Example of a case in which the first digestion was not adequate (Step 13), as a large portion of DNA was still >10 kb in size (D). More DpnII enzyme was added, and the incubation was extended by 3 h. This resulted in properly digested DNA (D*). Samples were loaded on a 0.6% gel. **d**, Examples of MC-4C PCR products (Steps 60 and 62). These PCRs inevitably vary in size distribution but should contain material between 1.7 and 5 kb. Each number indicates a different primer set, numbers with an asterisk indicate PCR before AMPure size exclusion (Step 60) and lanes without an asterisk contain the final product (Step 62). Samples 1, 2 and 3 show successful experiments with a smear of product extending to the desired size range. Lane 4 is an example of an experiment in which the Cas9 digestion partially failed, resulting in a large band (indicated with a red asterisk). Samples were loaded on a 0.6% agarose gel. Note, animal experiments were performed in compliance with the Dutch law on animal experimentation in line with the European Directive (2010/63/EU). Cell line identity was confirmed, and cells in culture were periodically tested for mycoplasma contamination.

- 15 Add 1,200 μ l of MilliQ and 200 μ l of 10 \times ligation buffer, mix well and then add 10 μ l of T4 DNA ligase (10 U/ μ l). Incubate overnight at 16 $^{\circ}$ C.

▲ CRITICAL STEP We achieve this temperature by placing a water bath in a room at 4 $^{\circ}$ C and setting the heat to 16 $^{\circ}$ C.

Ligation control ● Timing 2 h

- 16 Take 30 μ l from the sample from Step 15 as a ligation control (L) and add to it 60 μ l of 10 mM Tris pH 7.5 and 10 μ l of proteinase K (10 μ g/ μ l), incubate for 1 h at 65 $^{\circ}$ C while shaking at 600 r.p.m. and run the 20 μ l of the ligation control on a 0.6% (wt/vol) agarose gel stained with 0.5 μ g/ml ethidium bromide alongside the 20 μ l previously obtained digestion control from Step 12 for comparison. We recommend adding ethidium bromide (0.05 μ g/ μ l) to the running buffer to avoid destaining of the gel. A clear upwards shift in molecular weight should be visible in the 'L' sample.

See Fig. 3b for expected results.

? TROUBLESHOOTING

DNA isolation ● Timing 1 d and 3 h

- 17 If the ligation was successful, add 30 μ l of 10 μ g/ μ l proteinase K to the remaining cell suspension from Step 15 and incubate overnight at 65 °C while shaking at 600 r.p.m.
- 18 Split the DNA solution over two 2-ml tubes and add 1 ml of 2-propanol to each tube, mix well until the precipitating DNA becomes visible and incubate for 30 min at –20 °C. Spin down at 12,000g for 10 min in a tabletop centrifuge at 4 °C to pellet the DNA.
- 19 remove and discard the supernatant and add 1 ml of 80% (vol/vol) ethanol to the sample without resuspending the pellet. Spin down for 5 min at 12,000g at 4 °C to pellet the DNA. Perform this step two times.
- 20 Remove as much ethanol as possible without disturbing the pellet and then air-dry the pellet for 5–10 min.
- ! CAUTION Take care not to overdry the DNA pellet.
- 21 If you plan to proceed immediately with the second digestion, resuspend each pellet in 180 μ l of MilliQ water, incubate for 10 min at 50 °C and pool both samples into one 360- μ l sample. Otherwise, resuspend each pellet in 180 μ l of TE pH 7.5 buffer instead, pool both samples and store at –20 °C.

■ PAUSE POINT DNA template can be stored at least several months if the pellet is resuspended in TE (pH 7.5) buffer and stored at –20 °C

Second digestion ● Timing 4 h

- 22 Add 40 μ l of 10 \times restriction buffer (NEB 2.1) and 50 U of HindIII to the sample. Proceed in the same way if stored samples are thawed for further processing. Incubate for 3 h or more at 37 °C, then run 5 μ l of the sample (called ‘D2’) on a 0.6% (wt/vol) agarose gel stained with ethidium bromide (0.5 μ g/ml) to check the digestion. We recommend adding ethidium bromide (0.5 μ g/ml) to the running buffer to avoid destaining of the gel. See Fig. 3b for an example.
- 23 If the digestion is satisfactory (a clear downward shift of the smear is observable), heat-inactivate the enzyme by incubating the remaining sample at 65 °C for 20 min. Otherwise, add 50 more units of HindIII and incubate for a further 3 h before re-checking the digestion.

Circularization ● Timing 1 d and 3 h

- 24 Use a Qubit spectrometer to quantify the amount of template. We routinely obtain between 60 and 90 μ g template at this stage.
- 25 Set up a ligation reaction as outlined in the example below; for example, to 50 μ g template, add the following reagents and incubate overnight at 16 °C.

▲ CRITICAL STEP Perform Steps 26–32 in a 50-ml Falcon tube.

Reagent	Amount	Final concentration
131.5 ng/ μ l template (Step 23)	50 ng 380 μ l	5 ng/ μ l
10 \times ligation buffer	1 ml	1 \times
MilliQ water	8.6 ml	—
T4 DNA ligase (5 U/ μ l)	20 μ l	0.01 U/ μ l
Final	10 ml	—

- 26 To perform a Nucleomag PCR bead (p-bead) purification, first add 1 sample volume of 2-propanol to the ligated sample from Step 24 (add 10 ml of 2-propanol to the 1- ml example shown here).
- 27 Then add 1/100 sample volume of Nucleomag PCR beads (100 μ L beads for the 10-ml sample used here).

▲ CRITICAL STEP Make sure that beads are properly resuspended before use.

- 28 Incubate for at least 30 min at room temperature on a tube roller.
- 29 Put the sample on a magnetic separator until the solution becomes clear.
- 30 Remove and discard the supernatant, add 20 ml of 80% (vol/vol) ethanol, resuspend beads and put the tube back on the magnet.

- 31 Wait until all beads are cleared out of the solution, remove and discard the supernatant, and wash again with 80% (vol/vol) ethanol by resuspending the beads thoroughly and then putting the tube back on the magnet.
- 32 Wait until all beads are cleared out of the solution and remove and discard all but ~4 ml of the supernatant while keeping the tube on the magnet.
- 33 Remove the tube from the magnet, resuspend the beads in the remaining supernatant by pipetting, transfer 2 ml of ethanol/bead suspension to a 2-ml Eppendorf tube and place the 2-ml Eppendorf tube on a magnetic separator.
- 34 Remove and discard the ethanol when the solution becomes clear.
- 35 Add the remaining 2 ml of bead solution to the 2-ml tube and wait until the solution clears again.
- 36 Take 1 ml of the cleared supernatant, use it to rinse the remaining beads out of the 50-ml Falcon tube and then transfer the remaining beads to the 2-ml tube on the magnet rack. (Repeat this step if beads still remain in the Falcon tube.)
- 37 Discard the residual ethanol once the solution clears.
- 38 Briefly spin down the beads on a tabletop centrifuge or centrifuge them at 1,000g for 10 s at room temperature in a microcentrifuge. Transfer the tubes back to the magnet and remove and discard the residual ethanol. Briefly (~2 min) air-dry the beads (avoid over-drying).
- 39 Resuspend the beads in 500 μ l of 10 mM Tris pH 7.5 prepared in RNase-free water and incubate for 10 min at 50 °C while shaking at 600 r.p.m. on a heat block.
- 40 Place the tube back on the magnet rack for 5 min and transfer the supernatant to a new 1.5-ml Eppendorf tube.
- 41 When any residual beads become visible on the side of the tube when aspirating the supernatant, transfer all but the last 5 μ l of the supernatant and repeat Step 40 until no more beads are visible when removing the sample.
- 42 Quantify the template using a Qubit spectrophotometer and the DNA Broad Range Kit. We expect a total yield of between 40 and 60 μ g when starting with 20×10^7 cells.
■ **PAUSE POINT** The 4C template can be stored at –20 °C for several months.

In vitro Cas9 digestion ● Timing 4 h

▲ **CRITICAL** The Cas9 cleavage and inverse PCR steps (Steps 43–64) are viewpoint specific. Template processed in this way can no longer be used for alternative viewpoints. Only proceed with as much template as is needed for a viewpoint; usually, 20 μ g of 4C template (as quantified in Step 42) yields enough material to complete the protocol for one MC-4C viewpoint. Starting with 20×10^7 cells, it is usually possible to process two or three MC-4C viewpoints.

▲ **CRITICAL** We highly recommend testing gRNAs on a small scale on genomic DNA before use on MC-4C template. See Box 1 for instructions on preparing and testing gRNAs.

▲ **CRITICAL** Avoid RNase contamination from this point onwards.

- 43 For each of the three gRNAs, prepare a Cas9–gRNA pre-incubation according to the following recipe (for 20 μ g of template) and incubate for 30 min at room temperature:

Ingredient	Quantity (μ l)	Final concentration
10× NEB 3.1	10	1×
gRNA (100 ng/ μ l)	10	1 ng/ μ l
Cas9 enzyme (1 μ M)	5	50 nM
DEPC	75	—
Total	100	—

- 44 Add 10× NEB 3.1 buffer to the MC-4C template to be digested (from Step 42) to obtain a 1× buffer concentration (e.g., to 500 μ l of solution containing 20 μ g template, add 55 μ l of 10× NEB3.1).
▲ **CRITICAL STEP** When an efficiency test for the Cas9 digestion is to be performed, some (~500 ng) template should be left undigested as a control.
- 45 Add each of the three gRNA pre-incubation mixtures from Step 41 to the MC-4C template and incubate at 37 °C for 3 h.
- 46 Inactivate Cas9 by adding 1/25 volume of 10% (vol/vol) SDS and incubating at 70 °C for 10 min.

0.6× AMPure XP cleanup ● Timing 20 min

- 47 Add 0.6× the total volume of the sample of AMPure XP beads to the sample, mix thoroughly by vortexing and incubate for 5 min at room temperature.
- 48 Place on a magnetic separator and wait until the solution becomes clear.
- 49 Remove and discard the supernatant and add 500 µl of 80% (vol/vol) ethanol. Do not remove the tube from the magnet and do not mix the contents of the sample.
- 50 Wait until the solution becomes clear.
- 51 Repeat Steps 49 and 50.
- 52 Remove and discard the supernatant and briefly centrifuge the beads in a tabletop centrifuge (10 s at 1,000g at room temperature), then place the tube back on the magnet rack.
- 53 When the beads have moved from the bottom of the tube, remove and discard the residual ethanol.
- 54 Resuspend the beads in 500 µl of 10 mM Tris pH 7.5 and incubate for 10 min at room temperature while shaking at 600 r.p.m.
- ▲ **CRITICAL STEP** We recommend elution at 50 °C when the yield is critical (i.e., with precious template).
- 55 Briefly centrifuge the bead solution (10 s at 1,000g at room temperature) and place the tube back on the magnetic rack.
- 56 When the solution becomes clear, take the supernatant and place it in a new 1.5-ml tube.
- 57 (Optional) Check Cas9 digestion efficiency by qPCR, quantify the template using a Qubit spectrophotometer and perform a qPCR on 30 ng template. Compare the C_T values between primer sets flanking the Cas9 cut site and a fourth, unrelated primer set (that does not span a DpnII or HindIII restriction site), for Cas9-digested and Cas9-undigested template.

? TROUBLESHOOTING

- **PAUSE POINT** The MC-4C template can be stored for several months at −20 °C

Inverse PCR ● Timing 4 h

- 58 *Viewpoint-specific PCR*. Assemble 96 identical viewpoint-specific PCR reactions using PhireII hotstart polymerase. Add the viewpoint-specific primers designed according to the guidelines in the ‘Experimental design’ section. We usually make a pre-mix containing 5 µM forward primer and 5 µM reverse primer. For example, we used the following viewpoint primers for the mouse β-globin HS2 viewpoint and the human ‘CTCF-site K’ viewpoints:

β-globin HS2 FW	5′-CAGATGTTTTAGCTGTGACTGAT-3′
β-globin HS2 RV	5′-CTTGGACAGTGGTACTGCAATAATT-3′
CTCF-K FW	5′-AGCTGGACATTCTCAACTGC-3′
CTCF-K RV	5′-GACATGACGTTTGGCTCCATG-3′

Reagent	Amount (1×)	Amount (96×)
5× Phire buffer	5 µl	480 µl
10 mM dNTP	0.5 µl	48 µl
5 µM primer mix	2.5 µl	240 µl
Phire II hotstart polymerase	0.5 µl	48 µl
Template (Step 54)	100 ng	9.6 µg
MilliQ water	Up to 25 µl	Up to 2.4 ml
Total	25 µl	2.4 ml

- 59 Run the following program:

Cycle number	Denature (98 °C)	Anneal (57 °C)	Extend (72 °C)
1	20 s	—	—
2–31	10 s	20 s	1 min 30 s
32			5 min

- 60 Pool all 96 PCR reactions (2.4 ml), mix well and take a 500- μ l aliquot for size selection and sequencing.

Size selection and sequencing ● Timing 2 d

- 61 *Purification.* Perform a 0.6 \times AMPure XP bead purification on the 500- μ l aliquot as described in Steps 47–56. Elute in 75 μ l of TE pH 7.5.
- 62 Analyze 20 μ l of the purified PCR product on a 0.6% (wt/vol) agarose gel stained with 0.5 μ g/ml ethidium bromide. The result should be a smear that includes fragments above 1.5 kb and that is not dominated by strong ‘bands’ (Fig. 3d). We recommend also adding 0.5 μ g/ml ethidium bromide to the running buffer to avoid destaining of the gel.

? TROUBLESHOOTING

- 63 Perform a pippin size selection of the sample from Step 59 according to the manufacturer’s instructions (<https://www.cultek.com/media/wysiwyg/manuales/SAGE/PippinHT-Operations-Manual.pdf>). We routinely use the Pippin HT HTC1510 machine with a size-selection range between 1.5 and 8 kb on a 1.5% cassette. The input DNA can be spread over multiple ‘lanes’.
- 64 Perform a Nanopore library prep according to the manufacturer’s instructions for the 1D genomic DNA by ligation protocol (SQK-LSK109; https://community.nanoporetech.com/protocols/1d-lambda-control-sqk-lsk109/checklist_example.pdf) but (i) omit shearing the DNA, and (ii) adjust the volume of the DNA repair and end-repair steps for a higher input quantity when multiple pippin lanes are used (double the volume if two lanes are used).

Walkthrough of MC-4C pipeline ● Timing 15 min for the (downscaled) test dataset provided in Supplementary Data 2; 3 h for a real MC-4C dataset

▲ **CRITICAL** In this section, we provide a step-by-step description of how the MC-4C pipeline can be used to process sequenced reads in an MC-4C experiment. In this walkthrough, we assume that the user is using the Linux operating system. Minor modifications might be required in the commands used to follow this walkthrough using Mac OS X. Commands starting with ‘\$’ are executed in the command line.

- 65 *Setting up the pipeline.* Confirm that the required Python packages for MC-4C pipeline are installed:

```
$ pip install h5py numpy pandas pysam matplotlib
```

- 66 Download the latest version of the MC-4C pipeline from its git repository using:

```
$ wget https://github.com/deLaatLab/mc4c_py/archive/master.zip
$ unzip master.zip
$ cd./mc4c_py-master
```

- 67 Make an index for the reference genome of interest (mm9 in this example) using bwa:

```
$ bwa index ~/references/mm9.fa
```

- 68 *Preparing input data.* For demonstration and testing purposes, we prepared a downsampled MC-4C dataset. This test dataset can be downloaded from the following link. Download and then decompress it to a preferred location. This dataset consists of a small sequencing file (i.e., fq_BMaj-test.fastq.gz) and a corresponding configuration file (cfg_BMaj-test.cfg) that holds experiment-specific details of this experiment. After downloading the aforementioned test data, create a folder named ‘fastqs’ and move the obtained sequencing file (i.e., fq_BMaj-test.fastq.gz) to this folder:

```
$ mkdir -p./fastqs
$ mv ~/Downloads/fq_BMaj-test.fastq.gz./fastqs/
```

- 69 Create a folder named ‘configs’ and move the obtained configuration file (i.e., cfg_BMaj-test.cfg) to this folder:

```
$ mkdir -p./configs
$ mv ~/Downloads/cfg_BMaj-test.cfg./configs/
```


- 70 Define the paths for the reference genome, bwa and the reference index (made in Step 66) in the corresponding configuration file (i.e., `cfg_BMaj-test.cfg`). These paths follow the standard format of paths in the Linux operating system. For example, in our system these paths are defined as follows:

```
bwa /bin/bwa
bwa_index /bwa_index/mm9.fa
reference_fasta /reference_genomes/mm9.fa
```

- 71 Confirm that these paths are correctly defined by:

```
$ grep -E 'bwa|bwa_index|reference_fasta' ./configs/cfg_BMaj-test.cfg
```

▲ **CRITICAL STEP** In the MC-4C pipeline, reads and fragments that originate from these reads are assigned a unique identifier to facilitate their tracing throughout the pipeline. This is the first step in processing the reads produced in an MC-4C library.

- 72 *Assigning read identifiers to sequenced reads.* Assign a unique identifier to each sequenced read using the `setReadIds` module:

`$ python./mc4c.py setReadIds BMaj-test`. This module will produce the following output in the terminal:

```
%% Assigning traceable identifiers to reads...
Writing reads with traceable identifiers to:./reads/rd_BMaj-test.
fasta.gz
Total of [1] files are given as input.
1. Reading from:./fastqs/fq_BMaj-test.fastq.gz
processed 5,000 reads.
[i] Read identifier conversion is completed successfully.
```

▲ **CRITICAL STEP** Each sequenced read in an MC-4C experiment consists of multiple interacting fragments. To assist the aligner in recognizing and mapping these fragments to the reference genome, we opted to pre-split sequenced reads into fragments based on the restriction enzyme recognition sequence.

- 73 *Splitting reads into fragments.* Apply the `splitReads` module by executing the following command:

```
$ python./mc4c.py splitReads BMaj-test
```

This module will produce the following output in the terminal:

```
%% Splitting reads into fragments...
Reading reads from:./reads/rd_BMaj-test.fasta.gz
Writing fragments to:./fragments/frg_BMaj-test.fasta.gz
processed 5,000 reads and produced 36,165 fragments.
[i] Total of 5,000 reads and 36,166 fragments are produced successfully.
```

- 74 *Mapping fragments to the reference genome.* Map the produced fragments to the reference genome using the `mapFragments` module:

```
$ python./mc4c.py mapFragments BMaj-test
```

The pipeline will produce the following output in the terminal:

```
%% Mapping fragments to genome...
Reading fragments from:./fragments/frg_BMaj-test.fasta.gz
Writing mapped fragments to:./bams/bam_BMaj-test.bam
Running: ~/bin/bwa-0.7.17/bwa bwasw -b 5 -q 2 -r 1 -z 5 -T 15 -t 1 ~/
references/mm9.fa./fragments/frg_BMaj-test.fasta.gz | samtools view
-q 1 -hbS - >./bams/bam_BMaj-test.bam
[i] Fragments are mapped to genome successfully.
```

▲ CRITICAL STEP Owing to read errors of long-read sequencing, mapped fragment ends may not map over the restriction sites exactly. To facilitate downstream analyses, the mapped coordinates fragments are computationally extended to the nearest restriction site in the reference genome. In addition, to reduce extra read splits performed (due to erroneous restriction sites introduced in the read), adjacent fragments within a read that also map adjacently (<20 bp) in the genome are merged to produce a single fragment.

- 75 *Making an MC-4C dataset from mapped fragments.* Apply the **makeDataset** module:

```
$ python./mc4c.py makeDataset BMaj-test
```

Once called, the following output will be shown in the terminal:

```
% Creating an MC-4C dataset from mapped fragments...
Reading fragments from:./bams/bam_BMaj-test.bam
Database of restriction enzyme cut sites is not found. Scanning the
reference genome to create this database...
Searching in reference:~/references/mm9.fa
Scanning: chr1, chr10, chr11, chr12, chr13, chr14, chr15, chr16, chr17,
chr18, chr19, chr2, chr3, chr4, chr5, chr6, chr7, chr8, chr9, chrM, chrX,
chrY,
Writing processed fragments to a temporary file first:./datasets/
mc4c_BMaj-test_all.hdf5.tmp
processed 0 fragments in 0 reads.
processed 10,000 fragments in 1,640 reads.
processed 20,000 fragments in 3,311 reads.
Loading the temporary file:./datasets/mc4c_BMaj-test_all.hdf5.tmp
Writing dataset to:./datasets/mc4c_BMaj-test_all.hdf5
Removing temporary file:./datasets/mc4c_BMaj-test_all.hdf5.tmp
[i] MC4C dataset is created successfully.
```

▲ CRITICAL STEP Considering the presence of random fragment ligations in the MC-4C reads, it is extremely unlikely that an identical *trans* (or *far-cis*) fragment(s) will ligate to the viewpoint fragment. Therefore, the captured *trans* (and *far-cis*) fragments in the MC-4C reads can be used as a genomic unique molecular identifier (UMI) to identify PCR duplicates.

- 76 *Remove read duplicates.* To remove PCR duplicate reads in an MC-4C dataset, the **removeDuplicates** module can be called using:

```
$ python./mc4c.py removeDuplicates BMaj-test
```

This module will produce the following output in the terminal:

```
% Removing pcr duplicates from an MC-4C dataset...
Reading MC4C dataset from:./datasets/mc4c_BMaj-test_all.hdf5
Writing unique MC4C reads to:./datasets/mc4c_BMaj-test_uniq.hdf5
There are 4990 reads in the dataset.
Selected non-overlapping fragments with MQ >= 20: 4960 reads are left.
Selected reads #cis fragment > 1: 697 reads are selected.
Selected reads with #trans fragment > 0: 500 reads are selected.
Scanning for duplicated trans-fragments:
scanned 0 trans-fragments, 500 reads are still unique.
scanned 500 trans-fragments, 476 reads are still unique.
Result statistics (before --> after filtering):
#reads: 4,990 --> 488
#fragments: 20,843 --> 987
Writing dataset to:./datasets/mc4c_BMaj-test_uniq.hdf5
[i] PCR duplicates are removed from MC4C dataset successfully.
```

This is the final pre-processing stage in the MC-4C pipeline. In the current walkthrough, the resulting processed (and duplicate-removed) dataset should hold ~488 unique reads capturing 987 fragments.

- 77 (Optional) The user can use the commands implemented in the **QC** module of MC-4C pipeline to assess basic quality of produced MC-4C datasets. For example, the user can check the read-size distribution of the produced MC-4C dataset (i.e., 'BMaj-test') using the following command:

```
$ python./mc4c.py QC readSizeDist BMaj-test.
```

? TROUBLESHOOTING

Troubleshooting

Pre-sequencing troubleshooting

Step 1: Difficulties obtaining single-cell suspensions. The ability to obtain a single-cell suspension is crucial for homogenous crosslinking but differs among cell and tissue types. In some cases, more stringent treatment is needed—for example, by using tissue homogenizers, collagenase treatment or cell-straining devices. Spotting a small aliquot of the sample on a microscope slide and viewing the cells under a standard microscope can give an impression of the suspension homogeneity.

Step 13: Poor digestion efficiency. Digestion efficiency varies among cell types and may be inadequate using the standard procedure described here. The first course of action in such cases is to add more restriction enzyme (usually 100 U DpnII) and to incubate at 37 °C for 3 h. Repeat the digestion control step and compare the result with the earlier digestion control (Fig. 3c). If no improvement is visible, the problem is likely to be due to suboptimal permeabilization of nuclei. In such cases, it may help to re-collect the cells by centrifugation and repeat Steps 4–13 of the template preparation protocol. Extension of the lysis incubation time (Step 4) can also improve digestion efficiency, if incomplete digestions are a recurring problem with a particular cell type.

Step 16: Poor ligation efficiency. If ligation fails to show a clear upwards shift of the template on an agarose gel, the troubleshooting is similar to that of poor digestions: the first course of action is to add 40 U of ligase enzyme and fresh ATP to a final concentration of 1 mM and re-incubate the reaction at 16 °C overnight. If this does not improve ligation efficiency, repeat Steps 2–8 of the template preparation protocol to improve cell permeability, and then repeat Steps 15 and 16.

Step 57, Box 1 and step 11: Poor Cas9 digestion efficiency. In some cases, gRNA-targeted Cas9 fails to digest the target DNA. For this reason, we recommend first testing gRNAs on a small scale on genomic DNA. If gRNAs do not perform well, we recommend designing a different gRNA on the same target region. In cases in which gRNAs perform well on gDNA but not on the MC-4C template, we have seen improved performance when the MC-4C template was purified using a Qiagen PCR Purification Kit. However, this lowers the template yield. In this case, split the template over multiple Eppendorf tubes and divide 10 µg template per tube. Purify each aliquot with a different column so as not to exceed the maximum loading capacity of the columns.

Step 62: Unexpected PCR results. It is essential that the MC-4C PCR yields PCR products >1,500 bp (Fig. 3d) when analyzed on gel in Step 62. If this is not the case, then further PCR optimization is required. Redesigning primers and adjusting the annealing temperature may make a substantial improvement in large product yield, but in some cases shifting the viewpoint to a neighboring fragment is necessary. In other cases, the MC-4C template may still contain smaller fragments that are favored in the PCR reaction. These can be excluded to some extent by re-doing the 0.6× AMPure XP purification (as described in Steps 47–56). If strong bands are still visible in the PCR smear, as seen in the example in Fig. 3d, this indicates that one or multiple gRNAs did not work. If the bands are below the 1.5-kb size-selection threshold, they are excluded at a later stage by the pippin size selection, so if the smear does have a substantial amount of material >1.5 kb, then it can still be sequenced. However, if enough template is available, we recommend troubleshooting the gRNA performance (see Troubleshooting, Step 57, Box 1 and step 11 'Poor Cas9 digestion efficiency').

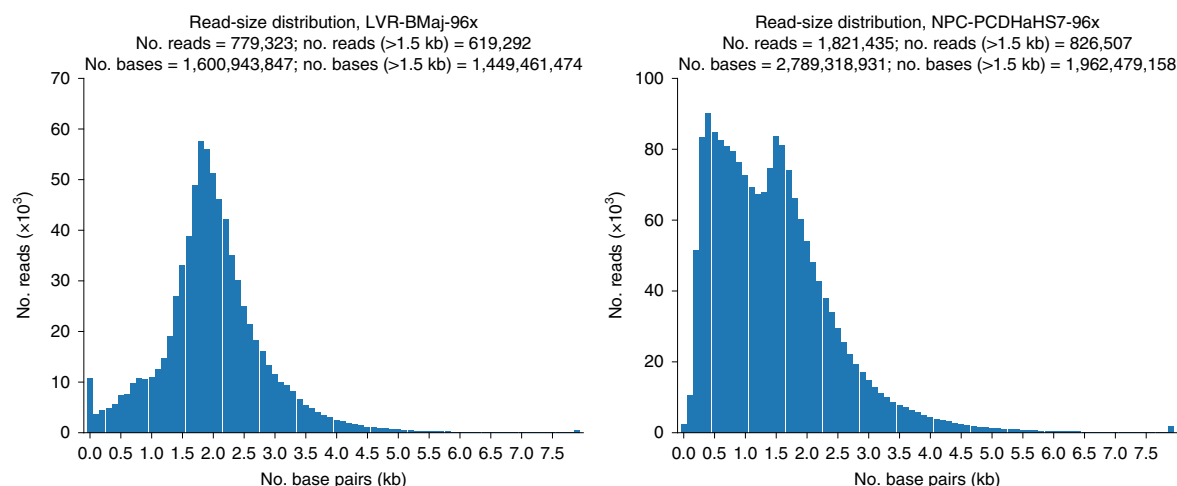


Fig. 4 | Read-size distribution of an MC-4C experiment. These plots show read-size distribution of reads in two MC-4C datasets produced from mouse liver cells (left) and neural precursor cells (right). The size distribution of raw reads should approximate a normal distribution of sizes, in which most reads are 1.5–8 kb in length. **a**, Read-size distribution of reads with a successful size selection. **b**, Read-size distribution of reads with poor size selection. This plot indicates that many small reads are sequenced. Note, animal experiments were performed in compliance with the Dutch law on animal experimentation in line with the European Directive (2010/63/EU).

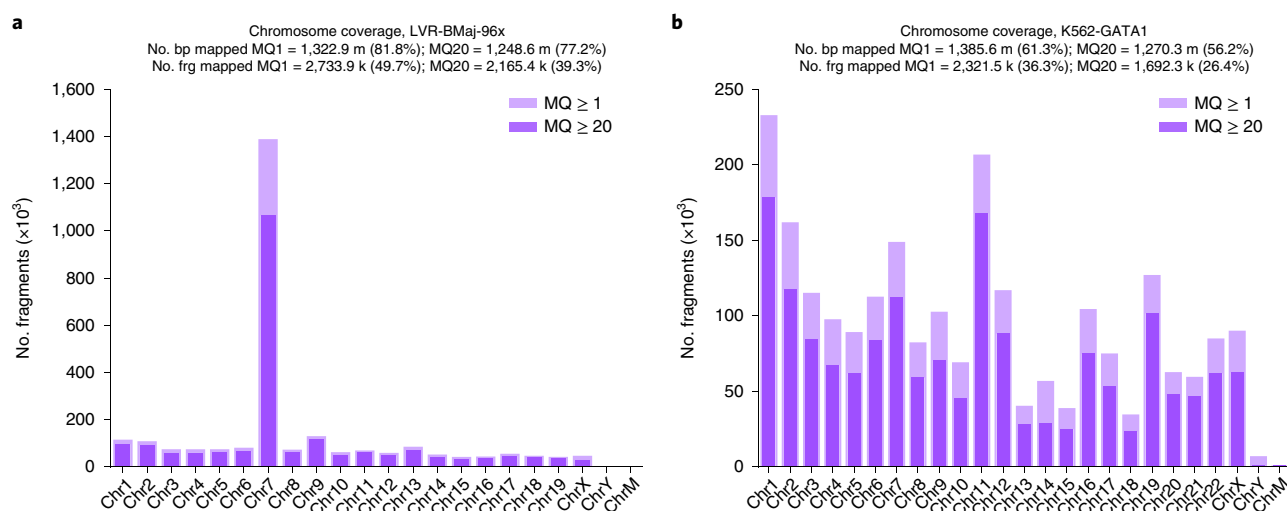


Fig. 5 | Chromosome coverage of an MC-4C experiment. These plots show coverage of mapped fragments over chromosomes in the genome in two MC-4C datasets produced from mouse liver cells (left) and K562 cell line (right). A single chromosome should show a substantial coverage, whereas others are minimally covered. **a**, Chromosome 7 (Chr7), the viewpoint chromosome, shows high coverage compared to other chromosomes. **b**, Coverage in viewpoint chromosome (ChrX) is not substantially different to other chromosomes, indicative of non-specific primers. The number of bases mapped with a mapping quality of at least 1 and of at least 20, and on the bottom line the number of fragments mapped with a mapping quality of at least 1 or at least 20. Note, animal experiments were performed in compliance with the Dutch law on animal experimentation in line with the European Directive (2010/63/EU).

Post-sequencing troubleshooting

Step 77: Poor size selection. In the MC-4C protocol, a 1.5–8 kb interval is used for selecting MC-4C PCR product before sequencing. However, after sequencing, we sometimes observed shorter reads. This could be caused by a poor size-selection procedure (Step 62). One can assess the efficiency of size selection by inspecting the read-size distribution of sequenced library using the **QC** module (see ‘Data quality control’ section in ‘Anticipated results’ for details). An efficient size selection should result in a normal distribution of reads sizes in the 1.5–8 kb interval. An example of such a distribution is given in Fig. 4a. Improper size selection can result in a library that still contains many short reads. An experiment in which short reads were unexpectedly sequenced is shown in Fig. 4b. Running the MC-4C pipeline on a sample with poor size selection will result in excessive read filtering, as a high

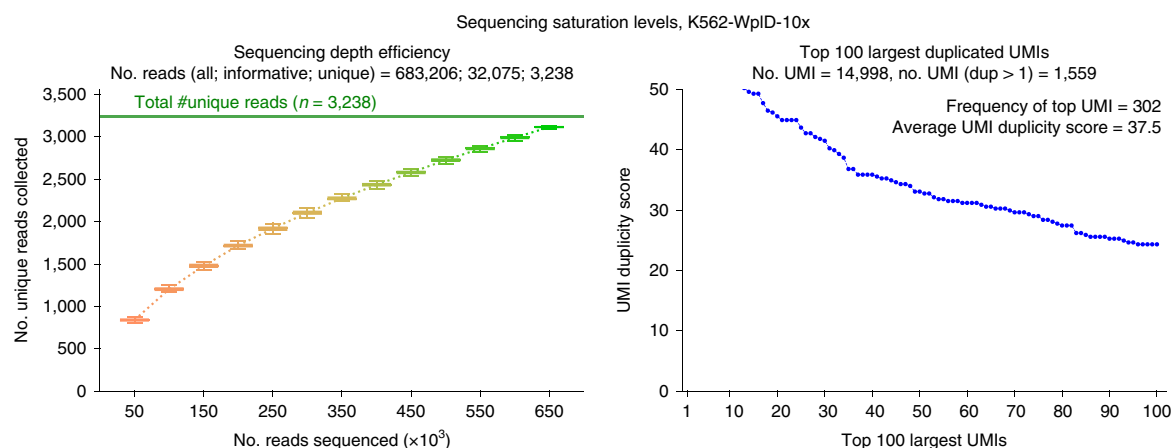


Fig. 6 | Sequence saturation plots. Example of an experiment in which MC-4C was performed using K562 cells. Frequency of observed UMIs shows that few UMIs are frequently duplicated. In this plot, the average duplicity score (shown on the top right) is 37.5, which indicates that diversity of amplified reads in the given experiment is relatively low.

Table 3 | Commands available in the QC module of the MC-4C pipeline

Command name	Function
readSizeDist	Plots size distribution of sequenced reads in terms of the number of base pairs sequenced
cirSizeDist	Plots circle-size distribution in terms of the number of fragments
chrCvg	Plots chromosome coverage of mapped fragments
overallProfile	Plots the overall profile of an MC-4C experiment
seqSaturation	Produces statistics to assess sequence saturation and UMI diversity

percentage of reads will contain <3 fragments. However, the remainder of reads can still be analyzed as usual.

Step 77: Off-target enrichment of MC-4C primers. The specificity of primers can be assessed by inspecting the chromosome coverage plot. Such a plot can be produced by calling ‘chrCvg’ that is implemented in the QC module:

```
$ python./mc4c.py QC chrCvg BMaj-test
```

An example of such a plot is provided in Fig. 5. When a high percentage (e.g., >30%) of reads is mapped to non-viewpoint chromosomes, we recommend redesigning the primers used to amplify (done in Step 52) to either shift the coordinates to different positions on the same viewpoint fragment or design primers for a different fragment close to the intended viewpoint.

Step 77: High UMI duplicity. A high UMI duplicity ratio for an experiment (i.e., >20; see Fig. 6) indicates a low library complexity. This may be caused by poor PCR performance, which can have several underlying causes. First, the template may not be properly circularized, which can potentially be mitigated by increasing the amount of MC-4C template per reaction. Second, PCR performance may be poor because of primer artefacts (self or hetero complementarity, secondary structures). We first recommend designing new primers (see PCR optimization under ‘Pre-sequencing troubleshooting’).

Timing

Main protocol: 7 d
Steps 2–5, crosslinking: 2 h hands-on time
Steps 6–11, first digestion: 6 h and an overnight incubation, 1 h hands-on time
Steps 12 and 13, digestion control: 2 h, 1 h hands-on time
Steps 14 and 15, first ligation: overnight incubation, 15 min hands-on time
Step 16, ligation control: 2 h, 1 h hands-on time

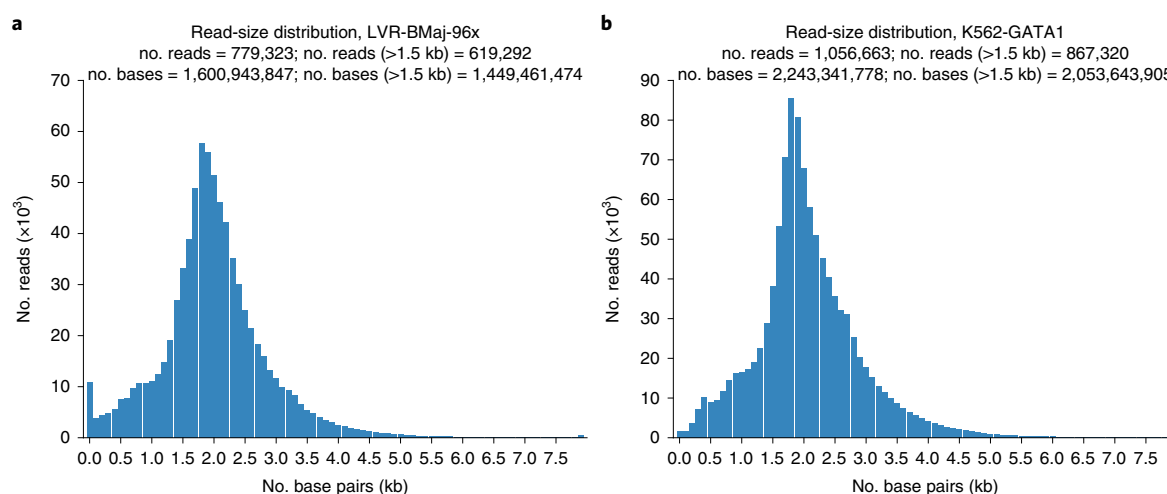


Fig. 7 | Read-size distribution of two MC-4C datasets. The top of each plot shows a series of characteristics for two MC-4C datasets produced from mouse liver cells (left) and K562 cell line (right). No. reads, total number of sequenced reads; no. reads (>1.5 kb), number of reads >1.5 kb; no. bases, total number of sequenced bases; no. bases (>1.5 kb), number of bases in reads over 1.5 kb. Note, animal experiments were performed in compliance with the Dutch law on animal experimentation in line with the European Directive (2010/63/EU).

Steps 17–21, reverse crosslinking and DNA isolation: overnight incubation and 3 h, 1 h hands-on time.

Steps 22 and 23, second digestion: 4 h, 30 min hands-on time.

Steps 24 and 25, second ligation: overnight incubation, 30 min hands-on time

Steps 26–42, template purification: 2 h, 1 h hands-on time

Steps 43–57, Cas9 cleavage: 4 h, 1 h hands-on time

Steps 58–62, PCR reaction and purification: 4 h, 1 h hands-on time

Steps 63 and 64, pippin size selection and sequencing: 10 h, 2 h hands-on time

gRNA preparation (2 d)

Box 1 (steps 1–5), synthesize gRNAs: 6 h, 1 h hands-on time

Box 1 (steps 6–10), set up Cas9 test digestion: 4 h, 1 h hands-on time

Box 1 (step 11), perform qPCR to confirm digestion: 4 h, 1 h hands-on time

Running MC-4C pipeline (3 h)

Step 71, **setReadIds** module: 30 min

Step 72, **splitReads** module: 45 min

Step 73, **mapFragments** module: 1 h

Step 74, **makeDataset** module: 15 min

Step 75, **removeDuplicates** module: 10 min

Anticipated results

Data quality control

The MC-4C pipeline implements comprehensive quality control functionalities in the **QC** module. It is recommended that the **QC** module is used to produce several summary reports that describe general statistics (e.g., read-size distribution) for an MC-4C experiment. The full list of commands available in this module is represented in Table 3.

These statistics help the user to assess the quality of an MC-4C experiment and also to identify the potential shortcomings in the library preparation that may have resulted in a suboptimal throughput in an MC-4C experiment.

Read-size distribution

A dataset with many short (i.e., <1.5 kb) reads is likely to have insufficient multi-contact complexity (i.e., <2 interactions, excluding the viewpoint). Therefore, it is recommended to initially check the size distribution of sequenced reads using the `readSizeDist` command in the **QC** module. Figure 7

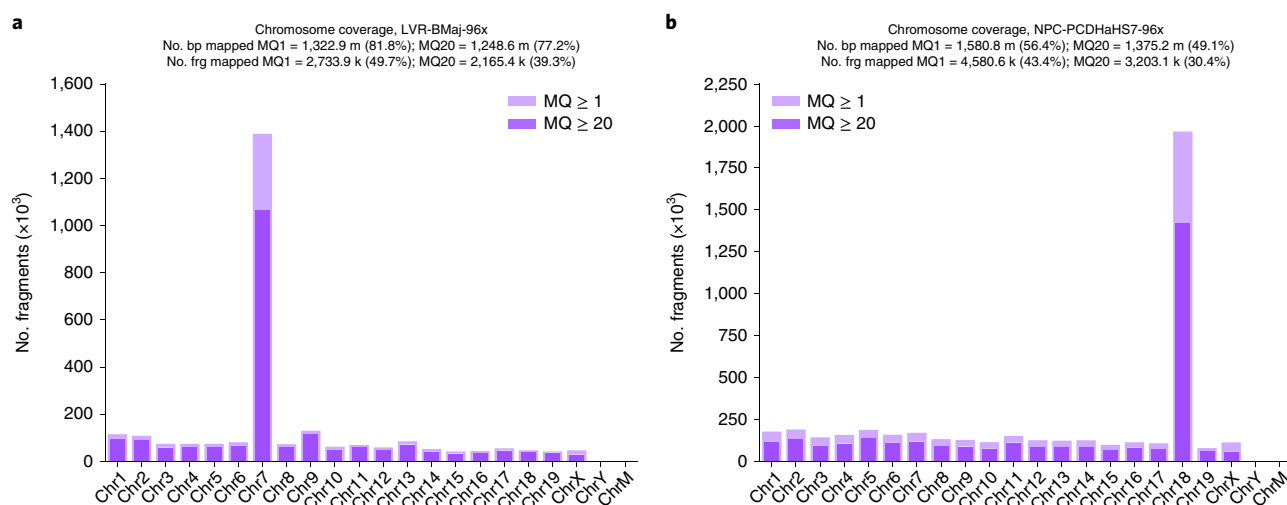


Fig. 8 | Chromosome coverage of two example MC-4C experiments. These plots show coverage of mapped fragments in two MC-4C datasets produced from mouse liver cells (left) and neural precursor cells (right). Light and dark purple bars represent the set of fragments mapped with mapping quality (MQ) of at least 1 and at least 20, respectively. The majority of fragments are mapped to the viewpoint chromosomes (Chr7 (**a**) and Chr18 (**b**)). Some characteristics are provided at the top of the graphs; the number of bases mapped with a mapping quality of at least 1 and of at least 20, and on the bottom line the number of fragments mapped with a mapping quality of at least 1 or at least 20. Note, animal experiments were performed in compliance with the Dutch law on animal experimentation in line with the European Directive (2010/63/EU).

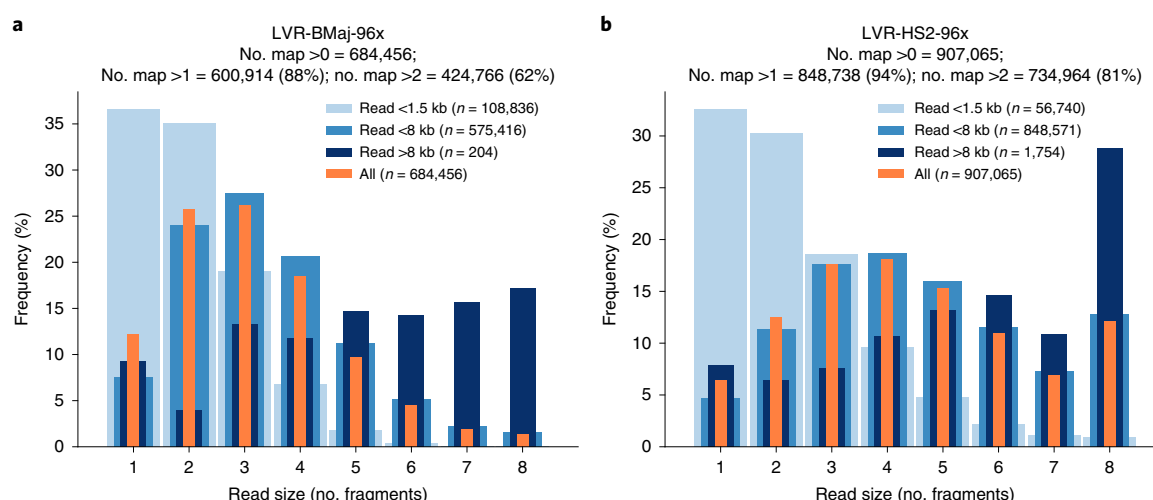


Fig. 9 | Read-size distribution of two MC-4C datasets. These plots show read-size distribution of reads in terms of mapped fragments produced from two MC-4C datasets prepared from mouse liver cells. Light blue bars represent the number of fragments in reads <1.5 kb, medium blue represents reads <8 kb and dark blue represents reads >8 kb. Orange bars indicate the distribution of all reads. The characteristics on top of the graph indicate the number of reads with >0 fragments mapped, with >1 mapped fragment and the number of reads with >2 fragments mapped. Note, animal experiments were performed in compliance with the Dutch law on animal experimentation in line with the European Directive (2010/63/EU).

represents the read-size distributions of two MC-4C datasets from mouse (mm9) and human (hg19). In a successful experiment, we expect a substantial number of reads to fall within the 1.5–8 kb interval.

See ‘Troubleshooting’ section for further actions if the distribution of read size in a given experiment is not as expected.

Chromosome coverage

As MC-4C is a targeted approach, it is expected that many fragments will map to the viewpoint chromosome. This is often dictated by the specificity of designed primers (see the ‘PCR primer design’ section), their amplification efficiency or the presence of contamination in the produced library. To validate these properties in an MC-4C dataset, the chromosome coverage plots drawn by the `chrCvg` command in the `QC` module should be examined. Figure 8 demonstrates two examples produced by this command.

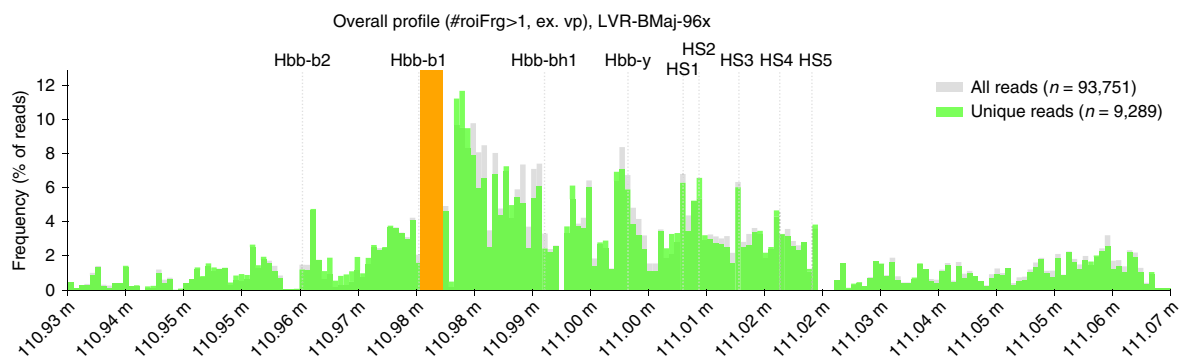


Fig. 10 | Overall profile of an MC-4C dataset. This plot shows the coverage profile of fragments mapped to the ROI and is produced from an MC-4C dataset prepared from the mouse liver cells. Gray bars indicate all mapped reads, and green bars indicate unique reads only. The number of reads in each category (all versus unique reads) is indicated in the legend. The x axis represents genomic coordinates where the corresponding fragments are mapped to. Note that the y axis is plotted as a percentage of total (sequenced or unique) reads. The orange bar indicates the viewpoint area. Labels refer to SOI in the locus of interest. Note, animal experiments were performed in compliance with the Dutch law on animal experimentation in line with the European Directive (2010/63/EU).

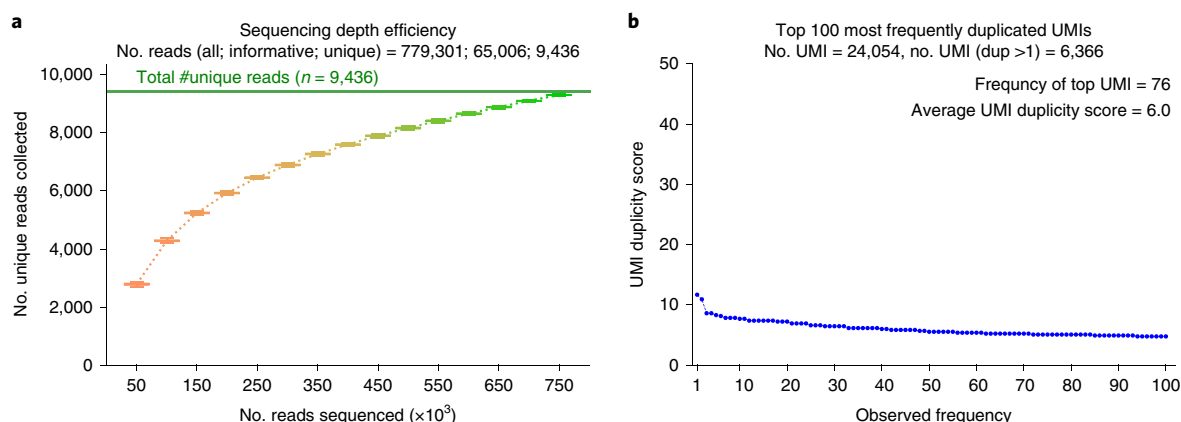


Fig. 11 | Quality check for sequencing saturation and UMI diversity. These plots assess frequency of uniquely captured interactions by analyzing frequency of observed UMIs. These plots are produced from an MC-4C dataset prepared from mouse liver cells. **a**, The number of reads sequenced in the MC-4C experiment is compared to the number of unique reads gained after analyzing the downsampled experiment. A sufficiently sequenced library shows a plateau that indicates that not many unique reads will be sequenced when more reads are obtained from that library. The color scheme refers to the number of reads used in the downsampled experiment. **b**, The complexity of the MC-4C library is investigated by plotting the frequency at which the top UMIs are observed in the MC-4C dataset, represented as a duplicity score. In addition, the average duplicity score of the top 100 UMIs is shown at the top of the plot. An average duplicity score <20 indicates a dataset with sufficient complexity. On top of the plot, the total number of UMIs is shown, as well as the number of UMIs that was sequenced more than once. Note, animal experiments were performed in compliance with the Dutch law on animal experimentation in line with the European Directive (2010/63/EU).

Table 4 | Commands available in the analysis module of the MC-4C pipeline

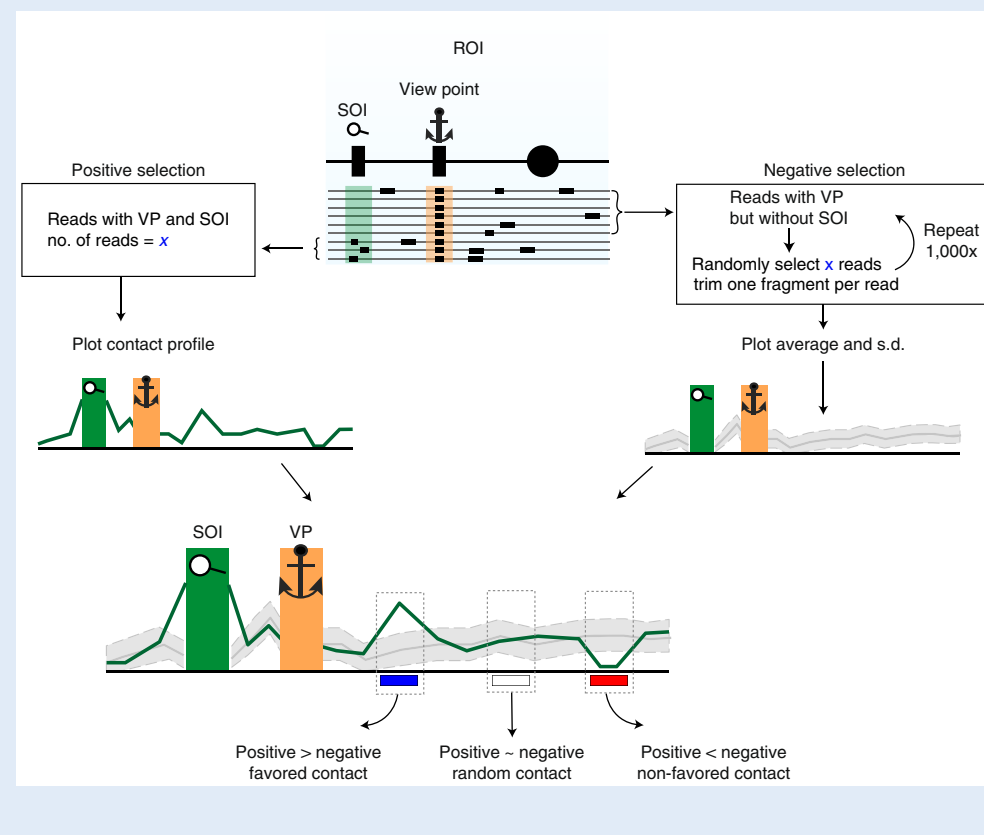
Command name	Function
atVpSoi	Performs three-way contact association test across all bins when a single SOI is interacting with the viewpoint
atSOISOI	Performs three-way contact association test for each pair of SOI
atAcrossROI	Performs three-way contact association test across all bins while each three consecutive bins in the ROI are considered as an SOI

Circle-size distribution

The `cirSizeDist` command enables inspection of the distribution of the number of fragments mapped per read. This is an important measure as multi-way interactions only follow from reads with three or more mapped fragments. The distribution of the number of fragments mapped per read can be calculated using only ROI fragments (fragments mapped within the ROI) or using only unique

Box 2 | VP-SOI analysis

This analysis is devised to investigate the three-way associations between viewpoint, SOI and any given third partner in the ROI. The corresponding details for this analysis are described extensively in our original publication¹³. In summary, we compare the chance of three-way interaction between the viewpoint, the SOI and a given site as the third partner (i.e., the positive set) to a background profile that reflects the three-way interactions of the viewpoint and the same site in the absence of the SOI (i.e., the negative set). To estimate the background, we randomly sample a number of reads from the negative set equal to the number of reads in the positive set. We remove one fragment from each sampled read in the negative set at random to compensate for the SOI fragment present in all reads in the positive set. This random sampling is repeated 1,000 times to generate an average background profile and corresponding s.d. The VP-SOI contact profile is then compared to the background profile, and a z-score is calculated. When the positive profile is higher than the background, a genomic region is more likely to interact with the viewpoint in the presence of the SOI. A positive profile similar to the background profile indicates that such region is equally likely to interact with the viewpoint with or without presence of the SOI. When the positive profile is lower than the background, a region is less likely to contact the viewpoint when it is interacting with the SOI. Figure adapted from Allahyar et al.¹³



reads (duplicates removed). Both are supported in the `cirSizeDist` command through the ‘--roi-only’ and ‘--uniq-only’ arguments (see Fig. 9 for an example of normal circle size distributions).

Overall MC-4C profile in the ROI

The three-way contact data in an MC-4C dataset can be visualized in a similar manner to traditional 4C data with the `overallProfile` command. An example of such a plot is given in Fig. 10. A typical overall profile shows coverage across the ROI where coverage decreases for regions that are farther away from the viewpoint.

Sequencing saturation level

Insufficient complexity (i.e., diversity of reads amplified) in an MC-4C library can result in a low yield of unique reads. To assess the obtained complexity of an MC-4C experiment, the `seqSaturation` command in the `QC` module can be used. This command produces two panels (Fig. 11). Initially, this command investigates the sequencing saturation by downsampling reads in a given MC-4C experiment and plotting the number of reads versus the number of unique reads (see Fig. 11, left panel). If the plotted line reaches a plateau (i.e., sequencing more reads does not yield more unique

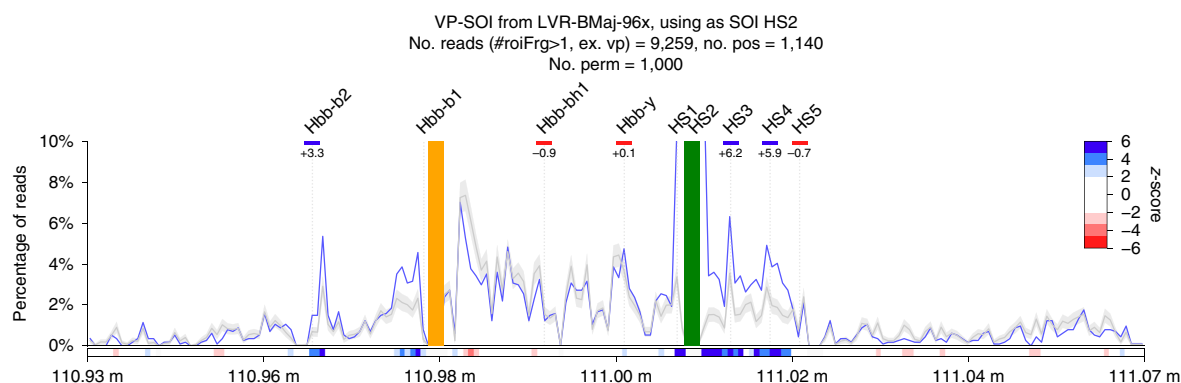


Fig. 12 | Viewpoint-SOI analysis for the *Hbb-b1* gene and HS2 as an SOI. This plot is produced from an MC-4C dataset prepared from mouse liver cells and shows the contact frequency of *Hbb-b1* (the viewpoint, orange bar) while it is interacting with the enhancer element HS2 (the SOI, green bar) as a profile colored in blue. This contact frequency is compared with the background profile (plotted in gray) generated from the interactions of *Hbb-b1* when HS2 is not present in the same read. From this comparison, z-scores are calculated and shown under the profile. The x axis represents the genomic coordinates where corresponding fragments are mapped. No. perm, number of permutations used to generate the background profile; no. pos, number of informative reads that cover the selected SOI; no. reads (#roiFrg>1, ex. vp), number of 'informative' reads (concatemers that contain two or more non-viewpoint fragments in the ROI). Note, animal experiments were performed in compliance with the Dutch law on animal experimentation in line with the European Directive (2010/63/EU).

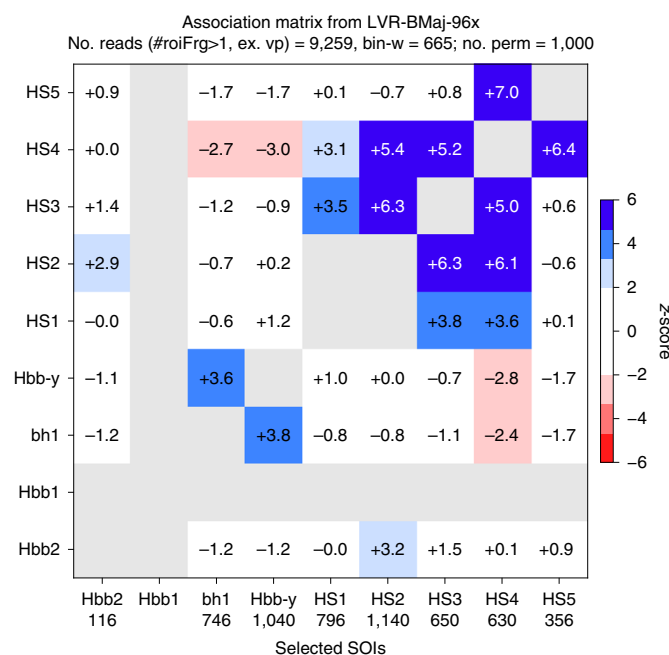


Fig. 13 | Association matrix of an MC-4C dataset. This plot is produced from an MC-4C dataset prepared from mouse liver cells. Each element in the matrix shows the z-score representing the (non)-favored interaction between the corresponding pair of SOIs when the viewpoint (*Hbb1* in this case) is present. Each number in a square indicates the corresponding z-score for each pair of SOI. The bottom axis indicates the number of reads containing both the viewpoint and an SOI (i.e., number of positive reads). No. perm, number of permutations used to calculate the background profile; no. reads (#roiFrg>1, ex. vp), total number of (informative) reads with two or more non-viewpoint fragments in the ROI. Note, animal experiments were performed in compliance with the Dutch law on animal experimentation in line with the European Directive (2010/63/EU).

reads), the user can conclude that the current library is saturated. In addition, this command calculates a UMI duplicity score (the frequency of an observed UMI divided by the total number of informative reads, multiplied by 1×10^5). By plotting the top 100 UMI duplicity scores, the library complexity can be estimated. The average duplicity score of all UMIs can be used to the same effect and is typically <20. Generally, a MC-4C dataset with sufficient diversity of UMIs should have an average duplicity score of <20. Refer to the 'Troubleshooting' section for further actions if the UMI duplicity score is substantially higher than 20.

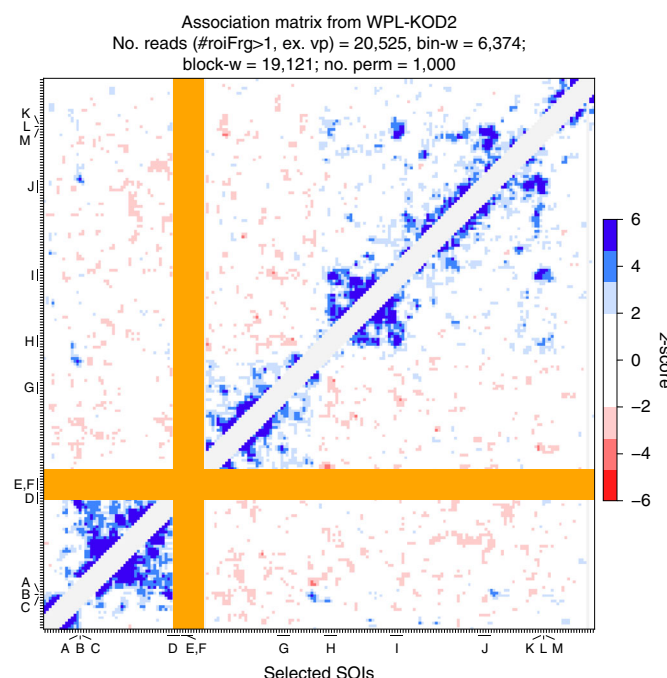


Fig. 14 | Association analysis across an ROI. This plot is produced from an MC-4C dataset prepared from HAP1 cells. Each dot represents the association between three consecutive bins and a single bin in the ROI. A positive z-score (cooperative relationship) is shown in blue, and negative z-scores (mutually exclusive relationship) are shown in red. Bin-w, bin width; block-w, size of SOI (equal to three bins); no. perm, number of subsampling; no. reads, number of informative reads. Note, animal experiments were performed in compliance with the Dutch law on animal experimentation in line with the European Directive (2010/63/EU).

Multi-contact analysis

Several types of three-way contact analysis are supported in the MC-4C pipeline (see Table 4 for a list of available analyses). Each analysis investigates whether one or several SOIs show elevated (favored) or reduced (non-favored) contacts in the locus of interest. An SOI in the MC-4C pipeline is defined as a region in which the user expects (non-)favored contacts to be formed with other SOIs (or bins) when the viewpoint is present in the captured concatemer. The three-way contact association analysis aims to identify contact predisposition between the viewpoint (V) and two other SOIs (e.g., X and Y), by comparing the number reads in two sets of reads, namely ‘positive’ and ‘negative’ sets. The positive set contains all reads that carry X, Y and V. The negative set contains all reads that carry X and V but are missing Y. A read is assumed to carry an SOI (i.e., X or Y) if that read contains a fragment that overlaps with the corresponding SOI region. To account for the conformation diversity in the population of cells in the positive set, we subsampled reads from the negative set (i.e., capturing X and V, but not Y) to the number of reads in the positive set. This procedure is repeated 1,000 times to estimate the frequency of contacts in the negative set (Box 2).

Furthermore, we implemented a correction for the fact that—by definition—reads in the positive set already contributed a fragment to SOI X. Therefore, the positive profile is effectively populated by fewer fragments compared to the negative set. To compensate for this, and ensure both negative and positive profiles are constructed based on the same number of fragments, one fragment from each negative read is randomly removed in every random subsampling of the negative set.

Finally, the mean and s.d. of the frequency at which SOI Y is observed in the negative set are calculated. Using these statistics, a z-score is determined to estimate the significance of the unfavored or favored contacts formed between V, X and Y. Although a modest (close to zero) z-score indicates a random contact frequency between X and Y when V is present, a positive or negative z-score implies a favored or unfavored contact between these three elements, respectively. Generally, we consider a z-score >4 or <-4 to be a significant association. However, z-score corrections must be considered if many sites are being investigated, to keep the false-discovery rate low.

In the MC-4C pipeline, the SOIs can be defined in annotation files (see the ‘Annotations’ section for details). SOI boundaries are determined by three bins centered around the defined coordinate in the annotation file. This restriction is implemented in the MC-4C pipeline to make sure that the SOI

size is always proportional to the ROI size. Our unpublished results (performed by A.A.) revealed that the false positive rate for (non)-preferential contacts identified by the MC-4C pipeline increases once the SOI and ROI size ratio deviates substantially from the suggested (3/200) ratio (data not shown).

Association analysis of a single SOI (`atVpSoi`)

This analysis is performed when the user is investigating whether the interaction propensity of the viewpoint and a given bin within the ROI changes when an SOI is present in the same read. This analysis is implemented in the `vpSoi` command within the **analysis** module. Figure 12 shows an example plot produced by this command. We note that a robust association analysis for a given SOI requires a minimum number of reads to contain the given SOI. As a guideline, we do not consider the *z*-scores reliable for any VP-SOI analysis in which fewer than 100 reads contained the given SOI.

Association test of loci (`atSOISOI`)

The MC-4C pipeline can also perform the three-way association analysis for every pair of SOIs and can visualize the results in a matrix of scores. This analysis can be performed using the `atSOISOI` command implemented in the **analysis** module. This command produces a heatmap of *z*-scores in which cooperative or mutually exclusive association between each pair of SOIs is represented by a positive or negative *z*-score, respectively. An example of such a heatmap is given in Fig. 13.

ROI-wide three-way contact analysis (`atAcrossROI`)

The MC-4C pipeline can provide an ROI-wide view of the three-way associations (viewpoint and two other areas in the ROI). This view is useful when SOIs in an ROI are not well characterized. This analysis can be performed using the `atAcrossROI` command within the **analysis** module. This command produces a heatmap of *z*-scores in which association between each consecutive three bins (considered as an SOI) versus all other bins in the ROI is assessed, and the results are represented by a positive (indicating cooperative relationship) or negative (indicating mutually exclusive relationship) *z*-score. An example of such a heatmap is given in Fig. 14.

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All raw sequencing data used in this study are available through the European Nucleotide Archive (<https://www.ebi.ac.uk/ena/data/view/PRJEB23327>). The processed data can be downloaded from the Mendeley Data repository (<https://doi.org/10.17632/wbk8hk87r2.3>). The test data used in the Walkthrough section of this protocol can be downloaded from github (https://github.com/deLaatLab/mc4c_test-data/archive/master.zip) and are included as Supplementary Data 2.

Code availability

The MC-4C pipeline (version 1.02, at the moment of protocol preparation), including all scripts used to produce figures in this protocol, is available under MIT License at github in `mc4c_py` repository (https://github.com/deLaatLab/mc4c_py) and is available directly as Supplementary Data 1. The data and results presented in this protocol were prepared using version 1.02 of the MC-4C pipeline.

References

1. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
2. Denker, A. & de Laat, W. The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev.* **30**, 1357–1382 (2016).
3. Brant, L. et al. Exploiting native forces to capture chromosome conformation in mammalian cell nuclei. *Mol. Syst. Biol.* **12**, 891 (2016).
4. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
5. Simonis, M. et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* **38**, 1348–1354 (2006).

6. Splinter, E., de Wit, E., van de Werken, H. J., Klous, P. & de Laat, W. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods* **58**, 221–230 (2012).
7. Davies, J. O. et al. Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat. Methods* **13**, 74–80 (2016).
8. Dostie, J. et al. Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16**, 1299–1309 (2006).
9. Li, G. et al. ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol.* **11**, R22 (2010).
10. Mumbach, M. R. et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, 919–922 (2016).
11. Fullwood, M. J. & Ruan, Y. ChIP-based methods for the identification of long-range chromatin interactions. *J. Cell. Biochem.* **107**, 30–39 (2009).
12. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
13. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
14. Sexton, T. et al. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
15. Allahyar, A. et al. Enhancer hubs and loop collisions identified from single-allele topologies. *Nat. Genet.* **50**, 1151–1160 (2018).
16. Bianco, S., Chiariello, A. M., Annunziatella, C., Esposito, A. & Nicodemi, M. Predicting chromatin architecture from models of polymer physics. *Chromosome Res.* **25**, 25–34 (2017).
17. Serra, F. et al. Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. *PLoS Comput. Biol.* **13**, e1005665 (2017).
18. Olivares-Chauvet, P. et al. Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. *Nature* **540**, 296–300 (2016).
19. Darrow, E. M. et al. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc. Natl Acad. Sci. USA* **113**, E4504–12 (2016).
20. Quinodoz, S. A. et al. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* **174**, 744–757.e24 (2018).
21. Beagrie, R. A. et al. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).
22. Oudelaar, A. M. et al. Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nat. Genet.* **50**, 1744–1751 (2018).
23. van de Werken, H. J. et al. 4C technology: protocols and data analysis. *Methods Enzymol.* **513**, 89–112 (2012).
24. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
25. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
26. Koranne, S. Hierarchical data format 5: HDF5. in *Handbook of Open Source Tools* 191–200 (Springer, 2011).
27. Walt, S., van der Walt, S., Chris Colbert, S. & Varoquaux, G. The NumPy array: a structure for efficient numerical computation. *Comput. Sci. Eng.* **13**, 22–30 (2011).
28. McKinney, W. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython* (‘O’Reilly Media, Inc.’, 2017).
29. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
30. Nakayama, T. et al. Cas9-based genome editing in *Xenopus tropicalis*. *Methods Enzymol.* **546**, 355–375 (2014).
31. Koh, C. M. Isolation of genomic DNA from mammalian cells. *Methods Enzymol.* **529**, 161–169 (2013).

Acknowledgements

We thank E. Vos and E. deWit and other laboratory members for their input at various stages of the development of MC-4C. This work was supported by an NWO VIDI grant (639.072.715) to J.d.R. and an NWO/CW TOP grant (714.012.002) and NWO VICI grant (724.012.003) to W.d.L., and by the NIH Common Fund Program, grant U01CA200147, as a Transformative Collaborative Project Award (TCPA; TCPA-2017-DE-LAAT).

Author contributions

C.V. and B.A.M.B. designed and performed experiments. C.V. and A.A. wrote manuscript and designed the figures. A.A. designed and performed the computational analysis, designed the figures and wrote the manuscript. P.H.L.K., M.J.A.M.V. and C.V.-Q. performed ‘C’ methods experiments. R.S. and A.A. implemented the pipeline in Python. I.J.R. performed and W.P.K. designed and supervised MinION sequencing experiments. G.G. helped with computational analysis. J.d.R. designed and supervised the computational analyses and pipelines and co-wrote the manuscript. W.d.L. conceived and supervised the study and wrote the manuscript.

Competing interests

C.V., B.A.M.B., P.H.L.K., M.J.A.M.V. and G.G. are shareholders of Cergentis. W.d.L. is founder and shareholder of Cergentis. W.P.K. and J.d.R. are co-founders and shareholders of Cyclomics.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41596-019-0242-7>.

Correspondence and requests for materials should be addressed to J.R. or W.L.

Peer review information *Nature Protocols* thanks Jian Ma 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: 12 March 2019; Accepted: 2 September 2019;

Published online: 13 January 2020

Related link**Key references using this protocol**

Allahyar, A. et al. *Nat. Genet.* **50**, 1151–1160 (2018): <https://doi.org/10.1038/s41588-018-0161-5>