We report a molecular assay, Methyl-HiC, that can simultaneously capture the chromosome conformation and DNA methylation in a cell. Methyl-HiC reveals coordinated DNA methylation status between distal genomic segments that are in spatial proximity in the nucleus, and delineates heterogeneity of both the chromatin architecture and DNA methylene in a mixed population. It enables simultaneous characterization of cell-type-specific chromatin organization and epigenome in complex tissues.

DNA methylation is dynamically regulated and exhibits characteristic patterns in different cell types. DNA methylation can be profiled by whole-genome bisulfite sequencing (WGBS) at base resolution, which shows that the methylation status of adjacent CpGs is often correlated. Since DNA is spatially organized into three-dimensional structures, distal genomic regions may be brought into proximity through chromatin folding. Thus, it is possible that spatially proximal DNA sequences may also exhibit coordinated DNA methylation. However, conventional WGBS has limited power to detect coordinated DNA methylation across large genomic distances due to short DNA fragments profiled in such assays.

The chromosome conformation capture technologies capture spatial proximity through restriction digestion and in situ ligation of proximal DNA segments. In principle, methylation status of cytosines on the ligated DNA is preserved in chromosome conformation capture experiments and can be detected through bisulfite sequencing approaches. We therefore developed Methyl-HiC, a method that combines in situ Hi-C and WGBS to simultaneously profile chromatin conformation and DNA methylation. It begins with a standard in situ Hi-C procedure, followed by an additional step of bisulfite conversion before library construction and paired-end sequencing (Supplementary Fig. 1a). The methylation status of cytosines and the pair-wise contact frequencies are then determined using custom software (Supplementary Fig. 1b–e) (see Methods for description).

We applied Methyl-HiC to mouse embryonic stem cells (mESCs) (Supplementary Fig. 1f) and observed a contact matrix highly similar to that of in situ Hi-C (Fig. 1a), with indistinguishable distribution of contact probabilities as a function of genomic distances (Fig. 1b). Similar sets of chromatin loops were detected from both assays (Fig. 1a, Supplementary Fig. 2a,b and Supplementary Table 1). Comparable topologically associating domains (TADs) were also identified from the two datasets (Fig. 1c and Supplementary Fig. 2c,d). Comparing Methyl-HiC results with WGBS data from the same cell type, Methyl-HiC data captured over 16 million CpGs in the mouse genome and the methylation levels of these CpGs were highly consistent with that from WGBS (Supplementary Fig. 2e and Fig. 1d). Notably, compared with WGBS, Methyl-HiC profiled about 20% fewer CpGs overall but relatively more CpGs in open chromatin regions such as promoters and enhancers, likely due to reduced access to heterochromatin by the restriction enzyme used in the procedure (Supplementary Fig. 2e,f). These results, taken together, demonstrate that Methyl-HiC can simultaneously and accurately profile both the general features of chromosomal architecture and >80% of the methylome in a biological sample.

Methyl-HiC allows us to test the hypothesis that cytosines linearly separated but positioned proximally in space may also have coordinated methylation status. We analyzed the CpG methylation on Methyl-HiC reads separated in chromatin loop anchors (Supplementary Fig. 3a–c). Indeed, the methylation status of CpGs in Methyl-HiC read pairs mapping to separated loop anchors showed a notable correlation (Fig. 1e and Supplementary Fig. 3d). Further, such correlation was not restricted to chromatin loops mediated by CTCF, as loop anchors not occupied by CTCF also showed substantial correlation of DNA methylation (Supplementary Fig. 3e). We further tested whether chromatin organizational features influence such coordinated methylation. Indeed, Methyl-HiC read pairs within the same TADs exhibited slightly higher DNA methylation concordance than those from two different TADs (Supplementary Fig. 3f). Although read pairs in either compartment A or B showed similar DNA methylation concordance (Supplementary Fig. 3g), CpGs from loop anchors showed higher methylation concordance in compartment A than those in compartment B (Fig. 1f). We also observed a weaker but obvious concordance for inter-chromosomal interactions (Supplementary Fig. 3h). Finally, we observed that the degree of DNA methylation concordance varies depending on the underlying chromatin states (Supplementary Fig. 4).

To further demonstrate the utility of Methyl-HiC, we used it to analyze chromatin conformation and DNA methylation together in single cells. Cellular heterogeneity presents a significant challenge for chromatin architecture analysis in complex tissues. Single-cell Hi-C has previously been performed to characterize dynamic chromatin organization in early embryogenesis and during cell cycle, and to uncover the heterogeneity of chromatin architecture within a population of cells. Single-cell Methyl-HiC (scMethyl-HiC) can in principle resolve the heterogeneity of both DNA methylation and chromatin organization from the same cells. To achieve this, we adapted the above method by

---

1Ludwig Institute for Cancer Research, La Jolla, CA, USA. 2Division of Human Genetics, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH, USA. 3Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA. 4Cincinnati Cancer Center, University of Cincinnati, Cincinnati, OH, USA. 5Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA, USA. 6Broad Institute of MIT and Harvard, Cambridge, MA, USA. 7Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, USA. 8Department of Cellular and Molecular Medicine, Center for Epigenomics, Institute of Genomic Medicine, Moores Cancer Center, University of California, San Diego, School of Medicine, La Jolla, CA, USA. 9These authors contributed equally: Guoqiang Li, Yaping Liu. *e-mail: biren@ucsd.edu
Methyl-HiC protocol for single-cell analysis. Briefly, after proximity ligation we sorted individual nuclei into 96-well plates where bisulfite conversion was carried out with each nucleus. The resulting DNA was then amplified for sequencing. We generated scMethyl-HiC data for 103 mESCs cultured in serum + leukemia inhibitory factor (LIF) condition (serum mESCs) and 47 mESCs cultured in LIF with GSK3 and MEK inhibitors (2i) condition (2i mESCs) (Supplementary Table 2). The contact probabilities of the aggregate scMethyl-HiC were comparable to the bulk dataset (Fig. 2a and Supplementary Fig. 5a–c), and to previously published single-cell Hi-C data of mESCs15 (Fig. 2b). The scMethyl-HiC experiments also generated DNA methylomes with an average of 567,380 CpGs per nucleus, comparable to previous single-cell methylome datasets18,14 (Supplementary Table 2). The average methylation levels of CpGs in serum and 2i mESCs were 61.9% and 23.4%, while the methylation levels of CpHs were 1.3% and 0.9% (Fig. 2c and Supplementary Table 2), respectively, which are also consistent with previous observations14. As in bulk Methyl-HiC data, DNA methylation at the chromatin loop anchors was coordinated in individual cells (Fig. 2d). These results showed that Methyl-HiC can simultaneously profile DNA methylation and chromatin architecture in single cells.

To reveal the heterogeneity of the cell population using the above scMethyl-HiC data, we first clustered all of the cells based on the DNA methylation of each cell alone. Interestingly, in addition to a clear separation between mESCs grown in serum and 2i conditions, the mESCs grown in serum condition were further divided into two subpopulations, consistent with a previous report14 (Fig. 3a). We compared these subpopulations with DNA methylation profiles from multiple cell lineages and found that cluster 3 showed a potential embryonic limb development trend (Fig. 3b). The aggregate chromatin contact matrices in each cluster also showed clear cluster-specific chromosomal architecture (Fig. 3d–e). For example, Epha4 and genes within the HoxD locus15 were differentially methylated and exhibited differential compartments between cluster 2 and cluster 3 (Fig. 3f and Supplementary Fig. 6e,f). These results suggest that scMethyl-HiC can be used to resolve cell-type-specific chromosomal architecture in heterogeneous cell populations.

We propose that scMethyl-HiC may be used to study cell-type-specific epigenome and chromatin organization in complex tissues.
Fig. 3 scMethyl-HiC reveals heterogeneity of cultured mESCs. a, tSNE visualization of unsupervised clustering results according to DNA methylation from individual scMethyl-HiC dataset (n=150). Methylation level is calculated in non-overlapping 1-Mb bins. b, Unsupervised clustering of aggregated DNA methylene in subgroups of serum mESCs with tissue-specific methemolmes in mouse embryonic development. Pearson’s correlation matrices from different cell clusters. Similar numbers of cells are randomly selected to plot the matrixes (n=24, 26 and 24). The left map shows the bulk in situ Hi-C matrices from serum and 2i mESCs, respectively. Pearson’s correlation coefficient is calculated under 1-Mb resolution. Color ranges have been set to the same scale. d, GO biological process terms predicted by GREAT (genomic regions enrichment of annotations tool) for DMRs (n=99) switching from compartment B in cluster 2 to compartment A in cluster 3. e, GO biological process terms of genes (n=672) switching from compartment B in cluster 2 to compartment A in cluster 3. P values are from modified Fisher exact test for gene-enrichment analysis. f, Snapshot of HoxD cluster genes in differential compartments between clusters 2 and 3. Hi-C matrices are from bulk in situ Hi-C of the same cell line to show the chromosome organization in nearby regions. The two tracks below the Hi-C matrix are the eigenvector decomposition. The positive/negative values correspond to compartment A/B, respectively. DMRs between cluster 2 and cluster 3 are also plotted. GO, gene ontology; tSNE, t-distributed stochastic neighbor embedding.

While the current scMethyl-HiC protocol is still limited by data sparsity, this limitation may be overcome in the future by alternative DNA methylation detection strategies.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41592-019-0502-z.

Received: 14 November 2018; Accepted: 24 June 2019; Published online: 5 August 2019

References
1. Iurlaro, M., von Meyenn, F. & Reik, W. Curr. Opin. Genet. Dev. 43, 101–109 (2017).
2. Lyko, F. Nat. Rev. Genet. 19, 81–92 (2018).
3. Guo, J. U., Su, Y., Zhong, C., Ming, G.-I & Song, H. Cell 145, 423–434 (2011).
4. Luo, C. et al. Science 357, 600 (2017).
5. Lister, R. et al. Nature 462, 315 (2009).
6. Guo, S. et al. Nat. Genet. 49, 635 (2017).
7. Shoemaker, R., Deng, J., Wang, W. & Zhang, K. Genome Res. 20, 883–889 (2010).
8. Dixon, JesseK, Gorkin, DavidU. & Ren, B. Mol. Cell 62, 668–680 (2016).
9. Lieberman-Aiden, E. et al. Science 326, 289 (2009).
10. Dixon, J. R. et al. Nature 485, 376 (2012).
11. Rao, Suhals, P. et al. Cell 159, 1665–1680 (2014).
12. Schmitt, A. D., Hu, M. & Ren, B. Nat. Rev. Mol. Cell Biol. 17, 743 (2016).
13. Kubo, N. et al. Preprint at bioRxiv https://doi.org/10.1101/118737 (2017).
14. Schwartzman, O. & Tanay, A. Nat. Rev. Genet. 16, 716 (2015).
15. Nagano, T. et al. Nature 547, 61–67 (2017).
16. Flyamer, I. M. et al. Nature 544, 110 (2017).
17. Ramani, V. et al. Nat. Methods 14, 263–266 (2017).
18. Smallwood, S. A. et al. Nat. Methods 11, 817 (2014).
19. Rodríguez-Carballo, E. et al. Genes Dev. 31, 2264–2281 (2017).
20. Liu, Y. et al. Nat. Biotechnol. 37, 424–429 (2019).
21. Diedenhofen, B. & Musch, J. PLoS ONE 10, e0121945 (2015).
22. Yang, T. et al. Genome Res. 27, 1939–1949 (2017).

Acknowledgements
We thank S. Kuan and B. Li for processing of sequencing. The study was supported in part by the NIH 4D Nucleome program (grant no. U54DK107977) (to B.R.), grant no. 1U01HG007610-01 (to M.Y.) and a start-up grant from CCHMC (to Y.L.).

Author contributions
B.R., G.L. and Y.L. conceived the study and prepared the manuscript. G.L. performed the Methyl-HiC experiments. Y.L. wrote the bioinformatics pipeline and performed the analysis with the guidance of M.K.Y.L., G.L. and Y.Z. performed the in situ Hi-C and WGBS analyses. M.Y. and R.F. helped on single-cell Hi-C data analysis. N.K. provided the list of loops with CTCF. G.L., Y.L. and B.R. wrote the manuscript.

Competing interests
B.R. is a cofounder of Arima Genomics, Inc.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41592-019-0502-z.

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

Correspondence and requests for materials should be addressed to B.R.

Peer review information: Nicole Rusk was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
Methods

Cell culture. The F1 Mus massus castaneus/Sv129/Sv (embryonic stem) cell line (F123) was obtained from Whitehead Institute as described previously23. Serum mESCs were cultured with irradiated mouse embryonic fibroblasts (Gibco, A34180) in medium with 10% DMEM, 15% Knock-out Serum Replacement (Gibco, 10828-028), 1X penicillin/streptomycin, 1X non-essential amino acids (Gibco, 11140-050), 1X GlutaMax (Gibco, 35050), 0.4 mM β-mercaptoethanol and 1,000 U ml⁻¹ LIF (Millipore, ESG1107). 2i mESCs were adapted from serum mESCs by passing cells in MEF (mouse embryonic fibroblast) and serum-free 2i medium, which contained 50% Neurobasal (Gibco, 21103-049), 50% DMEM/F12 (Gibco, 11320-033), 0.5% N-2 Supplement (Gibco, 17502-048), 1% B27+RA (Gibco, 17504-044), 0.05% BSA (Gibco, 15260-037), 1X penicillin/streptomycin, 2 mM glutamine (Gibco, 25030-081), 150 μM monothioglycerol (Sigma, M6145), 1,000 U ml⁻¹ LIF (Millipore, ESG1107), 1 μM MEK inhibitor (Stemgent, 04-0006) and 3 μM GS3K inhibitor (Stemgent, 04-0004). Serum mESCs were collected and plated on 0.2% gelatin-coated feeder-free plates for 10 min before collecting to remove feeder cell contamination.

Methyl-Hic. In situ Hi-C was performed according to a published protocol11. Briefly, 2 million cells were cross-linked with 1% formaldehyde for 10 min at room temperature. Reaction was then quenched with 0.2 M glycine. Cell pellets were washed with cold PBS and lysed with lysis buffer to get nuclei pellets. Nuclei were permeabilized with 0.5% SDS. DNA was in situ digested with 100 units of DpnII (NEB, R0543) overnight. The ends of restriction fragments were filled with biotinylated nucleotides (Invitrogen, 92542016) and in situ ligated with T4 DNA ligase (NEB, M0204). After reversal of cross-links, ligated DNA was ethanol precipitated and resuspended to about 0.001 base pairs (bp) by sonication (Covaris). Sonicated products were pulled down with streptavidin beads. Library construction was then performed on beads. After adapter ligation, beads were suspended in 20 μl TE buffer and subjected to bisulfite conversion with EZ DNA Methylation-Gold Kit (Zymo, D5005). Unmethylated lambda DNA (Promega, D1501) was sonicated and ligated with the same adapter for methyl-Hic control. The library was then spiked in at 0.5% before bisulfite conversion. Purified bisulfite-converted DNA was amplified with HiFi Hotstart Uracil Ready Mix (KAPA, KK2802).

scMethyl-Hic. In situ Hi-C was performed the same as above for bulk Methyl-Hic up to the steps of proximal ligation. After ligation, nuclei pellets were centrifuged and washed with PBS. Pellets were suspended in PBS stained with 1:200 DRAQ7 (CST, 7406S). The nuclei were FACS sorted, 1 nucleus at a time, into a 96-well plate containing 9 μl PBS in each well. Bisulfite conversion was then performed in each well with EZ-96 DNA Methylation-Direct Kit according to the manufacturer’s manual (Zymo, D5020). Unmethylated lambda DNA (0.5%) was spiked in before bisulfite conversion. Following bisulfite conversion of single nuclei, random priming of bisulfite-converted DNA with high-concentration Klenow fragments (Enzymatics, P706L) incorporated an internal indexed P5 adapter to 5’ ends of synthesized fragments, which could be used for downstream multiplexing capability. Exonuclease I (NEB, M0293) and Alkaline Alkaline Phosphatase (NEB, M0371) treatments were then performed to digest unused random primer and inactivate dNTPs, followed by an SPRI bead-based purification step. P7 adapters were then ligated to the 3’ end of single-stranded products by Adaptaise module (Swift, 30096). Library amplification was then performed using indexed primers that incorporate dual indexing to enable 96-plex sequencing. Amplified libraries were pooled together, subjected to size selection and library quantification. Please refer to the Supplementary protocol for more details.

WGBS. Genomic DNA was first extracted from mESCs (Qiagen, 69504). Genomic DNA (1–1.5 μg) was fragmented by sonication (Covaris), end-repaired, dA-tailed and ligated to cytosine-methylated Illumina Truseq adapter. Ligation product was subjected to bisulfite conversion reaction according to the manufacturer’s instructions (Zymo, D5005). Unmethylated λDNA (0.5%) was spiked in before the conversion. Bisulfite-converted DNA was then PCR amplified and purified.

Sequencing of DNA libraries. The quantification of sequencing libraries was determined by quantitative PCR and TapeStation DNA analyzer (Agilent Technologies). Pooling of multiplexed sequencing samples, clustering and sequencing were carried out as recommended by the manufacturer on an Illumina HiSeq 2500 or HiSeq 4000. For bisulfite-converted libraries, at least 50% of balanced libraries or Phix were multiplexed to overcome the imbalance of GC ratio. All libraries were sequenced in paired-end mode.

Analysis of WGBS data. Raw reads were first trimmed as paired-end reads using Trimomatic with default parameters to remove the adapters and low-quality reads. Trimmmed reads were aligned to mm9 using Bismark (v1.2.5). PCR duplications were removed with Picard (http://broadinstitute.github.io/picard/). CpG methylation levels were calculated by Bis-SPN in bissnp_easy_usage.pl with default parameters.

Analysis of in situ Hi-C data. Each of the end raw reads was mapped separately to the mm9 reference genome using BWA-mem. Filtered reads were then paired with in situ Hi-C script. The reads that mapped to same fragment were further removed. Contact matrices were generated at different resolutions using Juicer pipeline with KR (Knight and Ruiz) normalization and visualized using Juicebox9. Loops were then called by HiCCUPS10.

Methyl-Hic read mapping by Bhmem. Raw reads were first trimmed as paired-end reads using Trimomatic with default parameters to remove the adapters and low-quality reads. The reference genome was in silico converted to make C/T replacement (all Cs were converted to Ts) and G/A reference (all Gs were converted to As). Paired-end reads were mapped in single-end mode to each converted reference by BWA-MEM. Only uniquely mapped and mapping quality-passed reads on both ends were joined. Joint read pairs with more than 20 kb insertion size were considered as long-range interactions and subjected to the following interaction analysis as G/C methylation levels were calculated by Bis-SPN11. Only bases with quality score more than 5 were included in the downstream methylation analysis. More details were implemented in Bhmem.java.

Correlation analysis of methylation in Methyl-Hic datasets. The reads with large fraction of incompletely converted cytosines were filtered out as previously described5. Only read pairs were considered as those between the ends were kept for the analysis. For the regions of interests, such as hiCCUPS loops anchor regions, Pearson correlation coefficient (PCC) was calculated directly from the methylation level of each read of the read pair, not by the average methylation level at each end of anchor regions. Only read pairs spanning at least 20 kb genomic distance and extended ENM were considered for the analysis. As control, all of the read pairs within the same genomic region no matter whether they had long-range interaction or not were randomly shuffled 100 times and then subjected to PCC calculation. Fisher’s (1925) z, implemented in ‘cocor.indep.groups’ in R package ‘cocor’, was used to assess the significance between the observed PCC from interacted read pairs and the PCC of random read pairs. The scripts for this analysis are provided in MethylCorAcrossHiCCupsJava.

scMethyl-Hic analysis. Reads were mapped by the same Bhmem pipeline with additional parameters to adapt the single-cell protocol (G was converted to A in read 1 and C was converted to T in read 2; read 1 and read 2 mapped to different allele to simulate the motif Hi-C Hic). The library was then spiked in at 0.5% before bisulfite conversion. Purified bisulfite-converted DNA was amplified with HiFi Hotstart Uracil Ready Mix (KAPA, KK2802).

Data availability

Figures show merged data from all replicates. The bulk Methyl-Hic, single-cell Methyl-Hic and WGBS datasets generated in this study have been deposited in the Gene Expression Omnibus (GEO) under the accession number GSE19171. Previously published data used for this study are listed in the Supplementary Notes.

Code availability

All of the source code for Bhmem software is publicly available at Bitbucket via https://bitbucket.org/dnaase/bisulfitehics/src/master/

References

23. Gribnau, J., Hochdelinger, K., Hata, K., Li, E. & Jaenisch, R. Genes Dev. 17, 1909–1920 (2003).

24. Durand, N. C. et al. Cell Syst. 3, 99–101 (2016).

25. Liu, Y., Siegmund, K. D., Laird, P. W. & Berman, B. P. Genome Biol. 13, R61 (2012).
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.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
- An indication of 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 statistics 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
- Clearly defined error bars
- State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection
No software was used for data collection.

Data analysis
Published software for data analysis used throughout the study includes BWA-MEM (bwa-0.7.12), FastQC v0.11.4, picard-tools-1.118, BisSNP-0.82.jar, bismark_v0.14.5, bedtools2-2.24.0, samtools-1.2, Trimmomatic-0.33, bedGraphToBigWig, juicer_tools.1.7.6, juicebox_tools.7.5, sambamba_v0.6.3, bamtools-2.4.0, and SNPsplit_v0.3.0. All the source code for Bhmem software is publicly available at Bitbucket via https://bitbucket.org/dnaase/bisulfitehic/src/master/.

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 upon request. 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 Methyl-HiC, single cell Methyl-HiC, and WGBS data sets generated in this study have been deposited in the Gene Expression Omnibus (GEO) under the accession number GSE119171. Previously published data used for this study are listed in Supplementary Table 4.

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

| Sample size | Sample sizes were chosen to provide sufficient confidence to validate methodological conclusions of the applicability of single cell Methyl-HiC. |
| Data exclusions | In the single cell Methyl-HiC datasets, after removing low quality reads (not unique mapped, PCR duplicate), cells with less than 250,000 reads and 10,000 contacts were removed for the further analysis. The rational behind this was that the limited reads number from individual cells cannot represent the chromatin and DNA methylation status very well. The exclusion criteria were not pre-established. |
| Replication | Bulk Methyl-HiC were performed with three replicates. We merged all the replications together for the further analysis. The reproducibility analysis for single cell Methyl-HiC is performed in Supplementary Fig. 5d. All attempts at replication were successful. |
| Randomization | Not relevant to this study as we didn’t include experimental groups of samples/organism/participants. |
| Blinding | Not relevant to this study as we didn’t include group allocation. |

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a Involved in the study
☒ Unique biological materials
☒ Antibodies
☒ Eukaryotic cell lines
☒ Palaeontology
☒ Animals and other organisms
☒ Human research participants

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) The F1 Mus musculus castaneus × S129/SvJae mouse ESC line (F123) was a gift from the Whitehead Institute. Mouse embryonic fibroblasts were purchased from Gibco (Gibco, A34180).

Authentication Mouse embryonic fibroblasts were commercially available and authenticated by the manufacturer. F123 cell line was not authenticated to the best of our knowledge.

Mycoplasma contamination Cell line was not tested for mycoplasma contamination in this project.
Commonly misidentified lines
(See ICLAC register)
No commonly misidentified cell lines were used.

Flow Cytometry

Plots

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation
For single cell Methyl-HiC, In situ Hi-C was performed on F123 mESCs as same as above bulk Methyl-HiC till proximal ligation. After ligation, nuclei pellets were centrifuged and washed with PBS. Pellets were suspended in PBS stained with 1:200 DRAQ7 (CST, 74065).

Instrument
Sony, SH800

Software
Sony, SH800 software

Cell population abundance
The percentage of nucleus collected was around 40%, which was determined by DRAQ7 staining.

Gating strategy
Positive and negative boundaries were determined by DRAQ7 staining, which specified the nucleus in interphase.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.