nature protocols

PROTOCOL

https://doi.org/10.1038/s41596-019-0229-4

Mapping RNA–chromatin interactions by sequencing with iMARGI

Weixin Wu[1,4], Zhangming Yan[1,2,4], Tri C. Nguyen[1], Zhen Bouman Chen[3], Shu Chien[1,2] and Sheng Zhong[1,2*]

RNA–chromatin interactions represent an important aspect of the transcriptional regulation of genes and transposable elements. However, analyses of chromatin-associated RNAs (caRNAs) are often limited to one caRNA at a time. Here, we describe the iMARGI (in situ mapping of RNA–genome interactome) technique, which is used to discover caRNAs and reveal their respective genomic interaction loci. iMARGI starts with in situ crosslinking and genome fragmentation, followed by converting each proximal RNA–DNA pair into an RNA–linker–DNA chimeric sequence. These chimeric sequences are subsequently converted into a sequencing library suitable for paired-end sequencing. A standardized bioinformatic software package, iMARGI-Docker, is provided to decode the paired-end sequencing data into caRNA–DNA interactions. Compared to its predecessor MARGI (mapping RNA–genome interactions), the number of input cells for iMARGI is 3–5 million (a 100-fold reduction), experimental time is reduced, and clear checkpoints have been established. It takes a few hours a day and a total of 8 d to complete the construction of an iMARGI sequencing library and 1 d to carry out data processing with iMARGI-Docker.

Introduction

caRNAs are proposed to be a layer of the epigenome[1]. Interactions of caRNA with chromatin are essential for diverse molecular and cellular functions, including X-chromosome silencing[2], anchoring of nucleolus–chromosome interactions[3], homology-directed repair of telomeres[4] and RNA-mediated epigenetic inheritance[5]. At specific genomic loci, caRNA–DNA interactions also contribute to de novo DNA methylation[6], promotion[7,8] and suppression of transcription[9], as well as demarcation of active and silent chromatin domains[10].

The genomic interaction loci of a specific caRNA can be determined by a number of technologies, including capture hybridization analysis of RNA targets (CHART)[11], chromatin isolation by RNA purification (ChIRP)[12] and RNA antisense purification (RAP)[13]. However, these one-RNA-versus-the-genome technologies are not applicable to the discovery of new caRNAs. This challenge was recently addressed by a cohort of assays that can reveal caRNA–genome interactions in an all-RNA-versus-the-genome manner. These genome-wide RNA–chromatin interaction assays include MARGI[14] and its improved version iMARGI[15], chromatin-associated RNA sequencing (ChAR-seq)[16], and global RNA interactions with DNA by deep sequencing (GRID-seq)[17]. Both GRID-seq and ChAR-seq revealed a range of chromatin-bound RNAs, including nascent transcripts, chromosome-specific dosage compensation non-coding RNAs (ncRNAs), and trans-associated RNAs[16,17]. GRID-seq also revealed extensive interactions between mRNAs and enhancers[17].

MARGI and iMARGI revealed thousands of caRNAs, including both coding and ncRNAs[14,15,18]. These caRNAs are not only associated with the genomic sequences from which they are transcribed (to form proximal interactions) but can also attach to distal genomic sequences (to form distal interactions) on the same chromosomes or to other chromosomes (to form inter-chromosomal interactions). Surprisingly, transcription start sites (TSSs) have been identified as the preferred genomic loci targeted by non-coding caRNAs through distal and inter-chromosomal interactions[14]. A subsequent effort revealed that many non-coding caRNAs are tethered to chromatin by RNA polymerase II (Pol-II)-associated U1 snRNP (small nuclear ribonucleoprotein), which may offer a mechanistic explanation of the accumulation of caRNA on TSSs[19]. The amount of accumulated

[1]Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA. [2]Institute of Engineering in Medicine, University of California, San Diego, La Jolla, CA, USA. [3]Department of Diabetes Complications and Metabolism, Beckman Research Institute, City of Hope, Duarte, CA, USA. [4]These authors contributed equally: Weixin Wu, Zhangming Yan. *e-mail: szhong@ucsd.edu

caRNAs at each TSS correlated with the expression level of the gene whose transcription starts at that TSS[14], suggesting a regulatory role of the TSS-accumulated caRNAs. Consistent with this notion, suppression of TSS-accumulated caRNA leads to lower expression of the gene transcribed from the TSS[7]. Another unexpected consequence of the accumulation of distal and inter-chromosomal caRNAs at TSS is that the TSS-associated caRNAs become susceptible to being trans-spliced with their nearby nascent transcripts and, thus, create fusion RNAs[15]. These data highlight the power of iMARGI to generate novel hypotheses related to functions of caRNA. Comparisons of MARGI and iMARGI data in terms of genome-wide read pair distribution and numbers of discovered caRNAs can be found in ref. [15].

Development of the protocol

The central idea of the iMARGI technology is to convert each RNA–DNA interaction preserved in the crosslinked chromosomes into a unique DNA sequence and leverage paired-end sequencing as the high-throughput readout for mapping RNA–DNA interactions (Fig. 1a). RNA and DNA are first fragmented, and then a specifically designed linker sequence is introduced to ligate to RNA on one end and to DNA on the other end, thus forming RNA–linker–DNA chimeric sequences. The RNA–linker–DNA chimeric sequence is subsequently converted to double-stranded DNA (dsDNA) suitable for paired-end sequencing. Each end of a read pair is separately mapped to the genome, and pairing information of the two read ends is used to infer the original RNA–DNA interaction.

Critical to this experimental design is keeping track of which side of the linker corresponds to RNA and which side corresponds to DNA. To record this orientation information, the RNA must be specifically ligated to one designated side of the linker, and the fragmented DNA must be specifically ligated to the other side. These orientation-specific ligations are ensured by the design of the linker sequence (Fig. 1b).

To extract the orientation information, the linker sequence itself must be sequenced. However, because the linker resides in the middle of the desired ligation products in the form of RNA–linker–DNA, typical next-generation sequencing is not always able to sequence through the linker region, resulting in uninformative reads, which do not have the necessary orientation information. To address this challenge, iMARGI incorporates a strategy whereby the reverse transcribed strand (cDNA) of the ligation product (cDNA–linker–DNA) is circularized and then relinearized by cutting at the BamHI site within the linker sequence (Fig. 1c,d). As a result, the linker is cut into two halves, which are positioned at the two ends of the chimeric sequence, forming a left.half.Linker–cDNA–DNA–right.half.Linker (Fig. 1d). The two halves of the linker (left.half.Linker and right.half.Linker) are designed to be parts of the two NEBNext PCR primers for Illumina. Thus, the relinearized sequences can be directly amplified with the same PCR primers (blue and red sequences, Fig. 1d) during the preparation of a typical sequencing library.

Other important considerations of the iMARGI protocol include performing the ligation steps in intact crosslinked nuclei, maximizing ligation efficiencies, minimizing side products and removal of uninformative sequences, including unligated or partially ligated products. Compared to ligation in vitro, ligation in intact crosslinked nuclei avoids nonspecific interactions between RNAs and streptavidin beads. This is achieved by preparation and permeabilization of crosslinked nuclei, a combination of RNA and DNA end modifications, and specifically designed and modified ends of the linker sequence. To remove incomplete ligation products in the form of RNA–linker, ExoI exonuclease and T4 DNA polymerase with exonuclease activity are applied to cleave the biotinylated nucleotide in the linker sequence. The cleaved linker sequences cannot be pulled down with streptavidin and are thus discarded. Finally, to minimize the sequence bias of RNA–linker ligation, the 5′ end of the top strand of the linker sequence is composed of two random nucleotides ('NN' in Fig. 1b). Different RNA sequences tend to preferentially ligate to the linkers in the randomized pool with which they have the potential to form particular structures[20–22].

To minimize variation in data processing, we developed a complete data-processing pipeline with full documentation (https://sysbio.ucsd.edu/imargi_pipeline) (Fig. 2). This pipeline, called iMARGI-Docker, helps to improve analysis reproducibility by standardizing the data-processing steps. Based on the Docker technology, iMARGI-Docker can be executed on all mainstream Linux distributions without system-specific configuration, thus simplifying the computer technology requirements of the data analysis.

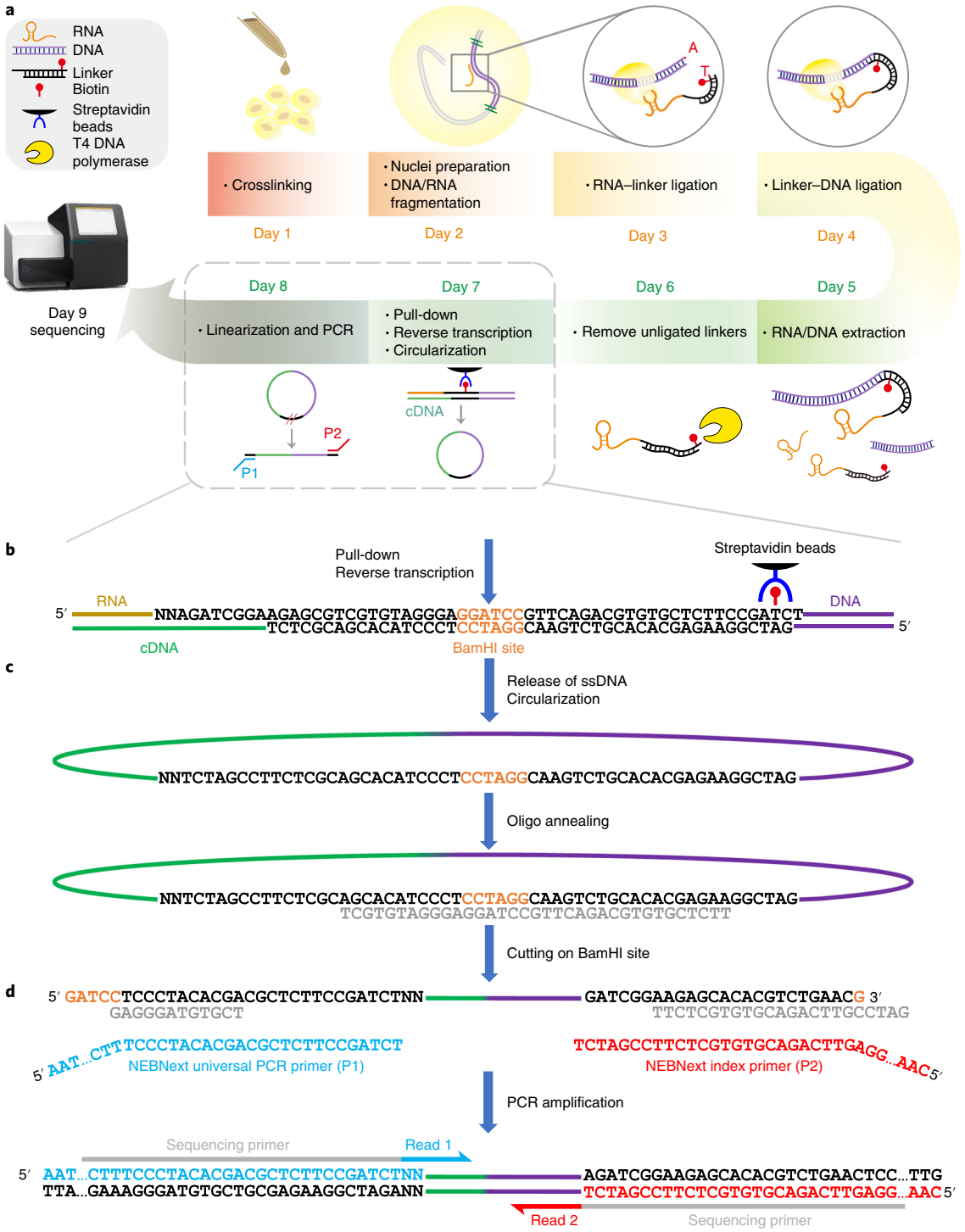

Fig. 1 | iMARGI protocol and linker design. a, A schematic overview of the iMARGI protocol. Experimental steps are carried out in nuclei (days 1–4; Steps 1–54) and subsequently in solution (days 5–8; Steps 55–113). **b**–**d**, An expanded view of the experimental steps on days 7 and 8. **b**, The linker sequence is composed of two strands of DNA with different lengths (black characters). The double-stranded region of the linker sequence contains a BamHI restriction site (orange). The top strand of the linker sequence is biotin-labeled (red pin). The double-stranded linker is ligated to genomic DNA (purple lines). cDNA (green line) is synthesized in the 3′ direction of the bottom strand. **c**, The bottom strand is released as single-stranded DNA (ssDNA), circularized and hybridized with an annealing oligo (gray characters). A cut on the BamHI site linearizes the circular DNA and puts the two halves of the bottom strand of the linker at the two ends of the ssDNA. **d**, The half linker at the 5′ end is identical to a fraction of the NEBNext universal PCR primer (blue characters), which contains Illumina's sequencing primer (gray bar) for Read 1 (blue arrow). The other half linker (at the 3′ end) is complementary to a fraction of the NEBNext index primer (red characters), which contains Illumina's sequencing primer (gray bar) for Read 2 (red arrow). P1, primer 1; P2, primer 2.

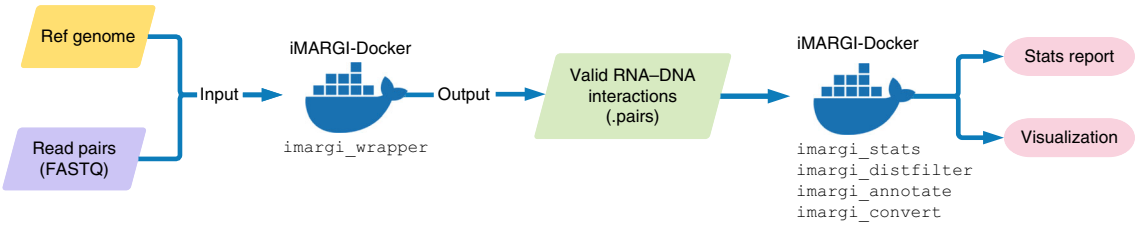

Fig. 2 | Computational workflow of iMARGI data analysis. Ref, reference.

Applications of the protocol

iMARGI does not require genome engineering or introduction of any exogenous molecules into cells before the cells are fixed. Thus, iMARGI is applicable to both cell lines and primary cells. Several millions of cells are sufficient for carrying out iMARGI analysis. Although we recommend starting with 5 million cells, our tests with 2–3 million cells generally produced high-quality data as well. We anticipate that the minimum required cell number will vary among cell types. Refer to Step 113 for determining the minimum amount of input cells needed.

Comparison to other methods

The three genome-wide RNA–chromatin interaction assays include iMARGI (and its predecessor MARGI[14]), GRID-seq[17], and ChAR-seq[16]. iMARGI and its predecessor MARGI share the linker design and most of the experimental steps. However, in iMARGI, the ligation steps are carried out in the nuclei instead of in solution. As a result, iMARGI requires fewer cells for the experiment while generating a larger number of informative sequence read pairs (see 'Comparison of iMARGI with MARGI' section in ref. [15]). We will devote the rest of this section to comparison of iMARGI, GRID-seq and CHAR-seq.

iMARGI has been tested on mammalian cells. MARGI and GRID-seq have been tested on fruit fly and mammalian cells, whereas ChAR-seq has been tested only with fruit fly cells[16] (Table 1). iMARGI and GRID-seq require several millions of cells, whereas ChAR-seq requires >100 million cells (Table 1). Considering these differences, hereafter we will emphasize the comparison between iMARGI and GRID-seq.

A main advantage of iMARGI is the maximized usable sequence length of the reads. Because the linker does not appear in the final read sequences, the entire sequence of each read pair can be used for sequence mapping, thus minimizing ambiguity in sequence alignment. By contrast, in GRID-seq the linker must be read through to determine which portion of the read sequence corresponds to RNA and which corresponds to DNA. To ensure that the linker sequence is included within the read sequence, GRID-seq uses an experimental strategy that trims the RNA side and the DNA side of the ligation product to 18–23 bp each[17]. Thus, only 18–23 bp of the sequence can be used for mapping. Sequences in this size range cannot always be uniquely mapped to the human genome. Even without allowing for any mismatch, >25% of the 20-bp sequences had ambiguous mapping to the human genome (see Fig. 1 in ref. [23]).

The second advantage of iMARGI is its minimized potential bias. iMARGI has a unique design in which the linker contains two random bases on the 5′ end of the top strand (Fig. 1b,c), which minimizes the sequence bias of the RNA ligase[15]. Linker sequences used in GRID-seq and ChAR-seq do not contain these random bases (Table 1). Furthermore, in iMARGI, the first-strand cDNA synthesis is carried out in solution, whereas in GRID-seq and ChAR-seq this process is carried out in the nuclei. It is possible that performing these enzymatic reactions in situ may result in partial inhibition by RNA structure or RNA binding proteins (Table 1).

In addition to the two above-noted advantages, the iMARGI experimental procedure is simplified in comparison to GRID-seq or ChAR-seq. Although iMARGI's linker (composed of two DNA strands) can be easily prepared by annealing, GRID-seq's RNA–DNA chimeric linker requires special synthesis (Table 2). Furthermore, in iMARGI, enzyme-based removal of unligated products is used, whereas GRID-seq relies on gel-based size selection, which can be time consuming and can cause loss of material. Another notable advantage is that iMARGI does not require second-strand DNA synthesis, which is required by GRID-seq and ChAR-seq. iMARGI converts the first synthesized DNA strand into a suitable form for PCR-based library construction, thus simplifying the

Table 1 | Comparison of iMARGI, GRID-seq and ChAR-seq

	iMARGI[15]	GRID-seq[17]	ChAR-seq[16]	Pros and cons
Applications				
Tested species	Human	Human, mouse, fruit fly	Fruit fly	iMARGI and GRID-seq have been tested in human cells
Input cell number	2–5 million	~2 million	100 million	
Data format				
Sequencing type	Paired end	Single end	Single end	
Usable sequence length for mapping RNA	Up to 100 bp	18–23 bp	65 bp on average	Short sequence length can result in ambiguity in mapping
Usable sequence length for mapping DNA	Up to 100 bp	18–23 bp	65 bp on average	
Experimental design				
Ligation	In nucleus	In nucleus	In nucleus	
Reverse transcription	In solution, with RNase inhibitor	In nucleus	In nucleus	Reverse transcriptase is less likely to be blocked by RNA-binding proteins or RNA secondary structure in solution (iMARGI)
Second-strand synthesis	Not applicable	Yes	Yes	iMARGI's ssDNA product can be directly PCR-amplified to produce sequencing library
Removal of incomplete ligation products	Exonuclease treatment combined with biotin selection	Gel-based size selection	No relevant experimental steps	iMARGI's enzyme-based method is easier to carry out than gel-based selection, which may cause loss of material
Ligation of adaptors	Not applicable	Y-shaped adaptors required	NEBNext hairpin adaptors required	iMARGI does not require adaptor ligation steps

Table 2 | Comparison of the linker sequences in iMARGI, GRID-seq and ChAR-seq

	iMARGI[15]	GRID-seq[17]	ChAR-seq[16]	Comment
Biotinylation	Biotinylated	Biotinylated	Biotinylated	Allows for selection of ligation products
Strands	A longer ssDNA (top strand) annealed to a shorter ssDNA (bottom strand)	ssDNA annealed to an RNA–DNA chimeric sequence	dsDNA with a 5′-overhang and a 3′-three-carbon spacer	GRID-seq's linker sequence requires special synthesis
5′ Adenylation	5′ Adenylated	5′ Adenylated	5′ Adenylated	For ligation to 3′ end of RNA
5′ Random bases	Contains two random bases on the 5′ end of linker top strand	None	None	5′ Random bases (NN) in the linker top strand can minimize biases of RNA ligase
Restriction sites	BamHI	2 × MmeI	DpnII	DpnII allows for carrying out Hi-C experiments in parallel
Complementarity to adaptors	Yes	No	No	Without sequencing the linker sequence, iMARGI is able to resolve the orientation of the RNA–linker–DNA. There is no need to ligate library construction adaptors in iMARGI

ssDNA, single-stranded DNA; dsDNA, double-stranded DNA.

experimental process. Finally, iMARGI does not require ligation of sequencing adaptors, which is required by GRID-seq and ChAR-seq. After splitting the linker, the two halves of iMARGI's linker sequences are compatible with library construction PCR primers. Using standard library preparation primers, iMARGI's sequence products can be directly amplified into a sequencing library.

iMARGI/MARGI, GRID-seq and ChAR-seq use different methods to assess false-positive rates. With MARGI and GRID-seq, mixed fruit fly and mammalian cells were used as input and the ligated sequences of the two species were used as false positives[14,17]. ChAR-seq was used with spike-in RNAs to estimate false positives[16].

For statistical analyses, GRID-seq provided a data normalization method to correct for background noises and identify specific RNA–DNA interactions[17]. In comparison, iMARGI/MARGI and

ChAR-seq did not provide a one-size-fits-all statistical method for identifying all RNA–DNA interactions. Instead, the authors opted for the statistical methods that best fit each biological question, including calling chromatin-enriched RNAs[16]; assessing enrichments of RNA–DNA interactions at TAD boundaries[16], TSSs[14] and the chromosomal regions near nuclear speckles[18]; and comparing global RNA–DNA interactions with genome-wide distributions of histone modifications[14,16], transcription factor binding intensities[18] and fusion RNAs[15].

Limitations of the protocol

The major limitation of the iMARGI technology is the requirement for millions of input cells. Another limitation is the relatively long experimental process, which requires 3–6 h per day and a total of 8 d to complete the library construction. Furthermore, owing to scarcity of independent experiments, it is difficult to globally evaluate the sensitivity and specificity of interactions detected by iMARGI. In addition, iMARGI does not distinguish 'active' interactions such as those of the RNAs bound to a chromatin-binding protein complex[24] and 'passive' proximity such as nascent RNA–chromatin interactions introduced by DNA looping[15]. Finally, we have not provided a unified statistical method for identifying all the caRNAs. We suspect that a unified statistical method cannot simultaneously optimize the sensitivity and the specificity, in light of the diverse modes of RNA–DNA interactions. It falls to the users to choose suitable statistical methods on the basis of their biological questions.

Experimental design

iMARGI starts with crosslinking cells and collecting nuclei, followed by fragmenting RNA and DNA in nuclei (Fig. 1a). A specifically designed linker sequence is introduced into the permeabilized nuclei to ligate it first with the fragmented RNA and subsequently with spatially proximal DNA. After these ligation steps, nuclei are lysed and crosslinks are reversed. Nucleic acids are purified and subsequently treated with exonucleases to remove any linker sequences that were not successfully ligated with both RNA and DNA. The desired ligation products in the form of RNA–linker–DNA are pulled down with streptavidin beads. The RNA part of the pulled-down sequence is reverse transcribed into cDNA, resulting in a complementary strand of (5′)DNA–linker–cDNA(3′) (Fig. 1b). Single-stranded DNA (ssDNA)–linker–cDNA is released from streptavidin beads, circularized and relinearized, producing ssDNA in the form of left.half.Linker–cDNA–DNA–right.half.Linker (Fig. 1c,d). The two halves of the linker (left.half.Linker and right.half.Linker) are templates for PCR amplification into the final sequencing library (Fig. 1d).

Collection of nuclei

Approximately 5 million cells are fixed in 1% (wt/vol) formaldehyde. We have successfully performed iMARGI with human embryonic kidney (HEK293T) cells, human foreskin fibroblasts (HFFc6), and human umbilical vein endothelial cells (HUVECs). To isolate cell nuclei, cells are incubated with a lysis buffer containing mild non-ionic detergent, which selectively disrupts the plasma membrane while keeping the nuclear membrane intact[25–27]. Nuclei are subsequently obtained by centrifugation. A white nuclear pellet can be observed by eye at the bottom of the tube after centrifugation. As a checkpoint for nuclear integrity, the nuclei are stained with DAPI and imaged. Confined DAPI signals in the nuclei are an indication of nuclear integrity (Fig. 3a–c).

Fragmentation of DNA and RNA in nuclei

The harvested nuclei are permeabilized by SDS treatment and subsequently incubated with RNase I and the restriction enzyme AluI for RNA and DNA fragmentation, respectively.

The recommended SDS concentration and treatment time have been optimized to balance the prevention of leakage of nuclear contents with sufficient enzyme penetration of the nuclei. The degree of nuclear leakage can also be quantified by the ratio of supernatant RNA (sRNA) to nuclear-retained RNA (nRNA) and the ratio of supernatant DNA (sDNA) to nuclear-retained DNA (nDNA) (checkpoint at Step 23; Fig. 3d–f). Sufficient enzyme penetration is reflected by obtaining the desired lengths of the fragmented DNA and RNA (checkpoint in Step 25). The desired lengths of fragmented DNA range from 200 to 1,500 bp (Fig. 4a). The desired lengths of fragmented RNA range from 180 to 500 nt (Fig. 4b).

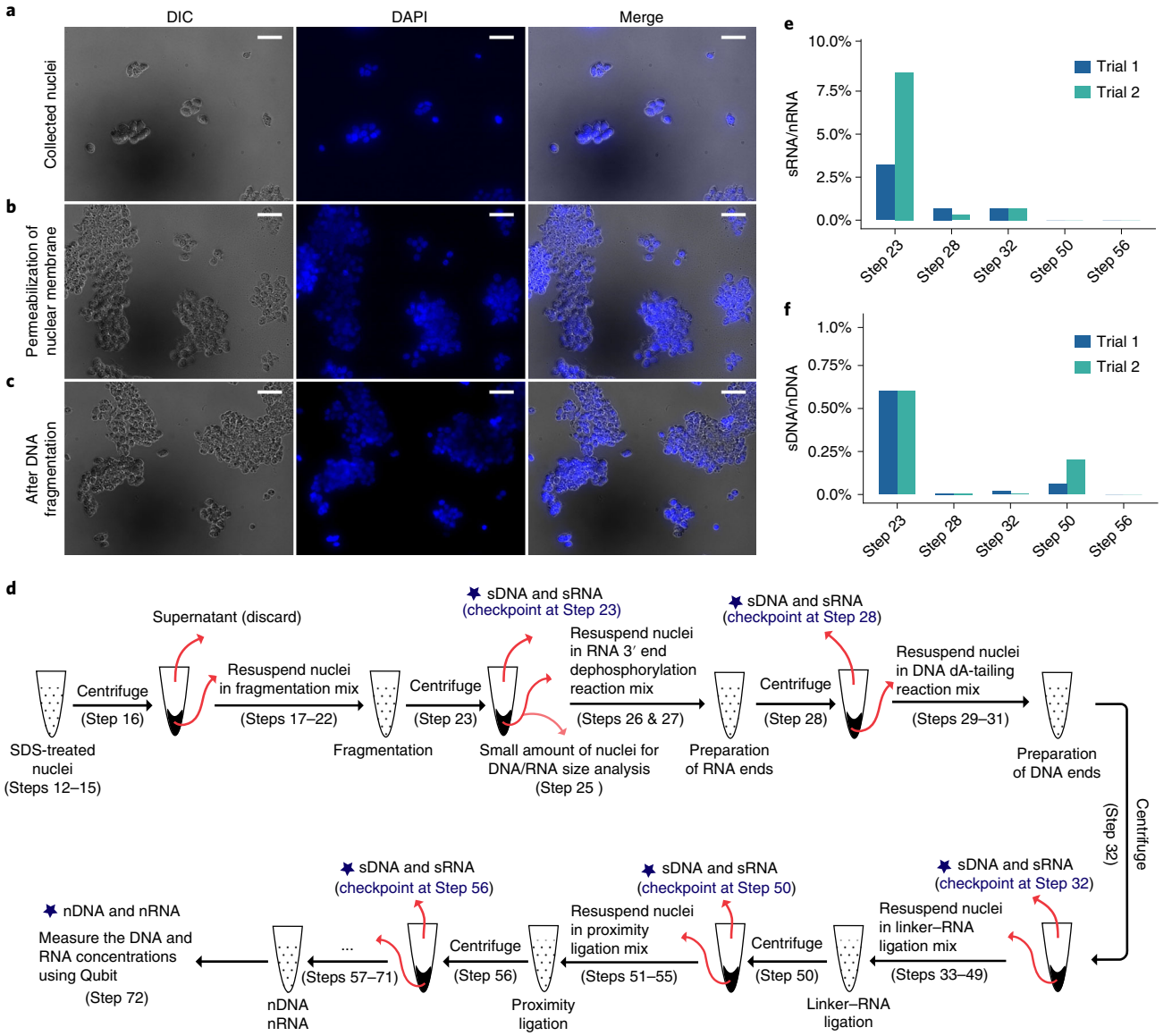

Fig. 3 | Checkpoints for nuclear integrity. a–c, Image-based checkpoints (Box 1). DAPI staining (blue) and differential interference contrast (DIC) images (gray) were taken after collection of nuclei (Step 9) (**a**), after permeabilization of the nuclear membrane and collection of nuclei (Step 17) (**b**), and after DNA fragmentation and collection of nuclei (Step 24) (**c**). Confined DAPI staining in DIC-defined nuclei is an indication of non-leaking nuclei. Scale bars, 50 μm. **d–f**, Quantitative assessment of nuclear integrity (Box 2). **d**, A schematic view of the experimental steps in which supernatant RNA (sRNA) and supernatant DNA (sDNA) are quantified. Nuclear-retained RNA (nRNA) and nuclear-retained DNA (nDNA) are quantified at Step 72. **e**, Ratios (*y* axis) between the amount of sRNA at each step (*x* axis) and the amount of nRNA measured at Step 72. **f**, Ratios (*y* axis) between the amount of sDNA at each step (*x* axis) and the amount of nDNA measured at Step 72. Trials 1 and 2 are two separate experiments that start with ~5 million and 3 million HEK293T cells, respectively. The HEK293T cell line used in panels in this figure has been authenticated and tested to ensure its identity and that it is free from mycoplasma contamination.

RNA–linker and linker–DNA ligation in nuclei

The top strand and the bottom strand of the linker are annealed to create suitable linkers for ligations. To prepare the fragmented RNA for ligation, T4 polynucleotide kinase (PNK) is applied to convert any 3′ phosphate groups on the fragmented RNA into 3′ hydroxyl (-OH) groups. RNA with 3′-OH is subsequently ligated to the top strand of the linker, which has an adenylated 5′ end, using T4 RNA ligase 2, truncated KQ. Unligated linkers are removed by extensive washes. To prepare the fragmented DNA for ligation, an 'A' base is added to the DNA strand with an exposed 3′ end using Klenow fragment, which lacks 3′–5′ exonuclease activity. The A-tailed DNA is subsequently ligated to the linker through sticky-end ligation.

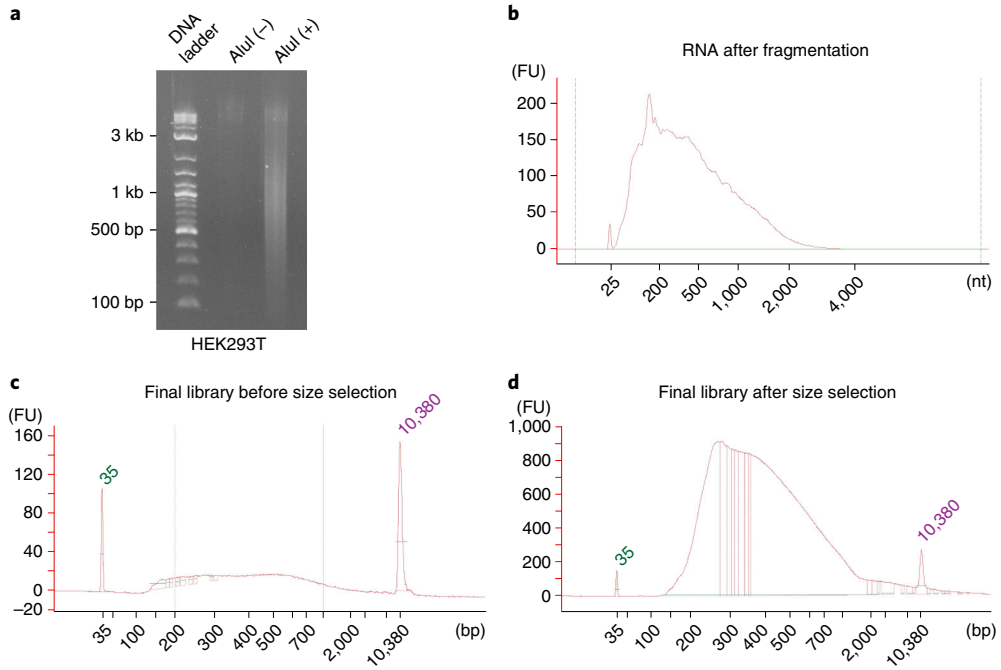

Fig. 4 | Size distributions of intermediate and final products from HEK293T cells. a, Size distributions of unfragmented (AluI (−)) and fragmented DNA (AluI (+)) resolved on a precast 1% (wt/vol) agarose E-Gel. 2-Log DNA ladder was used as ladder. **b**, Size distribution of fragmented RNA resolved on an Agilent Bioanalyzer. **c**,**d**, Size distributions of iMARGI sequencing library before (**c**) and after (**d**) size selection. The peaks at 35 and 10,380 bp are size markers that are not included in the sequencing library. The two vertical bars indicate 200 bp and 1,000 bp, respectively.

Collecting ligation products

After the ligation steps, nuclei are collected, washed and incubated with extraction buffer to extract nucleic acids and reverse crosslinks. The extracted nucleic acids include the desired ligation products in the form of RNA–linker–DNA, as well as incomplete products in the forms of RNA, DNA, and RNA–linker. The biotin on the linkers of the RNA–linker products is cleaved by exonucleases. After the cleavage, the desired ligation products are harvested by streptavidin beads, followed by stringent wash steps to remove any nucleic acids attached because of nonspecific binding.

Constructing the sequencing library

Reverse transcription is performed to produce cDNA, resulting in a complementary strand of cDNA–linker–DNA sequence (Fig. 1b). This complementary strand is released from the streptavidin beads by using denaturing buffer, which contains 0.1 M NaOH and 0.1 mM EDTA. The released strand is circularized using CircLigase (Fig. 1c). This single-stranded circular DNA is annealed with a hybridization oligo to create a double-stranded region containing a BamHI restriction site. A subsequent BamH1 cut linearizes the DNA, which splits the linker into two halves. The two halves of the linker at the two ends of the relinearized DNA are compatible with sequencing primers (Fig. 1d). The linearized DNA is amplified with NEBNext PCR primers for Illumina, size-selected (Fig. 4c,d) and subjected to paired-end sequencing.

Checkpoints

We designed checkpoints to gauge the success of several key experimental steps. These checkpoints examine the integrity of the extracted nuclei (Steps 9 and 23, Fig. 3), the efficiency of DNA and RNA fragmentation (Step 25) (Fig. 4a,b), the success of annealing the two strands of the linkers (Step 46), and the size distribution of the sequencing library (Steps 105 and 113) (Fig. 4c,d). Figure 5 shows results from testing different conditions to optimize DNA fragmentation. Empirical thresholds for quantitative evaluation of broken nuclei are given in Table 3. Moreover, we provide checkpoints to ensure sequencing quality (Step 115) and appropriate data processing (Steps 116, 117, 119 and 123).

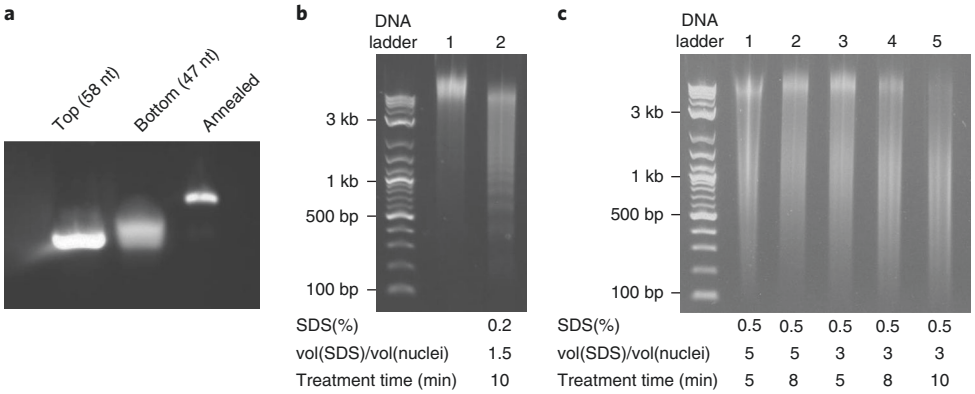

Fig. 5 | Checkpoints and troubleshooting. a, Checkpoint for annealing the two strands of the linker sequence. Sizes of the top strand (58 nt), bottom strand (47 nt) and the annealed product are resolved on a precast 2% agarose E-Gel. **b**,**c**, Troubleshooting for insufficient DNA fragmentation. Size distributions of unfragmented (control, lane 1 in **b**) and fragmented DNA (lane 2 in **b**, lanes 1–5 in **c**) under different parameters of nuclear permeabilization. 2-Log DNA ladder was used in the ladder lane. The corresponding parameters, including SDS concentration (SDS%), volume ratios (vol(SDS)/vol(nuclei)) and treatment time, are marked below each fragmented DNA lane. Lane 5 of **c** corresponds to the default parameters of the iMARGI protocol. The parameters resulting in optimal DNA fragmentation without disrupting nuclear integrity should be chosen.

Table 3 | Empirical thresholds for quantitative evaluation of broken nuclei

Step no. for retrieving sRNA and sDNA	sRNA/nRNA	sDNA/nDNA
23	10/100	1/100
28	1/100	1/100
32	1/100	1/100
50	1/100	1/100
56	1/100	1/100

These thresholds should be applied to the ratios of the measured amounts of sRNA and nRNA (sRNA/nRNA column) and the amounts of sDNA and nDNA (sDNA/nDNA column). The sRNA and sDNA should be obtained from the supernatant in Steps 23, 28, 32, 50 and 56 (rows). nRNA and nDNA should be obtained from Step 72. Any obtained ratio at any of these steps beyond the listed threshold of that step is a sign of having a non-negligible fraction of broken nuclei.

Data processing

We have automated the data-processing steps with iMARGI-Docker. The key steps include quality control (QC) of the sequencing data (Step 115), checking the hardware and software requirements (Step 116), QC of the mapped data (Step 123), and conversion of the mapped data to other data formats that are compatible with other popular data analysis and visualization tools (Steps 127–129).

Materials

Biological materials

- Cell lines of interest. We have used a variety of cell types successfully, such as human embryonic kidney (HEK) 293T cells (ATTC, CRL-1573), human foreskin fibroblasts (HFFc6) (4DN Cell Line Working Group), human umbilical vein endothelial cells (HUVECs) (Cell Applications, 200-05n) and human embryonic stem cells (hESCs) H1 (4DN Cell Line Working Group). All data shown in this paper were generated from HEK293T cells **! CAUTION** The cell lines used in your research should be regularly checked to ensure they are authentic and are not infected with mycoplasma.

Reagents

- Penicillin–streptomycin (Thermo Fisher Scientific, cat. no. 15140122)
- FBS (Thermo Fisher Scientific, cat. no. A2720803)
- DMEM (Genesee Scientific, cat. no. 25–500)
- mTeSR1 basal medium (400 mL; StemCell Technologies, cat. no. 85851)

- mTeSR1 5× supplement (100 mL; StemCell Technologies, cat. no. 85852)
- ReLeSR (StemCell Technologies, cat. no. 05872/05873)
- Corning Matrigel hESC-qualified matrix (Corning, cat. no. 354277)
- DMEM/F-12 (StemCell Technologies, cat. no. 36254)
- Formaldehyde (16% (wt/vol); Thermo Fisher Scientific, cat. no. 28908) **! CAUTION** Formaldehyde is toxic upon ingestion, inhalation or contact with skin. Handle formaldehyde in a chemical fume hood with caution. Protective gloves, lab coat and eye protection should be worn. Formaldehyde waste should be discarded in an appropriate container according to local regulations.
- UltraPure distilled water (DNase, RNase and protease free; Invitrogen, cat. no. 10977-015)
- Glycine (Fisher Scientific, cat. no. BP381-1)
- Protease inhibitor cocktail tablets (Roche, cat. no. 04693132001)
- Tris-HCl (1 M, pH 7.5; Invitrogen, cat. no. 15567027)
- Tris-HCl (1 M, pH 6.5; Teknova, cat. no. T1065)
- Nonidet P 40 (Sigma-Aldrich, cat. no. 74385-1L)
- CutSmart buffer (10×; New England Biolabs, cat. no. B7204S)
- SDS (20% (wt/vol); Invitrogen, cat. no. AM9820)
- Triton X-100 (Acros Organics, cat. no. 215682500)
- AluI (10 U/μL; NEB, cat. no. R0137L)
- RNasin Plus ribonuclease inhibitor (40 U/μL; Promega, cat. no. N2615)
- PBS (10×; Invitrogen, cat. no. AM9625)
- RNase I (50 U/μL; NEB, cat. no. M0243S)
- RNase A (10 mg/mL; Thermo Fisher Scientific, cat. no. EN0531)
- TURBO DNase (2 U/μL; Invitrogen, cat. no. AM2239)
- $MgCl_2$ (1 M; Invitrogen, cat. no. AM9530G)
- DTT (1 M in H_2O; Sigma-Aldrich, cat. no. 43816-10ML)
- T4 polynucleotide kinase (T4 PNK; 10 U/μL; NEB, cat. no. M0201L)
- NEBuffer 2 (10×; NEB, cat. no. B7002S)
- Klenow fragment (3′–5′ exo-; 5 U/μL; NEB, cat. no. M0212L)
- dNTP Solution Set (100 mM, 0.25 mL of each dNTP; NEB, cat. no. N0446S)
- 5′ DNA Adenylation Kit (NEB, cat. no. E2610L)
- Dynabeads MyOne Silane (40 mg/mL; Invitrogen, cat. no. 37002D)
- Isopropanol (Sigma-Aldrich, cat. no. I9516-500ML) **! CAUTION** Isopropanol is highly flammable, both as a liquid and as a vapor. It can cause irritation to the eyes and skin. Avoid inhalation of vapors, mist and gas. Handle with caution and ensure adequate ventilation.
- T4 RNA ligase reaction buffer (10×; NEB, cat. no. B0216S)
- T4 RNA ligase 2, truncated KQ (200 U/μL; NEB, cat. no. M0373L) **▲ CRITICAL** T4 RNA ligase 2, truncated KQ must not be replaced with other ligases, such as T4 RNA ligase 1 or T4 RNA ligase 2. T4 RNA ligase 2, truncated KQ is specifically used in this protocol to ligate the pre-adenylated 5′ DNA linker end to the 3′ OH end of the RNA in the absence of ATP to prevent RNA self-ligation. Other ligases are not compatible with this protocol.
- PEG 8000 (50% (wt/vol); NEB, cat. no. B1004S)
- DNA Ligase Reaction Buffer (10×; NEB, cat. no. B0202S)
- T4 DNA Ligase (2,000 U/μL; NEB, cat. no. M0202M)
- BSA (20 mg/mL; NEB, cat. no. B9000S)
- EDTA (0.5 M, pH 8.0; Ambion, cat. no. AM9261)
- Proteinase K (20 mg/mL; NEB, cat. no. P8107S)
- Phenol/chloroform/isoamyl alcohol (phenol/$ChCl_3$/IAA, 25:24:1, pH 7.9; Ambion, cat. no. AM9730) **! CAUTION** Phenol/$ChCl_3$/IAA (25:24:1) is toxic upon ingestion, inhalation or contact with skin. Handle phenol/$ChCl_3$/IAA (25:24:1) in a chemical fume hood with caution. Protective gloves, lab coat and eye protection should be worn. Phenol/$ChCl_3$/IAA (25:24:1) waste should be discarded in an appropriate container according to local regulations **▲ CRITICAL** This reagent must not be replaced with acidic extraction reagent (pH 4.5), which will remove DNA during extraction. The reagent used in this protocol ensures the extraction of both DNA and RNA.
- NaCl (5 M; Invitrogen, cat. no. AM9759)
- Sodium acetate (3 M, pH 5.2; VWR, cat. no. E521-100ML) **! CAUTION** Sodium acetate can cause skin irritation and serious eye damage. Avoid contact with eyes. Wear protective lab coat and gloves.

- Ethyl alcohol (pure, 200 proof, for molecular biology; Sigma-Aldrich, cat. no. E7023-500ML) **! CAUTION** Ethyl alcohol is highly flammable, both as a liquid and as a vapor. It can cause eye irritation. Avoid inhaling the vapors, mist or gas. Handle with caution and ensure adequate ventilation.
- Phase Lock Gel (heavy, 2 mL; VWR, cat. no. 10847-802)
- Qubit dsDNA BR assay kit (Invitrogen, cat. no. Q32850)
- Qubit dsDNA HS assay kit (Invitrogen, cat. no. Q32854)
- Qubit RNA BR assay kit (Invitrogen, cat. no. Q10211)
- Qubit RNA HS assay kit (Invitrogen, cat. no. Q32855)
- 2-Log DNA ladder (NEB, cat. no. N3200S)
- Exonuclease I (NEB, cat. no. M0293L)
- T4 DNA polymerase (3 U/μL; NEB, cat. no. M0203L)
- Tween 20 (Fisher Scientific, cat. no. BP337-500)
- Dynabeads MyOne streptavidin C1 (Invitrogen, cat. no. 65002)
- SuperScript III Reverse Transcriptase (kit includes SuperScript III Reverse Transcriptase 200 U/μL, 10 mM dNTP mix, 5× first-strand buffer and 0.1 M DTT; ThermoFisher, cat. no. 18080044)
- T4 polynucleotide kinase (PNK) reaction buffer (10×; NEB, cat. no. B0201S)
- Hydrochloric acid (HCl; 11.6–12 M; Sigma-Aldrich, cat. no. H1758-100ML) **! CAUTION** HCl can cause severe skin burns and eye damage. Wear protective lab coat, gloves, eye protection and face protection. Use in a well-ventilated area.
- Sodium hydroxide solution (NaOH; 10 M in H_2O; Sigma-Aldrich, cat. no. 72068-100ML) **! CAUTION** 10 M NaOH can cause severe skin burns and eye damage. Wear protective lab coat, gloves, eye protection and face protection. Handle with caution.
- ATP (10 mM; NEB, cat. no. P0756L)
- CircLigase ssDNA Ligase (100 U/μL; kit includes CircLigase 10× reaction buffer, 1 mM ATP and 50 mM $MnCl_2$; Epicentre, cat. no. CL4115K)
- BamHI-HF (20 U/μL; NEB, cat. no. R3136S)
- NEBNext High-Fidelity 2× PCR Master Mix (NEB, cat. no. M0541S)
- MinElute PCR Purification Kit (Qiagen, cat. no. 28006)
- MinElute Gel Extraction Kit (Qiagen, cat. no. 28604)
- RNA 6000 Pico Reagents Part I (Agilent Technologies, cat. no. 5067-1513)
- RNA 6000 Pico Ladder (Agilent Technologies, cat. no. 5067-1535)
- High Sensitivity DNA Kit(Agilent Technologies, cat. no. 5067-4626)
- NucBlue Fixed Cell ReadyProbes Reagent (Thermo Fisher Scientific, cat. no. R37606)
- DAPI
- Buffer RLT
- QIAquick PCR Purification Kit
- Liquid nitrogen **! CAUTION** Liquid nitrogen can cause severe skin burns and asphyxiation. Wear protective lab coat, gloves, eye protection and face protection. Use in a well-ventilated area.

iMARGI linkers and library PCR primers

- Linker top strand: 5′/Phos/NNAGATCGGAAGAGCGTCGTGTAGGGAGGATCCGTTCAGACG TGTGCTCTTCC*GA/iBiodT/CT3′ ('iBiodT' stands for internal biotin modification on dT) (Integrated DNA Technologies, custom order) **▲ CRITICAL** The C* indicates a phosphorothioated C base. This linker top strand should be purified by RNase-free HPLC and lyophilized. Dissolve the linker in UltraPure distilled H_2O and measure the concentration. Store at −20 °C for up to 2 years. The concentration of linker top strand used in our lab is 380 μM.
- Linker bottom strand: 5′/Phos/GATCGGAAGAGCACACGTCTGAACGGATCCTCCCTACACGA CGCTCT3′ (Integrated DNA Technologies, custom order) **▲ CRITICAL** This linker bottom strand should be purified by RNase-free HPLC and lyophilized. Dissolve the linker in UltraPure distilled H_2O and measure the concentration. Store at −20 °C for up to 2 years. The concentration of linker bottom strand used in our lab is 216 μM.
- Cut_oligo sequence: 5′TCGTGTAGGGAGGATCCGTTCAGACGTGTGCTCT/3InvdT/3′ ('3InvdT' stands for inverted dT at the 3′ end) (Integrated DNA Technologies, custom order) **▲ CRITICAL** This Cut_oligo strand should be purified by RNase-free HPLC and lyophilized. Dissolve the Cut_oligo strand in UltraPure distilled H_2O and measure the concentration. Prepare a final concentration of 10 μM Cut_oligo working stock using UltraPure distilled H_2O. Store at −20 °C for up to 2 years.

- NEBNext universal PCR primer for Illumina: 5′AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T3′ (NEB, cat. no. E7335S). Store at −20 °C for up to 1 year.
- NEBNext index primer for Illumina (index primers set 1): 5′CAAGCAGAAGACGGCATACGAGAT[6 bp barcode]GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC-s-T3′ (NEB, cat. no. E7335S). Store at −20 °C for up to 1 year.

Equipment

- Cell and tissue culture dishes (100 × 20 mm, Genesee Scientific, cat. no. 25-202; 150 × 20 mm, Genesee Scientific, cat. no. 25-203)
- BioLite cell culture treated flasks (75 cm^2; Thermo Scientific, cat. no. 12556010)
- Cell culture plates (6-well; VWR, cat. no. 10062892)
- Cell scrapers (Fisher Scientific, cat. no. 087711A)
- Incubator (set incubator to 37 °C, 5% CO_2; Thermo Scientific, cat. no. 13-998-086)
- Conical tubes (15 and 50 mL; VWR, cat. nos. 5250606 and 5250608)
- DNA LoBind Tubes (1.5 and 2 mL; Eppendorf, cat. nos. 022431021 and 022431048)
- Protein LoBind Tubes (5 mL; Eppendorf, cat. no. 0030108302)
- Dounce homogenizer (2 mL; Sigma-Aldrich, cat. no. D8938)
- ThermoMixer (ThermoMixer C; Eppendorf, cat. no. 5382000023)
- Microcentrifuge (Centrifuge 5424R; Eppendorf, cat. no. 5404000138)
- Centrifuge (Centrifuge 5810R; Eppendorf, cat. no. 022625501)
- Thermocycler (Mastercycler nexus gradient; Eppendorf, cat. no. 6331000025)
- E-Gel EX agarose gels (1% (wt/vol); Invitrogen, cat. no. G402001; 2% (wt/vol); Invitrogen, cat. no. G402002)
- E-Gel iBase Power System (Invitrogen, cat. no. 10001123)
- Qubit fluorometer (Invitrogen, cat. no. Q32866)
- NanoDrop Lite Spectrophotometer (Thermo Scientific, cat. no. ND-LITE)
- PCR tubes (0.2 mL; Applied Biosystems, cat. no. N8010540)
- PCR 8-well tube strips (0.2 mL; VWR, cat. no. 89401-570)
- DynaMag magnet (Thermo Fisher Scientific, cat. no. 12301D)
- Magnetic separation stand (Promega, cat. no. Z5342)
- 2100 Bioanalyzer (Agilent, cat. no. G2939BA)
- Blades (we use stainless-steel disposable scalpels for gel cutting; Integra Miltex, cat. no. 4-410)
- Gel imager (GelDoc-It TS2 310 Imager; UVP, cat. no. 71004-578)
- Rocker (UltraCruz 2D rocker; Santa Cruz Biotechnology, cat. no. sc-358757)
- Pipettes
- Filter tips
- Microscope slides
- Cover glasses for slides
- Standard inverted epifluorescence microscope (60× oil objective; numerical aperture (NA) = 1.42; Olympus, model no. IX83)
- Vortex
- Rotator (Mini LabRoller Rotator; Labnet, cat. no. H5500)
- Sequencer (Illumina, model no. HiSeq 4000)
- A computer or server with a minimum of 16 GB of memory running on any of the recommended Linux distributions, including Ubuntu, Debian, Fedora and CentOS ▲ **CRITICAL** A multi-core CPU is advantageous for accelerating the computing time of data processing; 16 or more cores are recommended. Enough hard disk space is necessary. 200 GB of free disk space is required for data processing of the example procedure.

Software

- Docker Engine - Community software (https://docs.docker.com/install/) version 18.09.5 is used here
- iMARGI-Docker software (https://github.com/Zhong-Lab-UCSD/iMARGI-Docker). In this paper, the current version, v.1.1.1, is pulled from Docker-Hub (https://hub.docker.com/r/zhonglab/imargi)
- An example iMARGI dataset (SRR8206679 in the NCBI Sequence Read Archive database; 361.2 million sequencing read pairs)

Reagent setup

1 M glycine

Add 3.74 g of glycine to a 50-mL tube and adjust the volume to 50 mL with UltraPure distilled H_2O. Dissolve glycine by inverting the tube. Use the solution immediately.

0.5 M glycine

Add 1.87 g of glycine to a 50-mL tube and adjust the volume to 50 mL with UltraPure distilled H_2O. Dissolve glycine by inverting the tube. Use the solution immediately.

50× protease inhibitor cocktail

Dissolve one tablet of protease inhibitor cocktail in 1 mL of UltraPure distilled H_2O. Prepare 50-µL aliquots and store at −20 °C for up to 12 weeks.

1× PBS

Dilute 1 mL of 10× PBS with 9 mL of UltraPure distilled H_2O. Store at room temperature (RT; 22–25 °C) for up to 6 months.

0.1 M DTT

Dilute 0.5 mL of 1 M DTT with 4.5 mL of UltraPure distilled H_2O. Store the solution at −20 °C for up to 12 months.

10 mM dATP

Dilute 50 µL of 100 mM dATP (from the dNTP Solution Set) with 450 µL of UltraPure distilled H_2O. Store the solution at −20 °C for up to 12 months.

1 mM NaOH

Dilute 1 mL of 10 mM NaOH solution with 9 mL of UltraPure distilled H_2O. Store at RT for up to 12 months.

1 M HCl

Dilute 1 mL of ~12 M HCl with 11 mL of UltraPure distilled H_2O. Store at RT for up to 12 months.

10% (vol/vol) Tween 20

Mix 10 µL of Tween 20 with 90 µL of UltraPure distilled H_2O. Store at RT for up to 12 months.

70% (vol/vol) ethanol

Mix 35 mL of 100% (vol/vol) ethanol with 15 mL of UltraPure distilled H_2O. Store at RT for up to 6 months.

20% (vol/vol) Nonidet P 40

Mix 20 µL of Nonidet P 40 with 80 µL of UltraPure distilled H_2O. Store at RT for up to 12 months.

1 M NaCl

Dilute 2 mL of 5 M NaCl solution with 8 mL of UltraPure distilled H_2O. Store at RT for up to 12 months.

1× cell lysis buffer

Mix 10 µL of 1 M Tris-HCl, pH 7.5, 10 µL of 1 M NaCl, 10 µL of 20% (vol/vol) Nonidet P 40, 20 µL of 50× protease inhibitor cocktail and 950 µL of UltraPure distilled H_2O. Use the cell lysis buffer immediately. The final composition of 1× cell lysis buffer is 10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 0.2% (vol/vol) Nonidet P 40 and 1× protease inhibitor cocktail.

1× CutSmart buffer

Dilute 1 mL of 10× CutSmart buffer with 9 mL of UltraPure distilled H_2O. Store the buffer at −20 °C for up to 6 months.

0.5% (wt/vol) SDS
Mix 5 µL of 20% (wt/vol) SDS, 100 µL of 1× CutSmart buffer and 95 µL of UltraPure distilled H_2O. Use the buffer immediately.

10% (vol/vol) Triton X-100
Mix 20 µL of Triton X-100, 100 µL of 1× CutSmart buffer and 80 µL of UltraPure distilled H_2O. Use the buffer immediately.

5× PNK phosphatase buffer, pH 6.5
Mix 350 µL of 1 M Tris-HCl, pH 6.5, 50 µL of 1 M $MgCl_2$, 100 µL of 0.1 M DTT and 500 µL of UltraPure distilled H_2O. Store the buffer at −20 °C for up to 6 months.

PNK wash buffer
Mix 1 mL of 1 M Tris-HCl, pH 7.5, 0.5 mL of 1 M $MgCl_2$, 0.1 mL of 100% (vol/vol) Tween 20 and 48.4 mL of UltraPure distilled H_2O. Store at RT for up to 6 months. The final composition of PNK wash buffer is 20 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$ and 0.2% (vol/vol) Tween 20.

Extraction buffer
Mix 25 µL of 1 M Tris-HCl, pH 7.5, 25 µL of 20% (wt/vol) SDS, 1 µL of 0.5 M EDTA, 10 µL of 5 M NaCl, 25 µL of proteinase K and 414 µL of UltraPure distilled H_2O to make a final volume of 500 µL. Freshly prepare before extraction and reverse crosslinking step. The final composition of the extraction buffer is 50 mM Tris-HCl, pH 7.5, 1% (wt/vol) SDS, 1 mM EDTA, 100 mM NaCl and 1 mg/mL proteinase K.

2× B&W buffer
Mix 300 µL of 1 M Tris-HCl, pH 7.5, 60 µL of 0.5 M EDTA, 12 mL of 5 M NaCl, 60 µL of 10% (vol/vol) Tween 20 and 55.8 mL of UltraPure distilled H_2O. Store at RT for up to 6 months. The final composition of the 2× B&W buffer is 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 2 M NaCl and 0.02% (vol/vol) Tween 20.

1× B&W buffer
Dilute 10 mL of 2× B&W buffer with 10 mL of UltraPure distilled H_2O. Store at RT for up to 6 months.

Denaturing buffer
Mix 100 µL of 1 M NaOH, 1 µL of 100 mM EDTA and 899 µL of UltraPure distilled H_2O. Store at RT for up to 12 months. The final composition of the denaturing buffer is 0.1 M NaOH and 0.1 mM EDTA.

High-salt biotin wash buffer
Mix 3 mL of 1 M Tris-HCl, pH 7.5, 240 mL of 5 M NaCl, 600 µL of 0.5 M EDTA, 600 µL of Tween 20 and 55.8 mL of UltraPure distilled H_2O. Store at 4 °C for up to 1 month. The final composition of the high-salt biotin wash buffer is 10 mM Tris-HCl, pH 7.5, 4 M NaCl, 1 mM EDTA and 0.2% (vol/vol) Tween 20.

Cell culture medium
For HEK293T cells, supplement DMEM with a final concentration of 10% (vol/vol) FBS and 1% penicillin–streptomycin. Store at 4 °C for up to 1 month.

Procedure

Crosslinking of cells in solution ● Timing 2 h

1 Grow cells in the appropriate culture medium to 90% confluency. Three to five million cells are enough to generate one iMARGI library. However, we recommend starting with more cells (i.e., ~2×10^7), crosslinking and making aliquots of 1×10^7 cells per tube to perform technical replicates, if necessary. For cells cultured in suspension or loosely adherent cells (e.g., HEK293T cells), we recommend crosslinking in a tube to minimize the loss of cells (option A). For adherent cells that attach well, we recommend crosslinking on a plate (option B).

(A) **Crosslinking cells in a tube**

(i) For adherent cells only, aspirate the medium from the tissue culture plates. Rinse the cells grown on cell culture plates with 7 mL of 1× PBS and put the plates on ice. Scrape the cells into the PBS that is already in the plates and transfer to a 50-mL tube. Wash the plates with 7 mL of 1× PBS once and combine the residual cells with the cells in the tube. Cell counting is recommended at this step. For suspension cells, start from Step 1A(ii).

(ii) Spin down detached adherent cells or cells in suspension at 500*g* for 3 min at 4 °C. Discard the supernatant. Resuspend the cells in 15 mL of 1× PBS.

(iii) Add 1 mL of fresh 16% (wt/vol) formaldehyde to a final concentration of 1% (wt/vol) and fix the cells at RT for 10 min with rotation.

(iv) Quench the crosslinking reaction by adding 4 mL of 1 M glycine to a final concentration of 0.2 M.

(v) Incubate at RT for 10 min with rotation and then on ice for 10 min.

(vi) Spin crosslinked cells at 2,000*g* for 5 min at 4 °C.

(vii) Discard the supernatant into an appropriate collection container by pipetting.

(viii) Rinse the cells once with 15 mL of ice-cold 1× PBS without resuspending the cell pellet.

(ix) Spin at 2,000*g* for 5 min at 4 °C to pellet the cells. Carefully discard the supernatant by aspirating.

PAUSE POINT Cells can be flash-frozen in liquid nitrogen and stored at −80 °C for at least 1 year. If frozen, gently thaw the cells on ice before proceeding to the rest of the protocol.

(B) **Crosslinking cells on a plate**

(i) Aspirate the medium from the tissue culture plates. Rinse the cells grown on cell culture plates with 5 mL of 1× PBS.

(ii) Add 7 mL of freshly made 1% (wt/vol) formaldehyde to the plates and incubate the cells at RT for 10 min with gentle rocking on a rocker.

(iii) Quench the crosslinking reaction by adding 4.6 mL of 0.5 M glycine to a final concentration of 0.2 M.

(iv) Incubate the cells at RT for 10 min with rotation and then on ice for another 10 min.

(v) Remove the solution by pipetting into an appropriate liquid waste container.

(vi) Place the plate on ice and rinse the cells twice with 5 mL of ice-cold 1× PBS.

(vii) Add 5 mL of 1× PBS to the plate. Scrape the cells from the plate and transfer to a 15-mL conical tube.

(viii) Add 5 mL of 1× PBS to the plate and transfer the cells left in the plate to the same 15-mL conical tube.

(ix) Spin at 2,000*g* for 5 min at 4 °C to pellet the cells. Carefully discard the supernatant by aspirating.

PAUSE POINT The cells can be flash-frozen in liquid nitrogen and stored at −80 °C for at least 1 year. If frozen, gently thaw the cells on ice before proceeding to the rest of the protocol.

Cell lysis and DNA fragmentation ● Timing 3 h plus overnight reaction

2 Resuspend one aliquot of crosslinked cells (~1 × 10^7 cells) from Step 1A(ix) or 1B(ix) in 1 mL of freshly prepared cold cell lysis buffer and incubate on ice for 15 min.

3 Transfer the mixture to a 2-mL Dounce homogenizer.

4 Homogenize the cells on ice with pestle A (from Dounce homogenizer kit). Slowly move the pestle up and down 10 times, incubate on ice for 1 min and then perform 10 more strokes. Incubate the cells on ice for a further 10 min.

▲CRITICAL STEP Pestle A is used in this step to homogenize cells. Pestle A has a looser fitting, which can help with the lysis of the cells while keeping the nuclei intact.

5 Weigh two empty 1.5-mL Eppendorf tubes.

6 Transfer the cell lysis mixture to one of the 1.5-mL tubes weighed in Step 5.

7 Centrifuge at 2,500*g* at 4 °C for 2 min to pellet the nuclei; discard the supernatant carefully.

8 Carefully resuspend the pelleted nuclei in 500 µL of 1× CutSmart buffer. Split 250 µL of the mixture with nuclei into the other 1.5-mL tube weighed in Step 5.

9 Centrifuge both tubes of nuclei from Step 8 at 2,500*g* at 4 °C for 1 min to pellet the nuclei; discard the supernatant carefully. After splitting the nuclei mixture in Step 8, the nuclei pelleted in this step

Box 1 | Checkpoint 1, for imaging nuclei integrity ● Timing 45 min

This checkpoint is recommended for researchers performing an iMARGI experiment for the first time. In this checkpoint, the integrity of nuclei will be examined under a microscope.

Procedure

1 Resuspend the remaining aliquot of pelleted nuclei (~5 × 10^6 cells after splitting in Step 8) in 500 µL of cold 1× PBS in a 1.5-mL tube.
2 Add one drop of NucBlue fixed-cell stain and incubate at RT in the dark for 10 min. After the incubation, pellet the nuclei by centrifugation at 2,000*g* for 5 min at 4 °C.
3 Resuspend the nuclei in 500 µL of cold 1× PBS and then dispense 10 µL of the nuclei mixture onto a microscope slide and cover it with a cover glass.
4 Image the slide with a 40× air or 60× oil objective with standard DAPI emission/excitation filters on a standard inverted epifluorescence microscope. Specifically, we used an Olympus IX83 inverted microscope with a 60× oil objective (NA = 1.42) with a 50-ms exposure at 1× gain. This DAPI-stained aliquot of nuclei can be subjected to Steps 10–19 in parallel with the other aliquot of nuclei from Step 9.
5 Repeat this checkpoint after permeabilization of the nuclear membrane (after Step 17) and DNA fragmentation (after Step 24) for the DAPI-stained aliquot of nuclei. The nuclei are considered intact if the DAPI signal is confined to the nuclei and overlaps with the differential interference contrast (DIC) images of the nuclei (Fig. 3a–c).

are from ~5 × 10^6 cells, which will be used in the rest of the protocol. The other half of nuclei pellet can be processed at the same time as a replicate, if necessary, or used to examine nuclei integrity as described in Box 1 (Checkpoint 1).

10 Weigh the 1.5-mL tube containing the nuclei pellet from Step 9 after discarding the supernatant.
11 Estimate the weight of nuclei pellet by subtracting the weight of the empty tube in Step 5 from the total weight of the tube and pellet in Step 10.
12 Gently resuspend the nuclei from Step 10 in 0.5% (wt/vol) SDS following this formula: volume of SDS (in microliters)/weight of nuclei (in milligrams) = 3. For example, if the measured weight of the nuclei pellet is 15 mg, add 45 µL of 0.5% (wt/vol) SDS to the nuclei pellet.
13 Incubate the SDS–nuclei mixture at 62 °C for 10 min in a ThermoMixer, shaking at 800 r.p.m.
14 Quench the SDS reaction by adding 1× CutSmart buffer and 10% (vol/vol) Triton X-100. The final SDS concentration will be 0.1% (wt/vol) and the final Triton X-100 concentration will be 1% (vol/vol).
15 Incubate at 37 °C for 15 min in a ThermoMixer, shaking at 800 r.p.m.
16 Pellet the nuclei by centrifuging at 2,500*g* at 4 °C for 1 min. Discard the supernatant carefully.
? TROUBLESHOOTING
17 Wash the nuclei by adding 500 µL of 1× CutSmart buffer to gently resuspend the nuclei, centrifuge at 2,500*g* at 4 °C for 1 min and discard the supernatant. Repeat the wash once.
? TROUBLESHOOTING
18 Resuspend the nuclei in 300 µL of the following chromatin digestion mix:

Reagent	Volume (µL)	Final concentration
H_2O	198	
10× CutSmart buffer	30	1×
AluI (10 U/µL)	70	2.3 U/µL
RNasin Plus (40 U/µL)	2	0.3 U/µL
Total volume	300	

19 Digest chromatin overnight at 37 °C in a ThermoMixer with shaking at 850 r.p.m. (pause 30 s for every 1 min shaking).

RNA fragmentation, preparation of RNA and DNA ends for ligation, and ligation of linker to RNA ● Timing 4 h plus overnight reaction

20 Dilute RNase I tenfold in 1× PBS.
21 Add 1 µL of diluted RNase I to the chromatin digestion mix from Step 19 and incubate at 37 °C for 3 min.
22 Place the tube on ice for at least 5 min.
23 Pellet the nuclei at 2,500*g* for 1 min at 4 °C. Discard the supernatant carefully or transfer it to a new 1.5-mL tube for nuclei integrity examination (Checkpoint 2, described in Box 2).

Box 2 | Checkpoint 2, for quantitative evaluation of nuclei integrity ● Timing 4 h

This optional checkpoint is recommended for researchers performing an iMARGI experiment for the first time. This checkpoint is complementary to the checkpoint performed at Step 9 (Box 1). In this checkpoint, the concentrations of RNA and DNA in nuclei and in the supernatant will be measured; these measurements will be used to quantitatively evaluate nuclei integrity.

Procedure

1 Remove the supernatants from Steps 23, 28, 32, 50 and 56 after the reactions in those steps have been completed (Fig. 3d) and transfer them to new tubes; add 10 μL of proteinase K, and incubate the tubes at 65 °C for 2 h to reverse the crosslinking.
2 Purify the nucleic acids using 50 μL of silane beads and elute with 50 μL of UltraPure distilled H_2O, but do not remove it from the beads. Refer to Steps 37–46 for silane bead purification. The sample volume will be the volume of the supernatant taken out after Steps 23, 28, 32, 50 and 56, which will be approximately the same as the reaction volume in Steps 18, 26, 30, 47 and 52.
3 Split the sample–bead mixture in half (into two tubes): treat one tube with 1 μL of RNase A to remove the RNA in the sample and the other tube with 1 μL of TURBO DNase + 2.5 μL of TURBO 10× reaction buffer to remove the DNA in the sample.
4 Incubate each tube at 37 °C for 30 min.
5 Purify the nucleic acids using the silane beads already in the mixture. Elute with 20 μL of UltraPure distilled H_2O and transfer the purified DNA and RNA to new tubes.
6 Measure the concentration of DNA and RNA, respectively, in each tube using a Qubit dsDNA HS assay kit or a Qubit RNA HS assay kit. Calculate the amount of purified sDNA and sRNA. Because the nucleic acids were split into two tubes after the first elution, one should multiply the amount of sDNA or sRNA by two to obtain the total amount of sDNA or sRNA leaked from nuclei.
7 Compare the amounts of sDNA and sRNA after each step with the amounts of the final nuclei-retained DNA (nDNA) and RNA (nRNA) from Step 72, by taking the ratios of these amounts (Fig. 3e,f). Small ratios indicate non-leaking nuclei, whereas large ratios indicate broken nuclei (see Table 3 for thresholds).

24 Wash the nuclei with 300 μL of PNK wash buffer, centrifuge at 2,500*g* at 4 °C for 1 min. Repeat the wash once.

25 *Checkpoint 3.* This step examines the efficiencies of chromatin fragmentation and RNA fragmentation. Transfer 10 μL of the nuclei pellet from Step 23 to a new tube. Add 100 μL of extraction buffer and then incubate the mixture at 65 °C for 2 h to reverse crosslinking. Purify the nucleic acids extracted from the nuclei, using 50 μL of silane beads. Elute the nucleic acids using 50 μL of UltraPure distilled H_2O, but do not remove the beads. Split the nucleic acid–bead mixture in half: add 1 μL of RNase A to one half; add 2.5 μL of TURBO 10× reaction buffer and 1 μL of TURBO DNase to the other half. Incubate at 37 °C for 30 min. Purify the nucleic acids using silane beads that are already in the mixture. Elute DNA or RNA, respectively. Incubate the beads with 15 μL of UltraPure distilled H_2O and elute as described in Step 46. Check the chromatin fragmentation pattern by loading eluted DNA onto a precast 1% agarose E-Gel. A smear of DNA fragments between ~100 bp and ~ 2 kb is considered to be efficient chromatin fragmentation (Fig. 4a). Check RNA fragmentation efficiency by loading eluted RNA on a Bioanalyzer. The degree of RNA fragmentation is deemed appropriate when most fragments range from 50 to 1,000 nt (Fig. 4b).
? TROUBLESHOOTING

26 Prepare the RNA 3′-end dephosphorylation reaction mix as follows:

Reagent	Volume (μL)	Final concentration
H_2O	148	
5× PNK phosphatase buffer, pH 6.5	40	1×
T4 PNK (10 U/μL)	10	0.5 U/μL
RNasin Plus (40 U/μL)	2	0.4 U/μL
Total volume	200	

27 Resuspend the nuclei from Step 24 in the reaction mix and incubate in a ThermoMixer at 37 °C for 30 min with shaking at 800 r.p.m.
28 Pellet the nuclei at 2,500*g* for 1 min at 4 °C. Discard the supernatant or transfer it to a new 1.5-mL tube for nuclei integrity examination as described in Box 2.
29 Wash the nuclei twice with 300 μL of PNK wash buffer on ice.

30 Prepare a DNA dA-tailing reaction mix as follows:

Reagent	Volume (μL)	Final concentration
H_2O	164	
10× NEBuffer 2	20	1×
Klenow fragment (3′–5′ exo-; 5 U/μL)	12	0.3 U/μL
10 mM dATP	2	0.1 mM
10% (vol/vol) Triton X-100	2	0.1 % (vol/vol)
Total	200	

31 Resuspend the nuclei pellet in the reaction mix and incubate in a ThermoMixer at 37 °C for 30 min with shaking at 800 r.p.m..

32 Pellet the nuclei at 2,500*g* for 1 min at 4 °C. Discard the supernatant or transfer it to a new 1.5-mL tube for nuclei integrity examination as described in Box 2.

33 Wash twice with 300 μL of PNK wash buffer; discard the supernatant from the second wash and keep the pelleted nuclei on ice.

34 Prepare the following adenylation reaction mix in a 0.2-mL PCR tube for the linker top strand:

Reagent	Volume (μL)	Final concentration
H_2O	3.5	
10× 5′ DNA adenylation reaction buffer[a]	3	1×
1 mM ATP[a]	3	0.1 mM
Linker top strand oligo (380 μM)	2.5	30 μM
50 μM Mth RNA ligase[a]	18	30 μM
Total volume	30	

[a]Included in 5′ DNA Adenylation Kit.

35 Incubate the reaction mixture in a thermocycler at 65 °C for 1 h and then at 85 °C for 5 min to inactivate the enzyme.

36 Add 4.2 μL (900 pmol) of the linker bottom strand (216 μM) to the adenylated linker top strand and incubate the samples in a thermocycler using the following program to anneal the top and bottom strands: 95 °C for 2 min; then 71 cycles of 20 s, starting from 95 °C and decrease the temperature by 1 °C each cycle, down to 25 °C, and hold at 25 °C.
▲ **CRITICAL STEP** Freshly prepare the adenylated linker top strand and anneal to the bottom strand in each experiment. The volumes of linker top strand and bottom strand needed in Step 34 depend on the actual concentration of the linker oligos.

37 Purify the annealed linker using 200 μL of silane beads. To do so, first transfer 200 μL of silane bead mixture to a 1.5-mL tube. Place on a magnetic stand, wait for the beads to attach to the magnet, and then discard the stock solution.

38 Resuspend the silane beads with 300 μL of buffer RLT, place the tube on a magnetic stand and wait for the beads to attach to the magnet. Discard the supernatant carefully.

39 Remove the tube from the magnetic stand.

40 Resuspend the beads in 3.5× the original sample volume of buffer RLT. The sample volume is 30 μL of reaction volume from Step 34 plus the 4.2 μL of bottom linker, which is 34.2 μL in total. Add the 34.2 μL of annealed linker sample from Step 36 to the bead suspension.

41 Add 4.5× the original sample volume of isopropanol to the bead–sample mixture. Mix well by inverting the tube five times.

42 Incubate the sample with the beads for 10 min at RT with rotation.

43 Place the tube on a magnetic stand and wait for the beads attach to the magnet. Discard the supernatant carefully.

44 Wash the beads with 300 μL of 70% (vol/vol) ethanol, wait for beads to attach to the magnet and discard the supernatant carefully.

45 Repeat the washing in Step 44 once. Discard the supernatant and let the beads air-dry for 5–10 min.

46 Elute the sample in 50 μL of UltraPure distilled H_2O. Place the tube on a magnetic stand, wait for the beads attach to the magnet and transfer the eluate to a new tube. All the later silane bead

purifications will follow the procedure outlined in Steps 37–46, unless otherwise mentioned. The beads and sample volume may vary and will be specified at each subsequent step.

Checkpoint 4. To check the annealing efficiency of linker top and bottom strands, separately dilute 1 µL each of linker top, linker bottom and annealed double-stranded (ds) linker sample from Step 46 with 19 µL of UltraPure distilled H_2O. Load 20 µL each of diluted linker top, linker bottom and ds linker onto a precast 2% agarose E-Gel. The successfully annealed ds linker will shift to a larger size and run at a higher molecular weight than the single-stranded top and bottom strands (Fig. 5a).

▲ **CRITICAL STEP** The final eluate collected after magnetic separation could be <50 µL. All the eluted linker will be added to the linker–RNA ligation reaction in Step 47.

47 Prepare the following linker–RNA ligation mix in a 2-mL tube:

Reagent	Volume (µL)	Final concentration
H_2O	Make up to 200 µL	
Linker (Step 46)	Entire eluate (≤50 µL)	
10× T4 RNA ligase reaction buffer	20	1×
50% (wt/vol) PEG 8000	80	20% (vol/vol)
10% (vol/vol) Triton X-100	2	0.1% (vol/vol)
RNasin Plus (40 U/µL)	2	0.4 U/µL
T4 RNA ligase 2, truncated KQ (200 U/µL)	10	10 U/µL
Total volume	200	

48 Resuspend the nuclei from Step 33 in the linker–RNA ligation mix and incubate at 22 °C for 6 h with intermittent mixing (900 r.p.m.; mix for 30 s, stop for 15 s and start mixing again) and then at 16 °C overnight with shaking at 850 r.p.m. in a ThermoMixer.

▲ **CRITICAL STEP** The reaction mix with PEG is very viscous. Pipette carefully to avoid residual reaction mix on the tip. We recommend using a 2-mL tube to facilitate mixing. Therefore, the nuclei need to be transferred from a 1.5-mL tube to a 2-mL tube at this step. Take 100 µL of reaction mix from Step 47 and add it to the nuclei pellet from Step 33. Resuspend the mixture and transfer all the nuclei to the 2-mL tube for reaction.

Proximity ligation ● Timing 45 min plus overnight reaction

49 Stop the linker–RNA ligation reaction by adding 20 µL of 0.5 M EDTA and incubation at 16 °C for 5 min.

50 Pellet the nuclei and discard the supernatant or transfer it to a new tube for examination of nuclei integrity as described in Box 2.

51 Wash five times with 500 µL of PNK wash buffer to remove free linkers as described in Step 17.

▲ **CRITICAL STEP** The nuclei may not pellet at the bottom of the 2-mL tube very securely. Carefully remove the supernatant after each wash to avoid losing nuclei.

52 Prepare the following proximity ligation mixture but do not add the T4 DNA ligase until Step 53:

Reagent	Volume (µL)	Final concentration
H_2O	1,660	
10× DNA ligase reaction buffer	200	1×
10 % (vol/vol) Triton X-100	20	0.1% (vol/vol)
BSA (20 mg/mL)	100	1 mg/mL
T4 DNA ligase (2,000 U/µL)	4	4 U/µL
RNasin Plus (40 U/µL)	16	0.5 U/µL
Total volume	2,000	

53 Before adding the T4 DNA ligase, resuspend the nuclei in the reaction mix' afterward, add the T4 DNA ligase.

▲ **CRITICAL STEP** The nuclei need to be transferred from the 2-mL tube to a 5-mL tube at this step. Take 500 µL of the reaction mix (excluding the T4 DNA ligase) from Step 52 and add it to the 2-mL tube with the nuclei pellet. Resuspend the nuclei and transfer all the nuclei to the 5-mL tube. Add the T4 DNA ligase to the 5-mL tube after transferring all the nuclei to the 5-mL tube.

54 Incubate the reaction mixture at 16 °C overnight with intermittent mixing in a ThermoMixer (650 r.p.m.; 30 s/min)
▲ **CRITICAL STEP** The proximity ligation is performed in a diluted condition in a 5-mL tube to minimize random ligation.

Reverse crosslinking and DNA/RNA extraction ● Timing 5 h

55 Stop the proximity ligation reaction by adding 200 µL of 0.5 M EDTA and incubation at 16 °C for 15 min.
56 Pellet the nuclei at 2,500*g* at 4 °C for 1 min. Discard the supernatant or transfer it to a new tube for nuclei integrity examination as described in Box 2.
57 Wash the nuclei pellet with 500 µL of 1× PBS, centrifuge at 2,500*g* at 4 °C for 1 min and discard the supernatant. Repeat the wash once.
58 Resuspend the nuclei in 250 µL of extraction buffer to extract the nucleic acid and to reverse the crosslinks.
▲ **CRITICAL STEP** The extraction buffer can form a white precipitate if put on ice. We recommend using the extraction buffer immediately after preparation.
59 Incubate the mixture at 65 °C on a ThermoMixer for 3 h with shaking at 800 r.p.m.
60 Add an equal volume of phenol/$ChCl_3$/IAA (25:24:1) (pH 7.9) to the reverse crosslinking mixture. Vortex for 1 min until the mixture turns white.
61 Pre-spin a tube of Phase Lock Gel at 1,500*g* for 1 min at RT. Add the mixture from Step 60 to the tube of Phase Lock Gel.
▲ **CRITICAL STEP** Phase lock gel could stick to the walls of the tube. Spin the tube of Phase Lock Gel before use.
62 Centrifuge the tube of Phase Lock Gel at 12,000*g* for 5 min at RT and then transfer the top aqueous phase to a new 1.5-mL tube.
▲ **CRITICAL STEP** Avoid touching the Phase Lock Gel with the pipette tip when transferring the top aqueous phase.
63 Add 250 µL of phenol/$ChCl_3$/IAA (25:24:1) (pH 7.9) to the 1.5-mL tube from Step 62. Vortex for 1 min until the mixture turns white.
64 Add the mixture to a new tube of Phase Lock Gel. Centrifuge at 12,000*g* for 5 min at RT, and then transfer the aqueous phase to a new 1.5-mL tube.
65 Perform ethanol precipitation by adding a 1/10th volume of 3 M sodium acetate (pH 5.2) and a 3× volume of ice-cold 100% (vol/vol) ethanol to the 1.5-mL tube from Step 64. Mix well by inverting the tube three times.
66 Incubate at −80 °C for 30 min.
■ **PAUSE POINT** The ethanol precipitation mix can also be incubated at −20 °C overnight.
67 Centrifuge the tube at 16,000*g* for 30 min at 4 °C.
68 Remove and discard the supernatant and gently add 1 mL of cold 70% (vol/vol) ethanol to the precipitated nucleic acid pellet. Do not disturb the pellet.
69 Centrifuge again at 16,000*g* for 10 min in 4 °C.
70 Remove and discard the supernatant and air-dry the pellet for 5–10 min.
71 Carefully resuspend the pellet in 50 µL of UltraPure distilled H_2O.
▲ **CRITICAL STEP** The dried nucleic acid pellet may become transparent and invisible in the tube. Add UltraPure distilled H_2O to the tube gently. Completely dissolve the nucleic acid pellet by pipetting.
72 Measure the DNA and RNA concentrations using a Qubit dsDNA BR assay kit and a Qubit RNA BR assay kit, respectively. A typical yield of DNA is ~10–15 µg for ~5 × 10^6 cells; the amount of RNA depends on the cell type.
■ **PAUSE POINT** Purified nucleic acids can be stored at −20 °C for up to 6 months.

Removal of biotin from unligated linkers ● Timing 3 h

73 Separate the purified nucleic acids into multiple 0.2-mL PCR tubes (PCR strips) containing a maximum of 8 µg of nucleic acid in each tube. Prepare the reaction mixture in each tube as follows:

Reagent	Volume (1×) (µL)	Final concentration
H_2O	To 144 µL	
10× NEBuffer 2	15	1×
BSA (20 mg/mL)	1	0.1 mg/mL
Exo I (20 U/µL)	5	0.7 U/µL

Table continued

(continued)

Reagent	Volume (1×) (μL)	Final concentration
RNasin Plus (40 U/μL)	2	0.6 U/μL
Nuclei acids (DNA + RNA; Step 72)	Variable	8 μg maximum each tube
Total volume	144	

74 Incubate the reactions in PCR strips at 37 °C for 30 min in a thermocycler.

75 Add the following mixture to each PCR tube to a final volume of 150 μL:

Reagent	Volume (1×) (μL)	Final concentration
10 mM dATP	1.5	0.1 mM
10 mM dGTP	1.5	0.1 mM
T4 DNA polymerase (3 U/ μL)	3	0.06 U/ μL
Total volume	6	

76 Incubate at 12 °C for 2 h in a thermocycler.

77 Combine the contents of all the tubes in the PCR strip into a single 1.5-mL tube and add a final concentration of 50 mM EDTA to stop the reaction.

PAUSE POINT The sample mixture can be stored at −20 °C overnight.

Biotin pull-down of RNA–DNA chimeric sequences, reverse transcription and ssDNA circularization ● Timing 5–6 h plus overnight reaction

78 Add 200 μL of streptavidin C1 beads to a 15-mL tube.

79 Wash the beads with 300 μL of 1× B&W buffer, place the tube on a DynaMag magnet for 1 min, and discard the supernatant. Repeat the wash three times.

80 Resuspend the beads with a volume of 2× B&W buffer equal to the volume of the sample from Step 77. Add the sample from Step 77 to the beads and incubate at RT for 30 min with gentle rotation.

81 Pull down the beads with a magnet and extensively wash the beads seven times, using 7 mL of high-salt biotin wash buffer for each wash.

▲CRITICAL STEP Washing the streptavidin beads with a high-salt biotin wash buffer is important for removing nonspecific nucleic acids that bind to the beads. Incubate the high-salt biotin wash buffer with the beads for 2 min on a rotator during each wash. Then place the tubes on a magnet, wait for the solution to turn clear, and discard the supernatant.

82 Resuspend the beads with 1mL of PNK wash buffer and transfer the bead mixture to a new 2-mL tube.

83 Put the 2-mL tube from Step 82 on a magnet and remove and discard the PNK wash buffer.

▲CRITICAL STEP High-salt biotin wash buffer left with the beads may affect later enzymatic reactions. Therefore, washing the beads with PNK wash buffer is essential to removing the residual high-salt biotin wash buffer completely before the enzymatic reactions.

84 Prepare the reverse transcription reaction mix as follows:

Reagent	Volume (μL)	Final concentration
H_2O	22	
5× First-strand buffer[a]	8	1×
10 mM dNTP[a]	2	0.5 mM
100 mM DTT[a]	2	5 mM
RNasin Plus (40 U/μL)	2	2 U/μL
SuperScript III Reverse Transcriptase (200 U/μL)[a]	4	20 U/μL
Total volume	40	

[a]Included in SuperScript III Reverse Transcriptase kit.

85 Resuspend the beads in the reverse transcription mix and incubate at 50 °C for 1 h with shaking at 800 r.p.m. in a ThermoMixer.

86 Pull down the beads with a magnet and discard the supernatant.
87 Wash the beads twice with 300 µL of 1× B&W buffer.
88 Wash the beads once with 300 µL of PNK wash buffer.
89 Prepare the DNA 5' end phosphorylation mixture as follows:

Reagent	Volume (µL)	Final concentration
H_2O	74	
10× T4 PNK reaction buffer, pH 7.6[a]	10	1×
T4 PNK (10 U/µL)	5	0.5 U/µL
10% (vol/vol) Triton X-100	1	0.1% (vol/vol)
10 mM ATP	10	1 mM
Total	100	

[a]Provided with T4 PNK.

90 Resuspend the beads in the phosphorylation mixture and incubate at 37 °C for 1 h.
91 Wash the beads with 300 µL of 1× B&W buffer twice. Remove and discard the 1× B&W buffer.
92 Resuspend the beads in 100 µL of denaturing buffer to dissociate single-stranded cDNA–DNA (ssDNA) chimeric sequences from the beads.
93 Incubate the tube at RT for 15 min, pull down the beads using a magnet and transfer the supernatant to a new tube.
▲ **CRITICAL STEP** After incubation in the denaturing buffer, ssDNA is dissociated from its biotinylated complementary strand and is released into the supernatant. Therefore, after magnetic separation, the supernatant should be collected.
94 Incubate the collected supernatant at 98 °C for 20 min to completely hydrolyze the complementary RNA strand and then neutralize the supernatant with 10 µL of 1 M HCl and 10 µL of 1 M Tris-HCl, pH 7.5.
95 Purify the resulting 120 µL of ssDNA using 100 µL of silane beads.
96 Elute the ssDNA from the beads with 20 µL of UltraPure distilled H_2O.
97 Prepare the ssDNA circularization mix in a 0.2-mL PCR tube as follows:

Reagent	Volume (µL)	Final concentration
ssDNA (from Step 96)	15	
CircLigase 10× reaction buffer[a]	2	1×
1 mM ATP[a]	1	0.05 mM
50 mM $MnCl_2$[a]	1	2.5 mM
CircLigase ssDNA Ligase[a] (100 U/µL)	1	5 U/µL
Total volume	20	

[a]From CircLigase kit.

▲ **CRITICAL STEP** All the ssDNA will be added to the mixture in Step 97. On the basis of our experience, 15 µL of ssDNA can be collected after eluting with 20 µL of UltraPure distilled H_2O.
98 Incubate the reaction mixture at 60 °C for 4 h in thermocycler and then heat-inactivate the enzyme by incubation at 80 °C for 10 min. Hold at 4 °C.
■ **PAUSE POINT** The reaction can be held at 4 °C in the thermocycler overnight.

Cut_oligo annealing, BamHI digestion and library amplification ● Timing 4 h

99 Prepare the following oligo-annealing mix:

Reagent	Volume (µL)	Final concentration
H_2O	23	
10× CutSmart buffer	3	1×
10 µM Cut_oligo	1	0.2 µM
Total volume	27	

100 Add the annealing mix to the 20 µL of circularized ssDNA from Step 98.

101 Anneal the oligos in a thermocycler using the same annealing program as described in Step 36.

102 Add 3 µL of BamHI-HF to the oligo-annealing mixture and incubate for 1 h at 37 °C in a thermocycler to linearize the DNA.

103 Purify 50 µL of the linearized DNA, using 50 µL of silane beads, and elute with 25 µL of UltraPure distilled H_2O.

104 Optimization of the number of PCR cycles used for library amplification is important for maximizing the complexity of iMARGI libraries and minimizing the PCR or amplification bias. We recommend optimizing the PCR cycle number by setting up a 50-µL test PCR reaction mix using 5 µL of linearized DNA, 25 µL of NEBNext High-Fidelity 2× PCR Master Mix, 1 µL of universal primer (10 µM), 1 µL of index primer (10 µM) and 18 µL of H_2O. Divide the PCR reaction mix into five 10-µL aliquots for PCR cycle screening. Five different cycle numbers (e.g., 10, 13, 15, 18, and 22 cycles) can be tested. Perform PCR as described in Step 107, but vary the cycle numbers.
? TROUBLESHOOTING

105 Determine the appropriate PCR cycle number by running the PCR products on an agarose gel or a Bioanalyzer.
Checkpoint 5. To visualize the size distribution of PCR products on a gel, load 1 µL of each PCR product onto a precast 1% agarose E-Gel. PCR cycle numbers resulting in fragments much larger than 1 kb should be avoided. Use the lowest PCR cycle number possible to avoid over-amplification. To check the size distribution of the products on a Bioanalyzer, dilute each PCR product to 50 µL using UltraPure distilled H_2O. Purify the diluted products using a QIAquick PCR Purification Kit according to the manufacturer's instructions and elute the QIAquick spin column with 50 µL of UltraPure distilled H_2O. Check the size distribution of the library by loading 1 µL of purified PCR product onto a Bioanalyzer, using the High Sensitivity DNA Kit. The majority of library after PCR is expected to range from 150 bp to 1,000 bp (Fig. 4c).
? TROUBLESHOOTING

106 After the PCR cycle number is determined in Steps 104 and 105, prepare the library amplification mixture as follows:

Reagent	Volume (µL)	Final concentration
H_2O	40	
Linearized DNA (Step 103)	5	
2× NEBNext High-Fidelity PCR Master Mix	50	1×
Universal primer (10 µM)	2.5	0.25 µM
Index primer (10 µM)	2.5	0.25 µM
Total volume	100	

107 Because the library amplification mixture contains more DNA template than the pilot PCR in Step 104, one can choose the optimized cycle number determined in Step 105 minus one cycle number in the final library generation PCR. Use the following PCR program:

Cycle no.	Denature	Anneal	Extend
1	98 °C, 30 s		
2–12[a]	98 °C, 10 s	65 °C, 30 s	72 °C, 30 s
13			72 °C, 5 min

[a]Use optimized cycle number. In our iMARGI experiments, we used 12 cycles for library generation.

PAUSE POINT Samples can be kept at 4 °C overnight. In our lab, we usually perform Steps 99–113 in 1 d. Therefore, we consider Steps 99–113 to be day 8.

Library size selection ● Timing 2 h

108 Purify the final PCR product using the MinElute PCR Purification kit according to the manufacturer's instructions; elute the MinElute column twice, each time with 12 µL of UltraPure distilled H_2O, thus obtaining ~24 µL in total.

109 Load the resulting ~24 μL of library onto a precast 2% agarose E-Gel to perform size selection. Mix 1 μL of 2-log DNA ladder with 19 μL of UltraPure distilled H_2O and run it on the 2% agarose E-Gel with the library, using E-Gel system.

110 Remove the agarose gel from the E-Gel cast and visualize the bands under UV light in the gel imager. Cut out the gel piece corresponding to a size ranging from 250 to 1,000 bp, collect the gel pieces, and transfer them to a 1.5-mL tube.
! CAUTION Wear a protective lab coat, gloves and an eye shield to protect yourself from UV radiation. Dispose of the rest of the gel in an appropriate container.

111 Extract DNA from the gel pieces, using a MinElute Gel Extraction kit according to the manufacturer's instructions; elute the MinElute column twice (12 μL of UltraPure distilled H_2O each time; ~24 μL in total).

112 Check the concentration of the final library by Qubit dsDNA HS assay kit.

113 *Checkpoint 6*. Run 1 μL of library on a Bioanalyzer, using the High Sensitivity DNA Kit, and check the quality and size distribution of the library (Fig. 4d). The final library concentration should be at least 1 ng/μL if the library average size is within the range from 380 to 410 bp. On the basis of our experience, with the correct number of cells, 12 PCR cycles in Step 107 will be sufficient to generate enough material for sequencing. To determine the minimum number of cells needed to generate an iMARGI library for sequencing, vary the cell number in Step 2 and repeat Steps 2–113. If the final library concentration is at least 1 ng/μL, as measured in Step 112, and the library average size is within the range of 380–410 bp, as measured at this step with the Bioanalyzer, the amount of input cells is sufficient.

Sequencing ● Timing variable, depending on sequencing facility

114 Sequence the library with 100 cycles of paired-end sequencing on an Illumina sequencer. We used a HiSeq 4000 instrument to sequence our libraries. Use a sufficient number of Illumina sequencing lanes and runs to produce 300 million or more read pairs.

115 *Checkpoint 7*. Check sequencing quality with the FastQC software[28]. The following minimum requirements must be reached: (i) in the Per Base Sequence Quality module, the lower quartile of the base quality score must be >10 in every position, and the median quality score must be >25; (ii) in the Per Base Sequence Content module, both the proportion of Cs at the first position and that of Ts at the second position of Read 2 must be >50%; (iii) in the Per Sequence Quality module, the average quality per read must be >27; (iv) in the Per Base N Content module, the proportion of Ns at any position should be <5%; and (v) in the Sequence Length Distribution module, read lengths of all reads must be identical. Do not rely on the other quality metrics in FastQC, because they are not compatible with iMARGI's experimental design. See general troubleshooting strategies for DNA sequencing (Troubleshooting section) if criteria i, iii, iv and v were not reached. See the Troubleshooting section if criterion ii was not reached.
? TROUBLESHOOTING

Data processing ● Timing 14 h

▲ CRITICAL Processing of iMARGI data will be carried out with a standard data-processing pipeline using iMARGI-Docker (https://sysbio.ucsd.edu/imargi_pipeline). The iMARGI sequencing dataset at the NCBI Sequence Read Archive (SRA) with accession no. SRR8206679 (ref. [15]) was used as the placeholder dataset in all commands in this section. Replace this placeholder dataset with your custom dataset.

116 *Checkpoint 8*. Ensure that your computer has 16 GB or more memory and is running one of the following Linux distributions: Ubuntu, Debian, Fedora or CentOS. Make sure that there is enough free space on the hard disk. Each set of 300 million sequencing read pairs requires 200 GB of free space on the hard disk.

117 *Checkpoint 9*. Check whether Docker has been installed by typing the following command into a Terminal window:

```
docker -v
```

If a Docker version is returned, for example, 'Docker version 18.09.5, build e8ff056', Docker has been installed. Otherwise, install Docker with root privilege, using the following command:

```
sudo curl -fsSL https://get.docker.com |sh -
```

118 Add non-root user to the Docker group with the following command and replace the placeholder `demo_user` with the Linux username of the user. Then you need to login again with the added user's account.

```
sudo usermod -aG docker demo_user
```

119 *Checkpoint 10*. Check whether Docker service has been started with the following command:

```
docker info
```

If a set of system information is returned, Docker service has been started; continue to Step 120. If the output shows 'Cannot connect to the Docker daemon', start Docker service with the command `sudo service docker start` on Ubuntu, Debian or Fedora, or with the command `sudo systemctl start docker` on CentOS.

120 Install iMARGI-Docker with the command:

```
docker pull zhonglab/imargi
```

121 Create a working directory. Download the example dataset (SRR8206679) from NCBI SRA and convert this dataset into FASTQ format. Move the reference genome (GRCh38/hg38) into the working directory.

```
mkdir ~/imargi_example
cd ~/imargi_example
mkdir data ref
docker run --rm -t -v ~/imargi_example:/imargi zhonglab/imargi fastq-
dump --gzip --split-3 SRR8206679
mv SRR8206679_*.fastq.gz ./data
wget https://www.encodeproject.org/files/GRCh38_no_alt_analysis_set_
GCA_000001405.15/@@download/GRCh38_no_alt_analysis_-
set_GCA_000001405.15.fasta.gz
gunzip -d GRCh38_no_alt_analysis_set_GCA_000001405.15.fasta.gz
mv GRCh38_no_alt_analysis_set_GCA_000001405.15.fasta ./ref
```

122 Run the iMARGI data-processing pipeline using an iMARGI-Docker container by specifying SRR8206679 as the input sequencing data (with the `-1` and `-2` flags) and GRCh38 as the reference genome (with the `-g` flag).

```
cd ~/imargi_example
mkdir ./output
docker run --rm -t -v ~/imargi_example:/imargi zhonglab/imargi imargi_-
wrapper.sh -r hg38 -N HEK_iMARGI -t 16 -g ./ref/GRCh38_no_alt_analysis_
set_GCA_000001405.15.fasta -1 ./data/SRR8206679_1.fastq.gz -2 ./data/
SRR8206679_2.fastq.gz -o ./output
```

The output files include the mapped RNA–DNA read pairs (final_HEK_iMARGI.pairs.gz). This is a compressed .pairs file. The .pairs file format is a standard file format for storing pairs of genomics locations and is defined by the NIH 4D Nucleome network[29].

123 *Checkpoint 11*. Open the summary output text file (pipelineStats.log) from the previous step. The first line should read 'Sequence mapping QC passed' or 'Sequence mapping QC failed'. Only proceed if sequence mapping QC has been passed. The sequence mapping QC is passed when both of the following two criteria are satisfied: (i) at least 50% of the read pairs had at least one end uniquely mapped to the genome (GRCh38/hg38) and (ii) at least 50% of the uniquely mapped non-duplicate read pairs were retained after the filtering steps. Note that the deduplication and filtering steps are encapsulated within the previous step, and the user does not have to execute specific commands for these operations.

(Optional) Further analysis ● Timing 3 h

124 Count the numbers of intra- and inter-chromosomal interaction read pairs.

```
docker run --rm -t -v ~/imargi_example:/imargi zhonglab/imargi imargi_
stats.sh -i ./output/final_HEK_iMARGI.pairs.gz -o ./output/report_
final_HEK_iMARGI.txt
```

125 Remove the proximal read pairs. The proximal read pairs are defined as the read pairs for which the RNA end and the DNA end are mapped to the same chromosome and are within 200,000 bp. Do not carry out this step if you want to analyze nascent transcripts.

```
docker run --rm -t -v ~/imargi_example:/imargi zhonglab/imargi imar-
gi_distfilter.sh -d 200000 -i ./output/final_HEK_iMARGI.pairs.gz -o ./
output/filter200k_final_HEK_iMARGI.pairs.gz
```

126 Annotate the interaction file (.pairs) with gene information from a GTF annotation file, such as GTF annotations built by GENCODE. The gene annotations corresponding to the mapped RNA end and the mapped DNA end of a read pair will be added to the output file as column 'gene1' and column 'gene2', respectively.

```
cd ~/imargi_example
wget ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/
release_24/gencode.v24.annotation.gtf.gz
gunzip -d gencode.v24.annotation.gtf.gz
mv gencode.v24.annotation.gtf ./ref
docker run --rm -t -v ~/imargi_example:/imargi zhonglab/imargi imar-
gi_annotate.sh -A gtf -a ./ref/gencode.v24.annotation.gtf -i ./out-
put/filter200k_final_HEK_iMARGI.pairs.gz -o ./output/annot_filter200k_
final_HEK_iMARGI.pairs.gz
```

(Optional) Conversion of .pairs file to other file types ● Timing 1 h

127 Convert the .pairs file into a .bedpe file[30] with the following command. Files in .bedpe format can be further analyzed by other software, including bedtools[30].

```
docker run --rm -t -v ~/imargi_example:/imargi zhonglab/imargi imargi_
convert.sh -i ./output/ filter200k_final_HEK_iMARGI.pairs.gz -f bedpe -o
./output/filter200k_final_HEK_iMARGI.bedpe.gz
```

128 Convert the .pairs file into .cool and .mcool files[31] with the following command. The two-dimensional contact matrix in a .mcool file can be visualized by HiGlass[32] as a heatmap.

```
docker run --rm -t -v ~/imargi_example:/imargi zhonglab/imargi imar-
gi_convert.sh -i ./output/ filter200k_final_HEK_iMARGI.pairs.gz -f cool
-o ./output/filter200k_final_HEK_iMARGI.cool
```

129 Convert the .pairs file into a GIVE interaction (.gin) file[33] with the following command. Each read pair in a .gin file can be visualized by GIVE[33] as a line connecting the mapped location of the RNA end (RNA lane, Fig. 6) and the DNA end (DNA lane, Fig. 6).

```
docker run --rm -t -v ~/imargi_example:/imargi zhonglab/imargi imar-
gi_convert.sh -i ./output/ filter200k_final_HEK_iMARGI.pairs.gz -f
give -o ./output/filter200k_final_HEK_iMARGI.gin
```

Troubleshooting

Troubleshooting advice can be found in Table 4 and Fig. 5a–c.

Table 4 | Troubleshooting table

Step	Problem	Possible reason	Solution
16	Obvious loss of nuclei pellets after SDS treatment or disappearance of nuclei pellets and presence of transparent gel-like precipitate at the bottom of the tube	Nuclei are lysed during SDS treatment in Step 12	Reduce the SDS concentration, volume to weight ratio (volume of SDS/weight of nuclei) or SDS treatment time. The recommended range for SDS concentration is 0.2–0.5% (wt/vol). The recommended range for (volume of SDS/weight of nuclei) is 1–5. The recommended treatment time is between 5 and 10 min at 62 °C. Carry out the protocol through Step 16 and observe the nuclei pellet at the bottom of the tube
17	It is difficult to resuspend the nuclei in wash buffer after SDS treatment	Nuclei become sticky after SDS treatment	Gently resuspend the nuclei and break up the nuclei clumps by pipetting. Pipette gently to avoid bursting the nuclei
25	Insufficient DNA fragmentation. Most DNA fragments are >1,000 bp (lane 2, Fig. 5b)	Inefficient SDS treatment in Step 12 or insufficient amount of restriction enzyme AluI used in Step 18	Test chromatin fragmentation efficiencies under different volume ratios (volume of SDS/weight of nuclei), SDS treatment time (Fig. 5b,c) or AluI concentrations. Carry out the procedure up to Step 25 and check the efficiency of DNA fragmentation. DNA in the size range of 100–3,000 bp indicates good fragmentation (Fig. 4a)
	Insufficient or overdigestion of RNA. If RNA fragments are mostly >1,000 nt and contain obvious ribosomal RNA peaks in the Bioanalyzer results, they are not fragmented very well; if RNA fragments are mostly between 20 and 100 nt, they are overdigested	The amount of RNase I used to fragment RNA was not optimal in Steps 20 and 21	Try a different fold dilution of RNase I in 1× PBS; the recommended RNase I fold dilution is from 100-fold to 10,000-fold, depending on the digestion results. Carry out the procedure up to Step 25 and check the efficiency of RNA fragmentation. A good example of RNA fragmentation is shown in Fig. 4b
104	Too many PCR cycles are required to generate enough material for sequencing (>14 cycles) or the number of cycles varies too much between technical replicates	Too little material after Step 103, which could be caused by starting with an insufficient number of cells or by low efficiency of DNA and linker proximity ligation. On the basis of our experience, the quality of annealed linker is critical to the efficiency of proximity ligation. If the single-stranded top linker is in excess after annealing, it will lead to lots of RNA ligated to the single-stranded top linker, thus preventing DNA from being ligated to the linker	Perform the procedure from the start with more cells Redo Steps 34–46 and the Step 46 Checkpoint. If the amount of excess top linker is comparable to the amount of annealed ds linker, lower the amount of top linker or increase the amount the bottom linker and perform Steps 34–46 again to ensure complete annealing of the linker by checking that there is no excess top linker left unannealed. A little excess bottom strand will not affect the DNA and linker ligation step
105	Library fragment sizes are too small (<500 bp) when visualized on an agarose gel	DNA and linker proximity ligation efficiency is too low	Perform Steps 34–46 again and optimize linker annealing as mentioned above
115	Criterion ii of this checkpoint was not reached	Insufficient DNA fragmentation	Test chromatin fragmentation efficiencies under different volume ratios (volume of SDS/weight of nuclei), SDS treatment time (Fig. 5b,c) or AluI concentrations (Step 25)

Timing

Step 1 (day 1), cell crosslinking: 2 h
Steps 2–19 (day 2), cell lysis and DNA fragmentation: 3 h hands-on time plus an overnight reaction
Steps 20–48 (day 3), RNA fragmentation, preparation of RNA and DNA ends for ligation, linker adenylation, and RNA−linker ligation: 4 h plus an overnight reaction
Steps 49–54 (day 4), proximity ligation: 45 min plus an overnight reaction
Steps 55–72 (day 5), crosslinking reversal and DNA/RNA extraction: 5 h
Steps 73–77 (day 6), removal of biotin from unligated linkers: 3 h

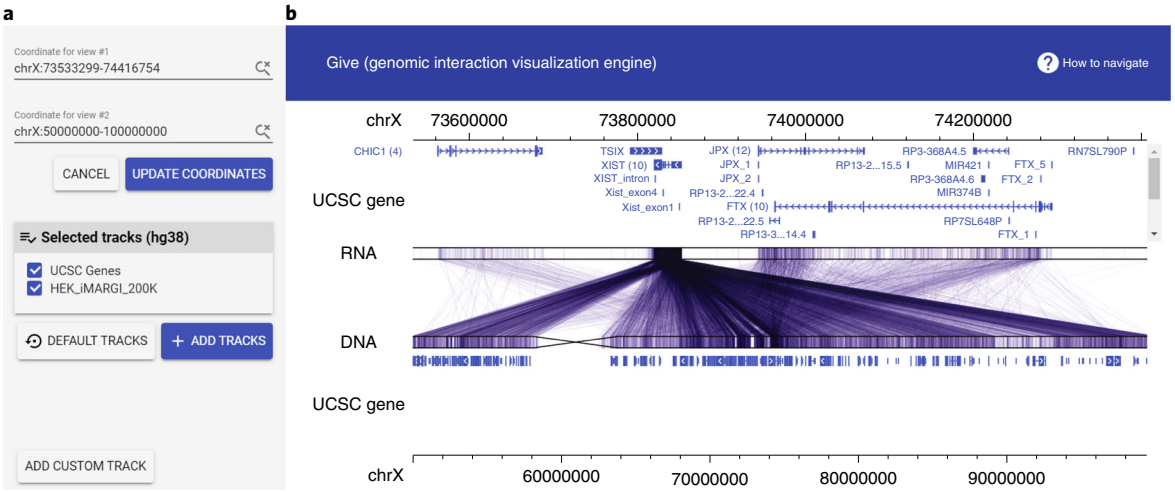

Fig. 6 | Visualization of iMARGI data in GIVE. a, User-defined genome coordinates and data tracks. **b**, A two-layer presentation of the genome. Each line corresponds to a mapped read pair, where the RNA end is mapped to the top layer (RNA lane) and the DNA end is mapped to the bottom layer (DNA lane). Users can adjust the coordinates of the two layers separately using the input box in **a**. The current setting shows human Chromosome X: 73,533,299–74,416754 in the top layer, covering the *XIST* gene and its genomic neighborhood, and Chromosome X: 50,000,000–100,000,000 in the bottom layer, covering 50 Mb of the X chromosome.

Steps 78–98 (day 7), biotin pull-down, reverse transcription, and ssDNA circularization: 5–6 h plus an overnight reaction
Steps 99–113 (day 8), Cut_oligo annealing, BamHI digestion, library amplification, size selection, and quality check: 6 h
Steps 114 and 115 (day 9 and beyond), sequencing: variable
Steps 116–123, data processing: 14 h
Steps 124–126, (optional) further analysis: 3 h
Steps 127–129, (optional) conversion of .pair files to other file types: 1 h
Box 1, checkpoint 1, for imaging nuclei integrity: 45 min
Box 2, checkpoint 2, for quantitative evaluation of nuclei integrity: 4 h

Anticipated results

The DNA sequences of a fully constructed iMARGI library are expected to be between 200 bp and 1,000 bp in length (Fig. 4d). The desired read depths depend on the research question. Owing to the diverse types of caRNAs and diverse modes of RNA–chromatin interactions, we are not able to recommend a uniform sequencing depth for all future studies. However, the *XIST*–X chromosome interaction and the interaction between the Malat1 lincRNA and the *NEAT1* gene locus could be reliably identified in HEK293T cells with ~30 million total read pairs[14]. In general, we recommend 300 million or more read pairs from paired-end sequencing. The first two bases of Read 2 are expected to be enriched with the CT dinucleotide corresponding to the AluI restriction sites. Although the minimum requirement in QC requires 50% or more C and T at each of the first two positions (see Checkpoint 7 at Step 115), high-quality libraries can reach 90% or more C and T at their respective positions. Although the minimum QC requirement at the mapping step is that 50% of the read pairs must have at least one end that is uniquely mapped and at least 50% of the uniquely mapped non-duplicate read pairs must be retained after the filtering steps (see Checkpoint 11 at Step 123), a high-quality library can reach 80% in both metrics.

Although RNA–chromatin interactions vary by cell type and cellular environment, the following observations are expected from a successful experiment: the number of uniquely mapped non-redundant iMARGI read pairs decreases as the distance between the two mapped ends increases, in an approximately linear manner on a log–log scale (see Fig. 2a in ref. [15]). Approximately 40% of the read pairs come from nascent transcripts and their nearby genomic sequences, as determined by the distance between the two mapped ends of a read pair. Prominent interactions between *XIST* and the X chromosome should be observed from differentiated female cells (see Fig. 3c of ref. [14]). The genome-wide distribution of small nuclear RNAs is expected to be non-uniform

(see Fig. 2 of ref. [18]), and the genomic regions enriched with small nuclear RNAs are expected to correlate with the A compartment, the nuclear compartment that is more active in transcription (see Fig. 5 of ref. [18]).

Reporting Summary

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

Data availability

An iMARGI dataset has been deposited into the NCBI Sequence Read Archive (SRA) under accession no. SRR8206679.

Code availability

The iMARGI-Docker software and its documentation are available at: https://sysbio.ucsd.edu/imargi_pipeline. The software is completely open source, under the BSD 2 license. The source code is available at https://github.com/Zhong-Lab-UCSD/iMARGI-Docker. The pre-built Docker image can be pulled from the Docker Hub. The version used in this paper is v.1.1.1.

References

1. Nguyen, T. C., Zaleta-Rivera, K., Huang, X., Dai, X. & Zhong, S. RNA, Action through Interactions. *Trends Genet.* **34**, 867–882 (2018).
2. Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131–137 (1996).
3. Yang, F. et al. The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biol.* **16**, 52 (2015).
4. Graf, M. et al. Telomere length determines TERRA and R-loop regulation through the cell cycle. *Cell* **170**, 72–85.e14 (2017).
5. Yu, R., Wang, X. & Moazed, D. Epigenetic inheritance mediated by coupling of RNAi and histone H3K9 methylation. *Nature* **558**, 615–619 (2018).
6. Watanabe, T. et al. Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus. *Science* **332**, 848–852 (2011).
7. Miao, Y. et al. Enhancer-associated long non-coding RNA LEENE regulates endothelial nitric oxide synthase and endothelial function. *Nat. Commun.* **9**, 292 (2018).
8. Place, R. F., Li, L. C., Pookot, D., Noonan, E. J. & Dahiya, R. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc. Natl Acad. Sci. USA* **105**, 1608–1613 (2008).
9. Morris, K. V., Chan, S. W., Jacobsen, S. E. & Looney, D. J. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* **305**, 1289–1292 (2004).
10. Rinn, J. L. et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
11. Simon, M. D. et al. The genomic binding sites of a noncoding RNA. *Proc. Natl Acad. Sci. USA* **108**, 20497–20502 (2011).
12. Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* **44**, 667–678 (2011).
13. Engreitz, J. M. et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
14. Sridhar, B. et al. Systematic mapping of RNA-chromatin interactions in vivo. *Curr. Biol.* **27**, 602–609 (2017).
15. Yan, Z. et al. Genome-wide co-localization of RNA-DNA interactions and fusion RNA pairs. *Proc. Natl Acad. Sci. USA* **116**, 3328–3337 (2019).
16. Bell, J. C. et al. Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts. *eLife* **7**, e27024 (2018).
17. Li, X. et al. GRID-seq reveals the global RNA–chromatin interactome. *Nat. Biotechnol.* **35**, 940–950 (2017).
18. Chen, W. et al. RNAs as proximity-labeling media for identifying nuclear speckle positions relative to the genome. *iScience* **4**, 204–215 (2018).
19. Yin, Y. et al. U1 snRNP regulates chromatin retention of noncoding RNAs. Preprint at bioRxiv, https://www.biorxiv.org/content/10.1101/310433v1 (2018).
20. Sun, G. et al. A bias-reducing strategy in profiling small RNAs using Solexa. *RNA* **17**, 2256–2262 (2011).
21. Jayaprakash, A. D., Jabado, O., Brown, B. D. & Sachidanandam, R. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res.* **39**, e141 (2011).
22. Sorefan, K. et al. Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence* **3**, 4 (2012).
23. Li, W., Freudenberg, J. & Miramontes, P. Diminishing return for increased Mappability with longer sequencing reads: implications of the k-mer distributions in the human genome. *BMC Bioinforma.* **15**, 2 (2014).

24. Zhao, J. et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell* **40**, 939–953 (2010).
25. Hadjiolov, A. A., Tencheva, Z. S. & Bojadjieva-Mikhailova, A. G. Isolation and some characteristics of cell nuclei from brain cortex of adult cat. *J. Cell Biol.* **26**, 383–393 (1965).
26. Suzuki, K., Bose, P., Leong-Quong, R. Y., Fujita, D. J. & Riabowol, K. REAP: a two minute cell fractionation method. *BMC Res. Notes* **3**, 294 (2010).
27. Nabbi, A. & Riabowol, K. Isolation of nuclei. *Cold Spring Harb. Protoc.* **2015**, 731–734 (2015).
28. Andrews, S. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).
29. Dekker, J. et al. The 4D nucleome project. *Nature* **549**, 219–226 (2017).
30. Quinlan, A. R. & Hall, I. M. J. B. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
31. Abdennur, N. & Mirny, L. Cooler: scalable storage for Hi-C data and other genomically-labeled arrays. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btz540 (2019).
32. Kerpedjiev, P. et al. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* **19**, 125 (2018).
33. Cao, X., Yan, Z., Wu, Q., Zheng, A. & Zhong, S. GIVE: portable genome browsers for personal websites. *Genome Biol.* **19**, 92 (2018).

Acknowledgements

We thank X. Wen for help with graphics; 4DN DCIC (http://dcic.4dnucleome.org) for discussions on the data processing pipeline; A. Zheng, A. Kaul, K. Faizi, N. Moshiri, R. Zhang, J. Ma, J. Chen, X. Huang and Z. Zhang for testing the iMARGI-Docker software; and K. Sriram and A. Chen for proofreading. This work was funded by DP1HD087990 (to S.Z.), NIH 4D Nucleome U01CA200147 (to S.C. and S.Z.), and NIH R00HL122368 and R01HL145170 (to Z.C.).

Author contributions

W.W. and S.Z. designed the research; W.W., T.C.N. and S.Z. developed the experimental method and protocol; Z.Y. and S.Z. developed the computational method and data analysis tools; W.W., Z.Y. and S.Z. wrote the manuscript; Z.C. and S.C. revised the manuscript.

Competing interests

S.Z. is a cofounder of Genemo Inc.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41596-019-0229-4.

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

Correspondence and requests for materials should be addressed to S.Z.

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

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

Received: 22 August 2018; Accepted: 23 July 2019;
Published online: 16 October 2019

Related links
Key references using this protocol
Sridhar, B. et al. *Curr. Biol.* **27**, 602–609 (2017) https://doi.org/10.1016/j.cub.2017.01.011
Yan, Z. et al. *Proc. Natl Acad. Sci. USA* **116**, 3328–3337 (2019) https://doi.org/10.1073/pnas.1819788116
Chen, W. et al. *iScience* **4**, 204–215 (2018) https://doi.org/10.1016/j.isci.2018.06.005

nature research

Corresponding author(s): Sheng Zhong

Last updated by author(s): 04/23/2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☒	☐ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	N/A
Data analysis	To minimize variations in data processing, we developed a complete data processing pipeline with full documentation (https://sysbio.ucsd.edu/imargi_pipeline) (Fig. 2). This pipeline, called iMARGI-Docker helps to improve analysis reproducibility by standardizing data processing steps. iMARGI-Docker can be executed on all mainstream operating systems without system-specific configuration, thus minimizing computer technology requirement of the data analysist. iMARGI-Docker source code is hosted at GitHub (https://github.com/Zhong-Lab-UCSD/iMARGI-Docker).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The example iMARGI sequencing data are deposited in NCBI Sequence Read Archive (SRA) and the accession number is SRR8206679.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Not applicable.
Data exclusions	Not applicable.
Replication	Not applicable.
Randomization	Not applicable.
Blinding	Not applicable.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Human embryonic kidney 293T (HEK293T) cells are purchased from ATCC, ATCC CRL-1573.
Authentication	The cells were authenticated by ATCC.
Mycoplasma contamination	HEK293 cells were purchased from ATCC, which have been thoroughly tested and authenticated to ensure their identity and that they are free from mycoplasma contamination (https://www.atcc.org/CellAuthenticationMatters.aspx). In addition, we carried out HiC-based DNA sequencing on this cell line. DNA sequencing also ruled out mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	None.