nature portfolio

https://doi.org/10.1038/s41580-022-00518-2

Supplementary information

Regulation, functions and transmission of bivalent chromatin during mammalian development

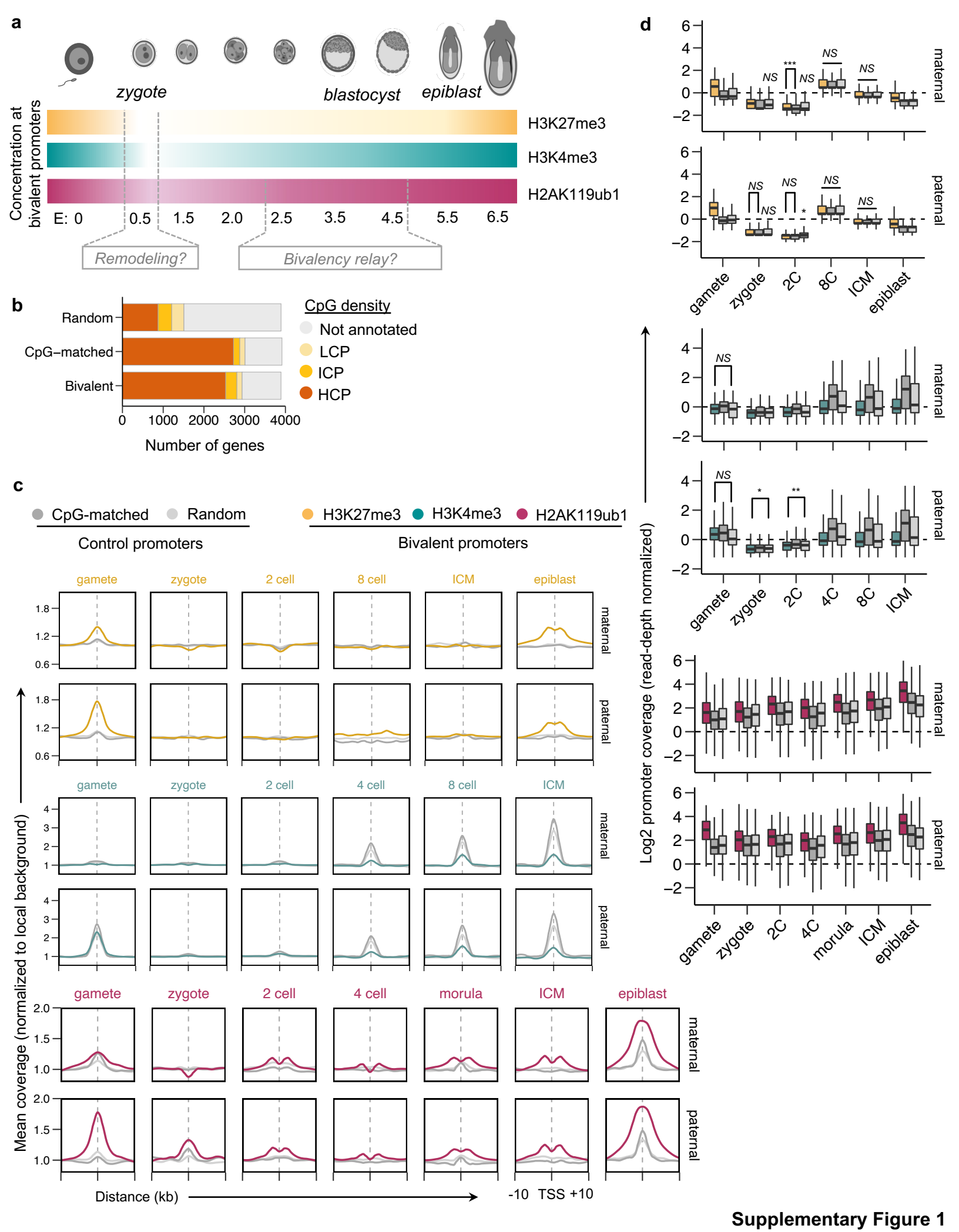
In the format provided by the authors and unedited

Supplementary Table 1. Genome-wide techniques for mapping histone modifications and chromatin-associated proteins.

Primary purpose	Technique	Description of method	Key refs
Direct detection of bivalency	Sequential ChIP, re-ChIP	- Stepwise ChIP of histone modifications to detect their overlap at the same chromatin fragments. - Caveats include large input requirement.	1,2
	reICeChIP-seq	- Combines spike-in nucleosome standards with cleavable anti-H3K4me3 fragment antigen-binding region and sequential IP. - Enables quantification of bivalency at single-nucleosome resolution.	3–5
	CIDOP-seq	Fuses histone-modification reading domains of different proteins to detect bivalency; used for H3K9me3 and H3K36me2/3 co-detection).	6
	cMAPs (IP and fluorescence)	- Fuses histone reading domains to detect bivalency with a fluorescent marker. - Can be used for IP-seq or imaging.	7
	Combinatorial detection	- Isolated nucleosomes are labeled with fluorescent adapters and captured on streptavidin-coated slides. - Single-molecule resolution; can be used for IP-seq or imaging.	8
	MulTI-Tag, Multi-CUT&Tag	Extends CUT&Tag (see below) to detect multiple targets with barcoded primary antibodies. Adapter-tagged secondary antibodies allow for distinguishing distinct IP targets.	9,10
Quantitative IP	ChIP-seq with exogenous spike-in	- Mixes material from different organisms before cells or chromatin IP or after DNA IP based on cell number rather than input amount. - Allows quantitative normalization of ChIP signal that accounts for global differences in enrichment levels. - Can be applied to ChIP-seq and other techniques presented in this table, although challenging with very low-cell-number techniques.	11–17
Proximity proteomics	ChromID	- Ligates multiple chromatin-reader protein domains to a biotin ligase, e.g. the CBX7 chromodomain and TAF3 bromodomain, to form a bivalency reader. - Profiles the chromatin protein interactome at targeted regions.	18

DNA replication	ChOR-seq	- Assesses the genomic location and quantifies levels of histone modifications or chromatin-associated proteins following DNA replication. - Employs pulse-chase labeling with 5-ethynyl-2'-deoxyuridine (EdU) to trace DNA replication.	19
	SCAR-seq	- Distinguishes chromatin occupancy between DNA strands following replication. - An EdU pulse labels nascent DNA, and biotin-captured (nascent) fragments are then separated into new vs parental strands.	20
Parental inheritance	Allele specific ChIP-seq	- Cells are collected from crosses between distinct lines, resulting in allele-specific SNPs that allow for distinguishing maternal from paternal strands. - Can be applied to other methods in this table.	21,22
Low input, in vivo	ULI NChIP-seq	Uses MNase to produce high-resolution maps of histone modifications from low inputs (as few as 1000 cells).	23
	STAR ChIP-seq*	Combines MNase digestion with streamlined DNA purification during library preparation to maximize recovery and minimize background.	21,22
	CUT&RUN, CUT&Tag	- MNase is targeted by antibodies against a chromatin-bound protein of interest to promote local digestion and mapping of the genomic regions. - Protocol modifications reduce number of steps to library preparation and enable single-cell analysis. - Some CUT&RUN methods allow for spike-in normalization using carryover *E. coli* DNA, which is an important consideration in low input technologies.	24–27

* Low-input ChIP-seq techniques exist beyond STAR ChIP-seq. We highlight STAR ChIP-seq as a representative technique because it was used in several studies central to our discussion of ChIP-seq in early embryos.



Supplementary Figure 1

Supplementary Figure 1. A Polycomb relay may preserve bivalent chromatin information during preimplantation development.

a) Schematic of H3K27me3, H3K4me3 and H2AK119ub1 dynamics over the promoters of bivalent genes during early development. The promoter signal of H3K27me3 is depleted[22], perhaps because the mark occurs in diffuse domains across the genome, including intergenic regions and over gene clusters. Similarly, H3K4me3 undergoes remodeling from non-canonical domains to sharper promoter peaks at ZGA[21,28] (see Fig. 3c). Recent studies found that H2AK119ub1 is remodeled at fertilization and then regained over PRC2 targets[29,30], but an analysis of bivalent promoters was not performed.

b) CpG density in the three gene sets used in the analysis: bivalent genes, random (non-bivalent) genes or a gene set with CpG density comparable to that of the bivalent genes (3868 genes in total), based on promoter CpG annotation from Ref.[31]. Bivalent genes were identified using sequential ChIP-seq in mouse ESCs[32]. HCP, high CpG promoter; ICP, intermediate CpG promoter; LCP, low CpG promoter.

c) Profile plots of the indicated histone modifications between gametes (oocyte or sperm) over bivalent or control promoters through implantation[21,22,29], separated by maternal and paternal alleles. Data are normalized to the coverage at 10kb upstream of the transcription start site (TSS), a way to compare promoter enrichment to local background signal and estimate peak-like enrichment. Analysis suggests that many H3K27me3 and H3K4me3 promoter peaks are lost at fertilization and gradually reemerge during preimplantation development. H3K27me3 enrichment over bivalent promoters is evident in the postimplantation (E6.5) epiblast, and H3K4me3 enrichment over bivalent promoters increases starting at about the 4-cell stage. H2AK119ub1 is also reduced from gametes to the zygote stage but then display persistent enrichment over both parental alleles through preimplantation, becoming more enriched at promoters in the epiblast. Notably, bivalent promoters appear distinct from random CpG-matched genes: H3K4me3 density is reduced and H2AK119ub1 density is increased over bivalent genes compared to controls, consistent with their expected lower expression[33–35].

d) Boxplots of the indicated histone modifications ±1kb from the TSS. Datasets are the same as in part (c) and are read-depth normalized to compare promoter enrichment to genome-wide levels at a given stage. Due to lack of spike-in controls it is difficult to quantitatively assess absolute levels between samples, but the ratio of bivalent versus control promoters estimates relative coverage over developmental time. The results are largely concordant with part (c), again capturing the remodeling of both parental chromatin landscapes at fertilization and the distinct kinetics of different modifications during early development. Bivalent promoters are significantly

different from CpG-matched controls at all time points except where indicated. NS, not significant; * P < 0.05, *** P < 0.001, or, if not indicated, P < 0.0001.

Statistics are from one-way ANOVA (comparing bivalent, random or CpG-matched within a developmental stage) or post-hoc pairwise comparisons (Tukey multiple comparisons test). Flat lines highlight non-significant ANOVA, and other values indicate pairwise comparisons of bivalent to either random or CpG-matched.

METHODS

Identifying bivalent regions

The set of bivalent genes in ESCs was based on ref.[32], resulting in a list of 3868 unique genes. UCSC RefSeq (from UCSC Table Browser) was used to identify the TSS for each gene. A random subset of the same size was selected from the mm10 genome using tidyverse slice_sample(n = 3868), excluding bivalent genes from the selection pool. CpG annotation is from ref.[31]. A control gene set with a comparable number of high-CpG density promoters (HCP) was selected for comparison.

Processing sequence data

Sequencing data are from NCBI GEO. We downloaded processed files that were separated by allele and already read depth normalized (reads per million mapped reads, RPKM). When necessary, conversion from mm9 to mm10 was performed using UCSC liftOver[36].

Histone modification	Method and cell type	Data format	Processing steps	GEO accession and ref
H3K27me3	STAR ChIP-seq	BED, 100bp bin size	mm9 to mm10; BED to bedgraph; bedgraph to bigwig	GSE76687, ref[22]
H3K4me3	STAR ChIP-seq	BED	mm9 to mm10; BED to bedgraph; bedgraph to bigwig	GSE71434, ref[21]
H2AK119ub1	CUT&RUN (oocyte – E6.5) STAR ChIP-seq (sperm)	Bigwig		GSE153529, ref[29]

Promoter profile plots

Promoter coverage was calculated per bin around the promoter (± 10kb), starting from normalized bigwig files as indicated above. Reads were counted using deeptools v3.3.0 computeMatrix

(reference-point with options --referencePoint center --outFileNameMatrix <filename> -b 10000 -a 10000 --binSize 100 -p max --smartLabels). The output matrix was used as input for plotProfile (option –outFileNameData), and the counts table resulting from this command was processed using tidyverse in RStudio. To compare relative enrichment compared to background signal, we normalized samples to their upstream-most bin for this analysis. In other words, coverage in each bin was divided by reads in the -10kb bin. Normalized data were then plotted in ggplot2.

Boxplots and statistics

For this analysis, promoters were defined as TSS ± 1kb. The slop tool from bedtools v2.28.0 was used to extend promoter locations (option -b 1000)[37]. Reads were counted over bivalent or random promoters using deeptools v3.3.0 multiBigwigSummary on the command line (options BED-file --outRawCounts <filename> --chromosomesToSkip chrY chrX chrM -p max)[38]. The resulting counts file (from --outRawCounts) processed using tidyverse in RStudio. Boxplots were produced using ggplot2 and statistics were performed using rstatix. One-way ANOVA was used to make within-group comparisons (bivalent, random and CpG-matched promoters) for each developmental stage. Post-hoc significance was determined by Tukey multiple comparisons test. For Tukey tests, we highlight comparisons of bivalent versus control promoters in Supplementary Figure 1d.

DATA AND CODE AVAILABILITY

All raw data are publicly accessible at the following GEO accession codes: GSE76687 (H3K27me3), GSE71434 (H3K4me3) and GSE153529 (H2AK119ub1).

Bash and R scripts, processed data used for plotting, and statistics are all available at https://github.com/tmacster/nrmcb_bivalency_plots.

REFERENCES:

1. Sen, S., Block, K. F., Pasini, A., Baylin, S. B. & Easwaran, H. Genome-wide positioning of bivalent mononucleosomes. *Bmc Med Genomics* **9**, 60 (2016).

2. Kinkley, S. *et al.* reChIP-seq reveals widespread bivalency of H3K4me3 and H3K27me3 in CD4+ memory T cells. *Nat Commun* **7**, 12514 (2016).

3. Grzybowski, A. T., Shah, R. N., Richter, W. F. & Ruthenburg, A. J. Native internally calibrated chromatin immunoprecipitation for quantitative studies of histone post-translational modifications. *Nat Protoc* **14**, 3275–3302 (2019).

4. Grzybowski, A. T., Chen, Z. & Ruthenburg, A. J. Calibrating ChIP-Seq with Nucleosomal Internal Standards to Measure Histone Modification Density Genome Wide. *Mol Cell* **58**, 886–899 (2015).

5. Shah, R. N. *et al.* Re-evaluating the role of nucleosomal bivalency in early development. *BioRxiv* https://doi.org/10.1101/2021.09.09.458948 (2021).

6. Mauser, R., Kungulovski, G., Keup, C., Reinhardt, R. & Jeltsch, A. Application of dual reading domains as novel reagents in chromatin biology reveals a new H3K9me3 and H3K36me2/3 bivalent chromatin state. *Epigenet Chromatin* **10**, 45 (2017).

7. Delachat, A. M.-F. *et al.* Engineered Multivalent Sensors to Detect Coexisting Histone Modifications in Living Stem Cells. *Cell Chem Biol* **25**, 51-56.e6 (2018).

8. Shema, E. *et al.* Single-molecule decoding of combinatorially modified nucleosomes. *Science* **352**, 717–721 (2016).

9. Meers, M.P., Llagas, G., Janssens, D.H., Codomo, C.H. and Henikoff, S. Multifactorial chromatin regulatory landscapes at single cell resolution. Preprint at *bioRxiv* https://doi.org/10.1101/2021.07.08.451691 (2021).

10. Gopalan, S., Wang, Y., Harper, N. W., Garber, M. & Fazzio, T. G. Simultaneous profiling of multiple chromatin proteins in the same cells. *Mol Cell* **81**, 4736-4746.e5 (2021).

11. Orlando, D. A. *et al.* Quantitative ChIP-Seq Normalization Reveals Global Modulation of the Epigenome. *Cell Reports* **9**, 1163–1170 (2014).

12. Hu, B. *et al.* Biological chromodynamics: a general method for measuring protein occupancy across the genome by calibrating ChIP-seq. *Nucleic Acids Res* **43**, e132–e132 (2015).

13. Bonhoure, N. *et al.* Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization. *Genome Res* **24**, 1157–1168 (2014).

14. Fursova, N. A. *et al.* Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. *Mol Cell* **74**, 1020-1036.e8 (2019).

15. Mierlo, G. van *et al.* Integrative Proteomic Profiling Reveals PRC2-Dependent Epigenetic Crosstalk Maintains Ground-State Pluripotency. *Cell Stem Cell* **24**, 123-137.e8 (2019).

16. Macrae, T. A. & Ramalho-Santos, M. The deubiquitinase Usp9x regulates PRC2-mediated chromatin reprogramming during mouse development. *Nat Commun* **12**, 1865 (2021).

17. Percharde, M., Bulut-Karslioglu, A. & Ramalho-Santos, M. Hypertranscription in Development, Stem Cells, and Regeneration. *Dev Cell* **40**, 9–21 (2017).

18. Villaseñor, R. *et al.* ChromID identifies the protein interactome at chromatin marks. *Nat Biotechnol* **38**, 728–736 (2020).

19. Reverón-Gómez, N. *et al.* Accurate Recycling of Parental Histones Reproduces the Histone Modification Landscape during DNA Replication. *Mol Cell* **72**, 239-249.e5 (2018).

20. Petryk, N. *et al.* Genome-wide and sister chromatid-resolved profiling of protein occupancy in replicated chromatin with ChOR-seq and SCAR-seq. *Nat Protoc* **16**, 4446–4493 (2021).

21. Zhang, B. *et al.* Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553–557 (2016).

22. Zheng, H. *et al.* Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. *Mol Cell* **63**, 1066–1079 (2016).

23. Brind'Amour, J. *et al.* An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat Commun* **6**, 6033 (2015).

24. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, e21856 (2017).

25. Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* **13**, 1006–1019 (2018).

26. Hainer, S. J., Bošković, A., McCannell, K. N., Rando, O. J. & Fazzio, T. G. Profiling of Pluripotency Factors in Single Cells and Early Embryos. *Cell* **177**, 1319-1329.e11 (2019).

27. Kaya-Okur, H. S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**, 1930 (2019).

28. Dahl, J. A. *et al.* Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **537**, 548–552 (2016).

29. Chen, Z., Djekidel, M. N. & Zhang, Y. Distinct dynamics and functions of H2AK119ub1 and H3K27me3 in mouse preimplantation embryos. *Nat Genet* **53**, 551–563 (2021).

30. Mei, H. *et al.* H2AK119ub1 guides maternal inheritance and zygotic deposition of H3K27me3 in mouse embryos. *Nat Genet* **53**, 539–550 (2021).

31. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).

32. Mas, G. *et al.* Promoter bivalency favors an open chromatin architecture in embryonic stem cells. *Nat Genet* **50**, 1452–1462 (2018).

33. Brookes, E. *et al.* Polycomb Associates Genome-wide with a Specific RNA Polymerase II Variant, and Regulates Metabolic Genes in ESCs. *Cell Stem Cell* **10**, 157–170 (2012).

34. Stock, J. K. *et al.* Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol* **9**, 1428–1435 (2007).

35. Mantsoki, A., Devailly, G. & Joshi, A. CpG island erosion, polycomb occupancy and sequence motif enrichment at bivalent promoters in mammalian embryonic stem cells. *Sci Rep* **5**, 16791 (2015).

36. Kuhn, R. M., Haussler, D. & Kent, W. J. The UCSC genome browser and associated tools. *Brief Bioinform* **14**, 144–161 (2013).

37. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)* **26**, 841–842 (2010).

38. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-5 (2016).