

Methyl-Micro-C: simultaneous characterization of chromatin accessibility, interactions, and DNA methylation

Leonardo Gonzalez-Smith, Claire Stevens, Huan Cao, Zexun Wu, Suhn K. Rhie 6*

Department of Cancer Biology and the Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, United States

*To whom correspondence should be addressed. Email: rhie@usc.edu

Abstract

Epigenomes, characterized by patterns of different signatures such as chromatin accessibility, chromatin interactions, and DNA methylation, vary across cell types and play a pivotal role in regulating gene expression. By mapping these signatures, the underlying mechanisms of development and diseases can be uncovered. However, many canonical epigenetic methods focus on mapping only one signature. Simultaneous measurement of epigenetic signatures from the same cell or tissue provides significant benefits for research, especially when resources are limited, and precise analysis is essential. Here, we report a technique called Methyl-Micro-C (MMC), which simultaneously profiles chromatin accessibility, chromatin interactions, and DNA methylation in the same sample. MMC enhances the resolution of chromatin interactions and the coverage of CpGs by combining MNase-mediated fragmentation with enzymatic conversion. This technique allows for the profiling of three-dimensional epigenomes, capturing consistent chromatin accessibility, chromatin interactions, and DNA methylation signals in an efficient manner. It is also relatively straightforward, allowing researchers to implement and apply it easily.

Introduction

Three-dimensional (3D) epigenomes vary across cell types, and their changes are linked to a range of diseases [1-3]. To profile epigenomes, whole genome bisulfite sequencing (WGBS) is used to characterize the methylation state of CpG regions throughout the entire genome, including regulatory elements that control gene expression [4, 5]. Recently, enzymatic methyl sequencing (EM-seq) has been developed [6]. This method provides improved coverage and sensitivity compared to WGBS by using enzymes instead of bisulfite treatment that can damage DNA. Besides DNA methylation, chromatin accessibility is used to determine the activities of genomic regions. Functional assays coupled with next-generation sequencing, such as DNase-seq, ATAC-seq, and MNase-seq, determine which regulatory elements are accessible or inaccessible [7]. To control transcription, physical interactions in 3D space occur between genomic regions that are far apart on the linear genome. Chromatin conformation capture (3C) and its derivative methods such as Hi-C utilize restriction enzymes followed by proximity ligation to identify chromatin interactions. Recently, Micro-C has been developed to further improve Hi-C by employing uniform MNase-mediated digestion. This advancement allows for higher-resolution of contact maps and loop calling as well as co-profiling chromatin accessibility [8-10].

Together, DNA methylation, chromatin accessibility, and chromatin interaction maps offer a comprehensive understanding of the 3D epigenome. However, profiling each of these features individually requires significant effort and often entails prohibitively expensive deep sequencing for accu-

rate comparative analyses. Simultaneous measurement of epigenetic signatures from the same cell or tissue can greatly benefit research, especially when resources are limited, and precise analysis is crucial. To address these challenges, here we developed Methyl-Micro-C (MMC).

Materials and methods

Cell culture

22Rv1 (CRL-2505) cells were obtained from ATCC (https://www.atcc.org) and were cultured according to the suggested protocols using RPMI (Corning, 10-040-CV) supplemented with 10% FBS plus 1% penicillin and 1% streptomycin. RC-77T/E cells were received from Johng Sik Rhim and were cultured using Keratinocyte SFM medium plus supplement reagents (Thermo Fisher, 17 005 042) as previously reported [11].

Methyl-Micro-C (MMC)

Cells were harvested, aliquoted into 1 million cell pellets, and snap frozen. Pellets were fixed by cross-linking, and lysed to isolate intact nuclei. Nuclei were treated with MNase (NEB, M0247S), and 1000 ng of digested chromatin underwent proximity ligation with a biotinylated bridge sequence. Two hundred nanograms of ligated DNA products underwent end preparation (NEB, E7546S), and methylated adapters (NEB, E7165AAVIAL) were added. After enriching ligated products, oxidation of methylated and hydroxymethylated cytosines as well as deamination of unprotected

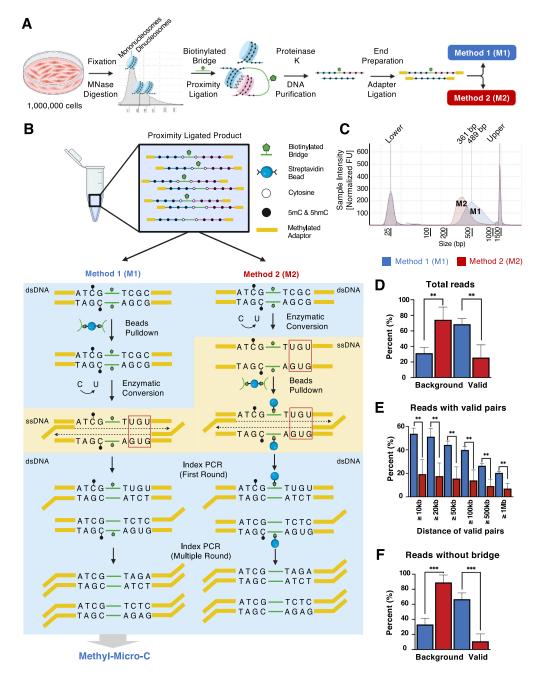


Figure 1. Methyl-Micro-C (MMC) workflow and optimization. (**A**) MMC starts with the step of harvesting cells and digesting nuclei with MNase to achieve an optimal 80:20 mono- to dinucleosomal ratio. Next step is a proximity ligation with a biotinylated bridge sequence. Ligated products are then purified and pulled down with streptavidin beads. Then, end preparation and methylated adapter ligation are performed. To optimize MMC, we tested Method 1 and Method 2 using 22Rv1 and RC-77T/E prostate cancer cells. (**B**) Method 1 and Method 2 use different orders of the following steps: streptavidin beads pulldown and elution to enrich interacting regions, oxidation of all 5mCs and 5hmCs, and deamination of unprotected cytosines to convert them to uracils. Next, Method 1 performs a dsDNA pulldown and elution off of streptavidin beads prior to enzymatic conversion, while Method 2 performs enzymatic conversion prior to a ssDNA pulldown, with index PCR being performed on the streptavidin beads. (**C**) The size of libraries generated from Method 1 (blue) and Method 2 (red) was measured. (**D**) Comparison of background (non-duplicate pairs <1 kb apart) generated from Method 1 and Method 2. (**E**) Chromatin interaction distance of valid pairs generated from Method 1 and Method 2. (**F**) Comparison of background and valid pairs between Method 1 and Method 2 using pairs that did not contain the bridge sequences. Method 1, which outperforms Method 2, is named MMC and deep-sequenced for further downstream analysis. Data with the same number of reads across replicates were used, and statistical significance was calculated using a t-test (*, *P-value* < .05; **, *P-value* < .01; ***, *P-value* < .001).

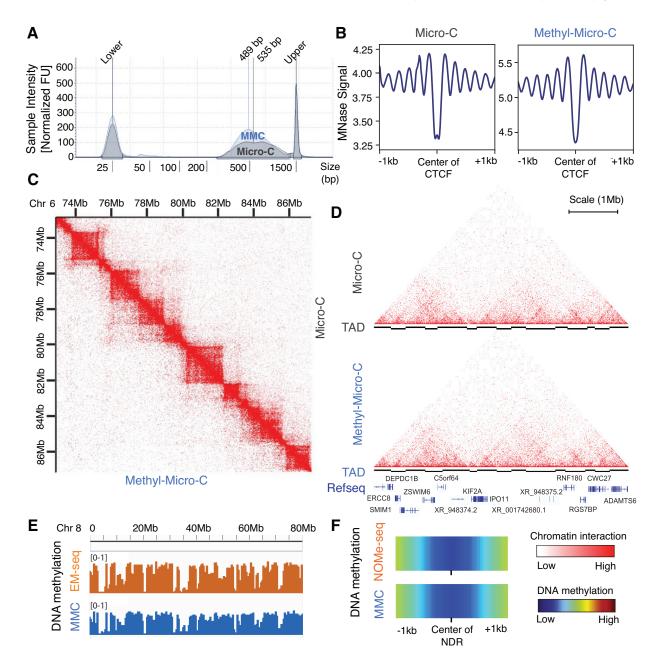


Figure 2. MMC yields similar results to Micro-C and whole genome methylation sequencing. (A) The size of libraries generated from Micro-C (gray) and MMC (blue) using 22Rv1 cells were measured. (B) Analysis of mapped reads centered around 22Rv1 CTCF sites was shown for Micro-C (left) and MMC (right) data. (C) Chromatin interaction heatmaps at 25 kb resolution are shown to compare Micro-C (top right) and MMC (bottom left). (D) Similar chromatin interactions and topologically associated domains (TADs) were detected between Micro-C (top) and MMC (bottom). (E) Comparison of DNA methylation levels on the p arm of chromosome 8 between EM-seq (top) and MMC (bottom) data generated from 22Rv1 cells. (F) DNA methylation levels around nucleosome depleted regions (NDRs) were measured using NOMe-seq (top) and MMC (bottom) data generated from 22Rv1 cells.

unmethylated cytosines was performed, using EM-seq conversion module (NEB, E7125S). MMC libraries were generated by performing index PCR and sequenced using Illumina sequencing systems, following the guidelines of the manufacturer. A more detailed MMC protocol is available in Supplementary Methods.

Enzymatic methyl sequencing (EM-seq)

The Enzymatic methyl sequencing (EM-seq) library was generated using EM-seq conversion module (NEB, E7125S) as previously reported [6], and the library was sequenced using Illumina sequencing platforms.

Nucleosome occupancy and methylome sequencing (NOMe-seq)

The NOMe-seq library was generated using M.CviPI methyltransferase (NEB, M0227B) to methylate GpCs as previously described [12]. Bisulfite treatment was used to convert unmethylated CpGs, and the library was generated using IDT xGen Methyl-seq kit (IDT, 10009824) and sequenced using Illumina sequencing platforms.

NOMe-HiC

22Rv1 cells were harvested and cross-linked with 1% formaldehyde for 10 minutes. The reaction was quenched with 0.2 M glycine, followed by ice-cold PBS (Phosphate-

buffered saline) washes. As previously described [13], 2 million cells were resuspended in nuclei isolation buffer and incubated on ice for 1 hour. Chromatin was then digested overnight with 100 U DpnII (NEB, R0543L) at 37°C. DpnII was inactivated, and the nuclei were cooled to room temperature for biotin fill-in. Proximal fragments were ligated in situ using T4 DNA ligase (NEB, M0202) and incubated with M.CviPI methyltransferase (NEB, M0227B). The reaction was stopped with PBS, and nuclei were reverse crosslinked overnight. DNA was then extracted, sonicated, and purified with AMPure beads (Beckman Coulter, A63882). Streptavidin beads (Invitrogen, 2877009) were used to pull down biotinylated DNA, and library preparation (NEB, E7645S with E7165AAVIAL) was performed on beads. Bisulfite conversion (Zymo Research, D5002) was performed overnight, and libraries were generated using a uracil-sensitive PCR mix

Data analysis

(Roche, 07959052001).

MMC, Methyl-HiC [14], and NOMe-HiC [13] data were mapped to the human genome (hg38) or mouse genome (mm10), depending on the cell line type, using bhmem [14]. Chromatin interaction data from MMC, Micro-C, Methyl-HiC, and NOMe-HiC data were processed using the 4DN pipeline [15] and HiC-Pro [16], as previously done [8]. Reads with bridge sequences from in-house MMC and Micro-C data were determined by searching for unique bridge sequences from the alignment file. Topologically associated domains (TADs) were identified using TopDom [17], and chromatin loops were identified using Mustache [18]. TADs and loops were compared among datasets, including those from the 22Rv1 Hi-C data previously generated in-house [2, 19]. HiCcompare [20] was used to compare chromatin interactions between datasets, and Juicebox [21] was used to visualize chromatin contact maps. To analyze DNA methylation, MMC, EM-seq, NOMe-seq, Methyl-HiC, and NOMe-HiC data were aligned to the human genome (hg38) or mouse genome (mm10), depending on the cell line type, using bwa-meth (https://github.com/brentp/bwa-meth) and quantified using BISCUIT [22]. Three replicates of MMC, Methyl-HiC, and NOMe-HiC data were subsampled to an equal number of reads and compared. Bis-tools [23] and IGV [24] tools were used to visualize DNA methylation signals. 22Rv1 CTCF ChIP-seq data, generated as part of the previous study [19], and CTCF motif data were used to identify robust and replicated CTCF binding sites. 22Rv1 nucleosome depleted region (NDR) data were obtained from previous studies [25, 26]. All datasets used for this study are listed in Supplementary Table S1.

Results and discussion

To simultaneously characterize chromatin activities and structures in the same sample, we leveraged Micro-C and EM-seq techniques and optimized the protocols to easily perform both using the same sample. We included eight steps, all of which can be completed relatively simply (Supplementary Fig. S1) compared to other techniques [13, 14, 27].

First, we harvested and cross-linked cells. Second, we used MNase to fragment nucleosomes genome-wide without sequence bias and lysed cells to isolate nuclei. Third, physically interacting regions were ligated with a biotinylated bridge se-

quence. Fourth, end repair and ligation of methylated adapters were performed. Fifth, ligated DNA was pulled down to enrich interacting regions. Sixth, 5mCs and 5hmCs were oxidized to 5gmCs using TET2 and T4-BGT. Seventh, unmethylated cytosines were deaminated using APOBEC, converting all unprotected cytosines to uracils. Lastly, index PCR, which converts all uracils to thymines, was used to generate a sequencing library.

To optimize the protocol, we tested two different workflows: Method 1 and Method 2. Both workflows included the same steps 1 through 4. However, subsequent steps were performed in a different order. To test the efficacy of each method, the same proximity ligated product from step 4 of the protocol was split, and both workflows were tested in parallel (Fig. 1A). In Method 1, we pulled down the ligated double-stranded DNA (dsDNA) with beads and performed enzymatic conversion (Fig. 1B). On the contrary, in Method 2, we enzymatically converted the ligated DNA, and the generated single-stranded DNA (ssDNA) was pulled down with beads. Libraries of both methods were subsequently made by performing index PCR.

To compare Method 1 and Method 2, we first examined the size of libraries and found that Method 1 retained a larger size (489 bp) than Method 2 (361 bp) (Fig. 1C). Next, we compared the percentage of background reads, defined as non-duplicate pairs <1 kb apart in the same chromosome, with the percentage of valid pairs, defined as proximity ligated non-duplicate pairs that were greater than or equal to 1 kb apart in the same chromosome as previously done [28]. We revealed that Method 1 had over 2.5-fold greater valid pairs than Method 2 (Fig. 1D). When we further analyzed the distance of valid pairs (i.e. \geq 10 kb to \geq 1 Mb), we found that Method 1 was able to capture more pairs that are farther apart than Method 2 (Fig. 1E) (t-test *P*-value < .01).

When we further examined whether the high background rate was related to the ligated DNA step, we found that, on average, Method 1 had a similar number of reads with bridge sequences (46.28%) compared to Method 2 (55.75%). When we compared the percentage of valid pairs from reads with bridge sequences and reads to those without bridge sequences for Method 1, the percentage of valid pairs did not change (68.60% versus 71.79%) (Supplementary Table S2). However, Method 2 did not display this same trend. When examining the reads that did not contain the bridge sequences, Method 2 displayed significantly higher background reads and lower valid pairs compared to Method 1 (Fig. 1F) (t-test P-value < .0005). This indicates that Method 1 enables increased enrichment of more valid pairs, while Method 2 had a higher proportion of its background reads nested within sequences that did not contain the bridge. Therefore, we named Method 1 as MMC (Supplementary Fig. S1 and Supplementary Methods) and further deep-sequenced the libraries for downstream

Next, we compared MMC with Micro-C. When we first examined the sizes of Micro-C and MMC libraries, we found a similar average library size around 500 bp (Fig. 2A). By visualizing the chromatin accessibility around CTCF binding sites using Micro-C and MMC data, we observed that MMC can detect similar nucleosome phasing to Micro-C (Fig. 2B). To further visualize chromatin interactions, side-by-side heatmaps were displayed, and we found similar chromatin interactions, indicating the ability of MMC to detect

similar chromatin interactions (Fig. 2C and Supplementary Fig. S2). Both methods also detected comparable TADs (Fig. 2D and Supplementary Fig. S3).

Besides mapping chromatin accessibility and interactions, MMC has the advantage of simultaneously profiling the DNA methylome. Comparing DNA methylation levels between EM-seq and MMC, we found that overall DNA methylation patterns were similar to each other (Fig. 2E and Supplementary Fig. S4). By examining DNA methylation levels near NDRs, we revealed that both NOMe-seq and MMC methods found unmethylation at NDRs, demonstrating that MMC can also profile DNA methylation of regulatory elements (Fig. 2F).

MMC implements MNase digestion instead of the restriction enzyme-mediated fragmentation, used in the other simultaneous profiling methods such as Methyl-HiC and NOMe-HiC (Supplementary Fig. S5). Therefore, MMC allows for increased resolution with less cellular input. It can also be optimized and repeated with the same cross-linked chromatin using various MNase concentrations to optimally capture mono-, di-, and tri-nucleosome fragments. Moreover, MMC is more time-efficient as the unmethylated cytosine conversion step is not dependent on an overnight incubation with bisulfite. In addition to its efficiency, MMC captured a higher percentage of valid pairs (70.09%) compared to Methyl-HiC (48.48%, t-test P-value < .05) and NOMe-HiC (34.44%,t-test P-value < .001) (one-way ANOVA P-value < .005) (Supplementary Fig. S6). When we further compared NOMe-HiC and MMC data generated in-house, MMC captured significantly more valid pairs (65.96%) than NOMe-HiC (35.72%) (t-test *P*-value < .05). Furthermore, MMC captured more CpG sites involved in chromatin loops compared to NOMe-HiC (t-test *P*-value < .05) (Supplementary Fig. S7).

In conclusion, MMC allows for the simultaneous profiling of chromatin accessibility, chromatin interactions, and DNA methylome from the same sample. This technique offers significant benefits, when comparing multi-omics data from the same sample. Moreover, MMC is particularly advantageous when samples are limited and precise analysis is required within a constrained budget, without the need for deep sequencing each data type individually.

Acknowledgements

We acknowledge the Center for Advanced Research Computing (CARC) at the University of Southern California for providing computing resources that have contributed to the research results reported in this study (https://carc.usc.edu). We thank Ethan Nelson-Moore, Shannon Schreiner, and the other Rhie lab members for useful discussions and insightful suggestions.

Author contributions: Leonardo Gonzalez-Smith (Investigation [equal], Methodology [equal], Visualization [equal], Writing—original draft [supporting], Writing—review & editing [supporting]), Claire Stevens (Data curation [lead], Formal analysis [equal], Investigation [supporting], Methodology [supporting], Validation [equal], Visualization [equal], Writing—original draft [supporting], Writing—review & editing [supporting]), Huan Cao (Investigation [equal], Methodology [equal], Validation [equal], Visualization [supporting], Writing—original draft [supporting], Writing—review & editing [supporting]), Zexun Wu (Data curation [supporting], Formal analysis [equal], Validation [equal], Visualization]

tion [equal], Writing—review & editing [supporting]), and Suhn K. Rhie (Conceptualization [lead], Formal analysis [supporting], Funding acquisition [lead], Investigation [supporting], Methodology [supporting], Supervision [lead], Writing—original draft [lead], Writing—review & editing [lead]).

Supplementary data

Supplementary data is available at NAR Genomics & Bioinformatics online.

Conflict of interest

None declared.

Funding

This work was supported in part by grants K01CA229995, R21CA260082, R21CA264637, R21HG011506, and P30CA014089 from the National Institutes of Health, W81XWH-21-1-0805 grant from the Department of Defense, and pilot grants from the USC Keck School of Medicine, the USC Center for Genetic Epidemiology, and the USC Norris Comprehensive Cancer Center.

Data availability

All of the next generation sequencing datasets generated in this study have been submitted to the NCBI GEO under accession number GSE276687. The in-house scripts of bioinformatics analysis are available at https://zenodo.org/records/15320273 and https://github.com/rhielab/MethylMicroC.

References

- 1. Rhie SK, Schreiner S, Witt H *et al.* Using 3D epigenomic maps of primary olfactory neuronal cells from living individuals to understand gene regulation. *Nat Methods* 2018;4:1046–55. https://doi.org/10.1126/sciadv.aav8550
- Rhie SK, Perez AA, Lay FD et al. A high-resolution 3D epigenomic map reveals insights into the creation of the prostate cancer transcriptome. Nat Commun 2019;10:4154. https://doi.org/10.1038/s41467-019-12079-8
- 3. Wang Y, Song F, Zhang B *et al.* The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol* 2018;19:151. https://doi.org/10.1186/s13059-018-1519-9
- Zhou W, Dinh HQ, Ramjan Z et al. DNA methylation loss in late-replicating domains is linked to mitotic cell division. Nat Genet 2018;50:591–602. https://doi.org/10.1038/s41588-018-0073-4
- Lister R, Pelizzola M, Dowen RH et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 2009;462:315–22. https://doi.org/10.1038/nature08514
- Vaisvila R, Ponnaluri VKC, Sun Z et al. Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. Genome Res 2021;31:1280–9. https://doi.org/10.1101/gr.266551.120
- Lee BH, Rhie SK. Molecular and computational approaches to map regulatory elements in 3D chromatin structure. *Epigenetics Chromatin* 2021;14:14.
- 8. Lee BH, Wu Z, Rhie SK. Characterizing chromatin interactions of regulatory elements and nucleosome positions, using Hi-C, Micro-C, and promoter capture micro-C. *Epigenetics Chromatin* 2022;15:41.

- 9. Hsieh T-HS, Cattoglio C, Slobodyanyuk E et al. Resolving the 3D landscape of transcription-linked mammalian chromatin folding. Mol Cell 2020;78:539-53. https://doi.org/10.1016/j.molcel.2020.03.002
- 10. Krietenstein N, Abraham S, Venev SV et al. Ultrastructural details of mammalian chromosome architecture. Mol Cell 2020;78:554-65. https://doi.org/10.1016/j.molcel.2020.03.003
- 11. Theodore S, Sharp S, Zhou J et al. Establishment and characterization of a pair of non-malignant and malignant tumor derived cell lines from an African American prostate cancer patient. Int J Oncol 2010;37:1477-82.
- 12. Rhie SK, Schreiner S, Farnham PJ. Defining regulatory elements in the human genome using nucleosome occupancy and methylome sequencing (NOMe-seq). Methods Mol Biol 2018;1766:209-29. https://doi.org/10.1007/978-1-4939-7768-0_12
- 13. Fu H, Zheng H, Chen X et al. NOMe-HiC: joint profiling of genetic variant, DNA methylation, chromatin accessibility, and 3D genome in the same DNA molecule. Genome Biol 2023;24:50. https://doi.org/10.1186/s13059-023-02889-x
- 14. Li G, Liu Y, Zhang Y et al. Joint profiling of DNA methylation and chromatin architecture in single cells. Nat Methods 2019;16:991-3. https://doi.org/10.1038/s41592-019-0502-z
- 15. Dekker J, Belmont AS, Guttman M et al. The 4D nucleome project. Nature 2017;549:219-26. https://doi.org/10.1038/nature23884
- 16. Servant N, Varoquaux N, Lajoie BR et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol 2015;16:259. https://doi.org/10.1186/s13059-015-0831-x
- 17. Shin H, Shi Y, Dai C et al. TopDom: an efficient and deterministic method for identifying topological domains in genomes. Nucleic Acids Res 2016;44:e70. https://doi.org/10.1093/nar/gkv1505
- 18. Roayaei Ardakany A, Gezer HT, Lonardi S et al. Mustache: multi-scale detection of chromatin loops from hi-C and Micro-C maps using scale-space representation. Genome Biol 2020;21:256. https://doi.org/10.1186/s13059-020-02167-0
- 19. Guo Y, Perez AA, Hazelett DJ et al. CRISPR-mediated deletion of prostate cancer risk-associated CTCF loop anchors identifies

- repressive chromatin loops. Genome Biol 2018;19:160. https://doi.org/10.1186/s13059-018-1531-0
- 20. Stansfield JC, Cresswell KG, Vladimirov VI et al. HiCcompare: an R-package for joint normalization and comparison of HI-C datasets. BMC Bioinformatics 2018;19:279. https://doi.org/10.1186/s12859-018-2288-x
- 21. Durand NC, Robinson JT, Shamim MS et al. Juicebox provides a visualization system for hi-C contact maps with unlimited zoom. Cell Syst 2016;3:99-101. https://doi.org/10.1016/j.cels.2015.07.012
- 22. Zhou W, Johnson BK, Morrison J et al. BISCUIT: an efficient, standards-compliant tool suite for simultaneous genetic and epigenetic inference in bulk and single-cell studies. Nucleic Acids Res 2024;52:e32. https://doi.org/10.1093/nar/gkae097
- 23. Lay FD, Liu Y, Kelly TK et al. The role of DNA methylation in directing the functional organization of the cancer epigenome. Genome Res 2015;25:467-77. https://doi.org/10.1101/gr.183368.114
- 24. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 2013;14:178-92. https://doi.org/10.1093/bib/bbs017
- 25. Chen Z, Wu D, Thomas-Ahner JM et al. Diverse AR-V7 cistromes in castration-resistant prostate cancer are governed by HoxB13. Proc Natl Acad Sci USA 2018;115:6810-5. https://doi.org/10.1073/pnas.1718811115
- 26. Tang F, Xu D, Wang S et al. Chromatin profiles classify castration-resistant prostate cancers suggesting therapeutic targets. Science 2022;376:eabe1505. https://doi.org/10.1126/science.abe1505
- 27. Lee D-S, Luo C, Zhou J et al. Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. Nat Methods 2019;16:999-1006. https://doi.org/10.1038/s41592-019-0547-z
- 28. Akgol Oksuz B, Yang L, Abraham S et al. Systematic evaluation of chromosome conformation capture assays. Nat Methods 2021;18:1046-55. https://doi.org/10.1038/s41592-021-01248-7