LETTER

Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre–implantation embryos

Xiaoyu Liu[1,2,3]*, Chenfei Wang[1]*, Wenqiang Liu[1]*, Jingyi Li[1]*, Chong Li[1], Xiaochen Kou[1], Jiayu Chen[1], Yanhong Zhao[1], Haibo Gao[1,3], Hong Wang[1], Yong Zhang[1], Yawei Gao[1] & Shaorong Gao[1,2,3]

Histone modifications have critical roles in regulating the expression of developmental genes during embryo development in mammals[1,2]. However, genome-wide analyses of histone modifications in pre-implantation embryos have been impeded by the scarcity of the required materials. Here, by using a small-scale chromatin immunoprecipitation followed by sequencing (ChIP–seq) method[3], we map the genome-wide profiles of histone H3 lysine 4 trimethylation (H3K4me3) and histone H3 lysine 27 trimethylation (H3K27me3), which are associated with gene activation and repression[4,5], respectively, in mouse pre-implantation embryos. We find that the re-establishment of H3K4me3, especially on promoter regions, occurs much more rapidly than that of H3K27me3 following fertilization, which is consistent with the major wave of zygotic genome activation at the two-cell stage. Furthermore, H3K4me3 and H3K27me3 possess distinct features of sequence preference and dynamics in pre-implantation embryos. Although H3K4me3 modifications occur consistently at transcription start sites, the breadth of the H3K4me3 domain is a highly dynamic feature. Notably, the broad H3K4me3 domain (wider than 5 kb) is associated with higher transcription activity and cell identity not only in pre-implantation development but also in the process of deriving embryonic stem cells from the inner cell mass and trophoblast stem cells from the trophectoderm. Compared to embryonic stem cells, we found that the bivalency (that is, co-occurrence of H3K4me3 and H3K27me3) in early embryos is relatively infrequent and unstable. Taken together, our results provide a genome-wide map of H3K4me3 and H3K27me3 modifications in pre-implantation embryos, facilitating further exploration of the mechanism for epigenetic regulation in early embryos.

Epigenetic modifications are highly dynamic and change extensively during pre-implantation embryo development[6–9]. However, genome-wide analyses of histone modifications in pre-implantation embryos remain unclear because of the limited number of cells present in pre-implantation embryos. Recently, an ultra-low-input micrococcal nuclease-based native chromatin immunoprecipitation (ULI-NChIP) protocol was developed, allowing us to profile genome-wide histone modifications in pre-implantation embryos[3]. In the preliminary experiments, 1,000 or 200 embryonic stem (ES) cells were used to profile H3K4me3 and H3K27me3 modifications, and the results were very similar to the published data[10] (Extended Data Fig. 1a, b).

To obtain a genome-wide landscape of the histone modifications during zygotic genome activation (ZGA) and the first cell lineage differentiation in mouse pre-implantation embryos, we mapped the H3K4me3 and H3K27me3 modifications in embryos from the two-cell stage to the blastocyst stage (separated inner cell mass (ICM) and trophectoderm (TE)). Mouse metaphase II (MII) oocytes, mouse ES cells (mES cells) and trophoblast stem cells (mTSCs) were also analysed

(Fig. 1a, b, see the Methods section for the details). As expected, we observed H3K4me3 enrichment on well-known housekeeping genes, such as the *Actb* and *Copb2*, and increased H3K27me3 signal on somatic developmental genes, such as the *Hox* gene cluster[11] (Fig. 1c), and all replicates were highly comparable (Extended Data Fig. 1c–f).
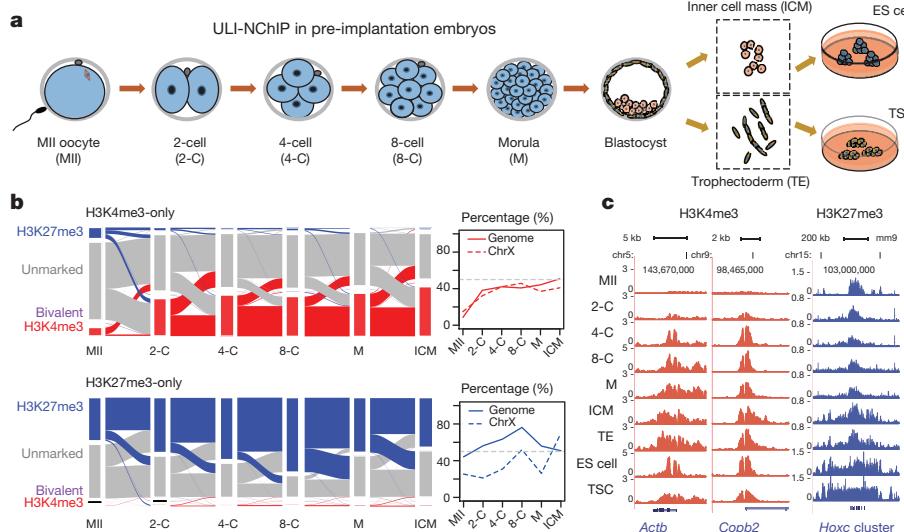
To analyse the chromatin state throughout pre-implantation development, we performed ChromHMM analysis[12]. In MII oocytes, some non-canonical flat H3K4me3 domains with low fold enrichment were widely observed. These flat H3K4me3 domains were reduced in the two-cell stage and were rarely observed in the later stages, but H3K4me3 domains with relative greater fold enrichment were established during development (Extended Data Fig. 2a). Because we mainly focused on the reestablishment of histone modifications during ZGA and the first cell lineage differentiation in pre-implantation embryos, a stringent threshold for H3K4me3 signals in ChromHMM analysis was applied to remove the flat H3K4me3 domains in this study. Subsequently, the genome was summarized to four chromatin states based on the profiles of two histone modifications: H3K4me3-only regions, H3K27me3-only regions, bivalent regions, and non-marked regions (Fig. 1b, Extended Data Figs 2b, c, 3a, b, f–i and Supplementary Table 1). In this process, the global number of H3K4me3-only regions increased in the two-cell stage when compared to the MII oocytes. This number fluctuated mildly and was dynamic mainly on relatively short, weak, and TSS (transcription start site)-distant domains (Fig. 1b and Extended Data Fig. 3c–e); in contrast, the number of H3K27me3-only regions increased to a lesser extent at each stage (Fig. 1b). The quick establishment of H3K4me3 compared to H3K27me3 is consistent with corresponding protein expression (Extended Data Fig. 2d). During the transition from the morula to the inner cell mass (ICM)/trophectoderm (TE) and the derivation of cell lines, apparent differences in these two histone modifications were also observed, and the bivalent regions increased remarkably, especially in the ES cells (Extended Data Fig. 3f–i; Supplementary Table 1). Furthermore, the continuous increase of H3K27me3 on the X chromosome from the TE to the female TSCs reflects that the repressive epigenetic marks are established during X chromosome inactivation (Extended Data Fig. 3f–i)[13,14].

Next, we focused on the histone modification dynamics at promoter regions. Compared to the MII oocytes, the number of H3K4me3-marked promoters increased dramatically at the two-cell stage, and many promoters retained this modification in the following stages till blastocyst stage. In contrast, H3K27me3 exhibited a massive loss in the two-cell-stage embryos and a greater degree of dynamics, especially in the transition from the morula to the ICM/TE; the H3K27me3-marked promoters in embryos were mainly enriched in terms that related to further development (Fig. 2a, Extended Data Fig. 4a, b and Supplementary Table 2).

In previous observations of tissue or cell lines, non-methylated CpG island (CGI) chromatin was usually enriched with H3K4me3 but not

Figure 1 | Dynamics of the histone modification landscape in early mouse embryos. a, Schematic for the pre-implantation embryos and cell line samples used in ULI-NChIP analysis. **b**, Alluvial plots (left panel) showing the global dynamics of H3K4me3-only and H3K27me3-only regions during early embryo development. Each line represents a 200-bp bin defined on the ChromHMM categories, and the total regions are those classified as this state in at least one analysed stage. The global dynamics (solid line) and chromosome X dynamics (dashed line) were plotted separately (right panel). **c**, Genome browser view of the H3K4me3 density in the *Copb* and *Actb* range and the H3K27me3 density at the *Hoxc* region in pre-implantation embryos. The ChIP–seq RPM (reads per million mapped reads) was smoothed based on the mean of 5 pixels using the UCSC genome browser.
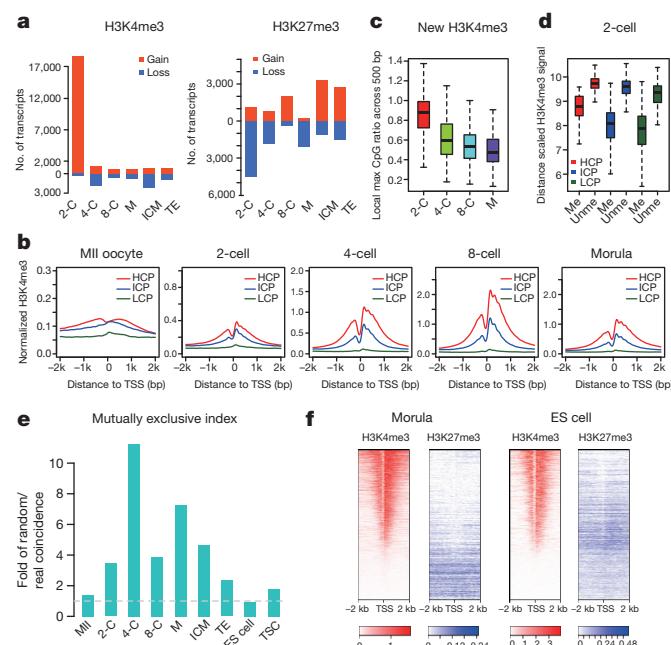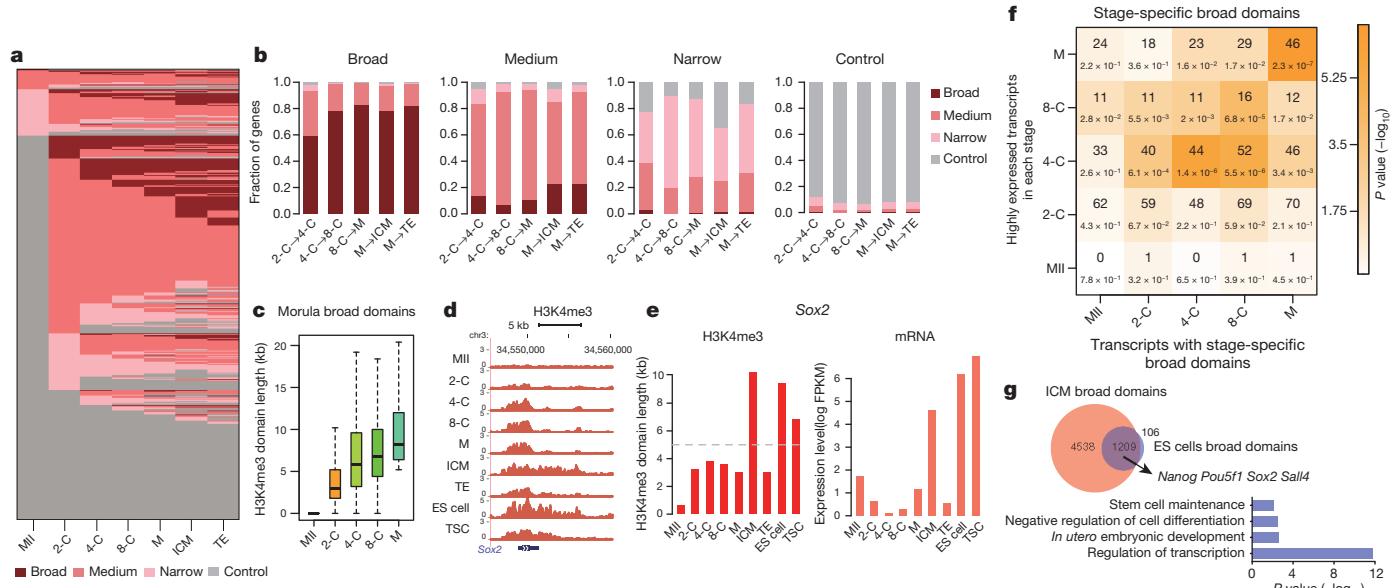
H3K27me3[15]. We then investigated the sequence preferences of these two histone marks in early embryos. All promoters were divided into high, intermediate and low CpG promoters (HCPs, ICPs and LCPs) based on the local maximum CpG ratio, as previously described[16]. The H3K4me3 signals showed a constant positive correlation with the CpG ratio in promoters (Fig. 2b, Extended Data Figs 4c, e and 5d). The CpG ratio of newly formed H3K4me3 domains in promoters was reduced during development to the morula stage (Fig. 2c). In contrast, the H3K27me3 signals showed a negative correlation with the CpG ratio (Extended Data Figs 4d, f and 5a, e), and the CpG ratio of newly H3K27me3 modified promoters increased from the two-cell to the morula stage (Extended Data Fig. 5b). The combined analysis on the DNA methylation and histone modifications suggested that DNA methylation negatively correlated with H3K4me3 modifications (Fig. 2d and Extended Data Fig. 5f), which was similar to observations in ES and other cell types[16–20]. However, the general correlation between DNA methylation and H3K27me3 in promoters was not obvious (Extended Data Fig. 5c, g). These results indicate that during early embryo development, the establishment of H3K4me3 is distinct from that of H3K27me3. The negative correlation between H3K4me3 and H3K27me3 in promoter regions also supported this hypothesis (Extended Data Fig. 5h). We calculated the mutually exclusive index between these two marks and found that this index increased in early cleavage stages but decreased in the ICM/TE, and further decreased in ES cells and TSCs. This observation suggests that the incompatible relationship between H3K4me3 and H3K27me3 weakens during the first cell lineage differentiation and the derivation of respective cell lines (Fig. 2e, f).

The H3K4me3 signals usually cover a 1–2 kb region around the TSS[21]. However, recent studies have determined that broader H3K4me3 domains are usually preferentially enriched in genes that are essential for the function and identity of a given cell type[21,22], and that these broad H3K4me3 domains may be associated with increased transcription[22]. To further investigate the potential function of H3K4me3 breadth in early embryo development, we classified the H3K4me3 domains around TSSs into three groups: broad, exceeding 5 kb; medium, approximately 1–5 kb; and narrow, less than 1 kb. The promoter regions without H3K4me3 signal were used as controls (Extended Data Fig. 6a–c and Supplementary Table 3). We found that far more broad H3K4me3 domains were present in the early embryos than in the derived cell lines or MII oocytes (Extended Data Fig. 6a–c). Moreover, the high dynamic of broad H3K4me3 occurred mainly between broad and medium across continuous stages and barely changed between broad and narrow or control (Fig. 3a, b and Supplementary Table 3). This observation indicates that broad H3K4me3 domains are established or removed gradually. We focused on the broad H3K4me3 domains in the


Figure 2 | Establishment of H3K4me3 and H3K27me3 in early mouse embryos. a, Numbers of RefSeq transcripts that gained or lost H3K4me3 (left) or H3K27me3 (right) marks in different stages were calculated by comparing to the previous stages. **b**, Profiles of the average H3K4me3 signal density on the HCPs, ICPs and LCPs. Definitions of the three classes of promoters are detailed in the Methods section. H3K4me3 signal density was calculated using H3K4me3 RPM with 50-bp resolution. **c**, Averaged local maximum of the CpG ratio of newly established H3K4me3-marked transcripts. The newly established H3K4me3-marked promoters were defined as promoters belonging to the non-marked state in the previous stage and marked by H3K4me3 in the present stage. **d**, Averaged H3K4me3 signals of the HCPs, ICPs and LCPs in the two-cell stage classified by methylation level. The published whole-genome bisulphite sequencing data were used in this analysis[20]. Genes with averaged promoter methylation levels ≥0.6 were defined as methylated, and those with levels ≤0.25 as unmethylated. The box plots in **c**, **d**, show boxes, with the middle line of the box indicating the median, outer edges representing the first and the third quartiles, and the whiskers stand for 1.5 × interquartile range below the lower quartile and above the upper quartile. **e**, Mutually exclusive index of the H3K4me3 and H3K27me3 modifications during early embryo development. A dashed line with an index equal to 1.0 indicates that no exclusive effects exist. **f**, Distribution of the H3K4me3 and H3K27me3 signals in the morula stage and ES cell samples. Each row represents a promoter region of ±2 kb around the TSS and ranked by the H3K4me3 read counts in promoters. Colours indicate ChIP–seq RPM.

Figure 3 | Broad H3K4me3 is associated with transcription regulation and cell identity. a, Dynamics of the H3K4me3 domain length on all of the RefSeq promoters throughout early embryo development. The promoters were classified based on the breadth of marked H3K4me3: broad (dark red), medium (red), narrow (pink) and control (grey). **b**, Tendencies of the four types of H3K4me3 domains in promoters in the next stage of early embryo development. Each panel represents a specific type of H3K4me3 domain from left to right. Each bar represents the types and fractions of the promoters in the next stage. **c**, Establishment of broad H3K4me3 domains in promoters at the morula stage. RefSeq promoters with broad H3K4me3 domains at the morula stage were chosen, and the associated H3K4me3 domain lengths in previous stages were plotted. **d**, Genome browser view of the H3K4me3 broad domains established at the *Sox2* gene

during early embryo development. **e**, Bar plot of the H3K4me3 domain length at the *Sox2* promoter region (left panel) and its corresponding expression level (right panel) in the embryo, ES cell and TSC samples. The dashed line represents the cut-off for the broad domains. The expression level was evaluated using an averaged $\log_2$ FPKM. **f**, The stage-specific H3K4me3 broad domains showed strong correlation with the highly expressed genes of the same stage. The stage-specific H3K4me3 broad domains were defined as genes with broad domains only in the corresponding stages. Highly expressed genes were defined as genes with an averaged $\log_2$ FPKM ≥ 3. Overlapped gene numbers and P values are indicated in the corresponding lattice. **g**, Broad domains inherited from the ICM to ES cells. A gene ontology analysis was performed on the overlapped genes, and representative genes were labelled.

morula stage and observed that their length indeed increased gradually as development progressed (Fig. 3c and Extended Data Fig. 6d). To investigate the function and possible pruning mechanism of broad H3K4me3 domains in early embryos, we microinjected a small interfering RNA (siRNA) mixture of H3K4me3 demethylases (*Kdm2b*, *Kdm5a*, *Kdm5b*, *Kdm5c* and *Kdm5d*)[23] into MII oocysts followed by intracytoplasmic sperm injection (ICSI). We found that knock-down of H3K4me3 demethylases, especially *Kdm5b*, impaired blastocyst formation on 3.5 and 4.5 days post-activation, but had little effect on the development to the morula, which was consistent with a previous study of *Kdm5b* using porcine pre-implantation embryos[24] (Extended Data Fig. 6e). We further analysed the profile of H3K4me3 and blastocyst quality with *Kdm5b* inhibition and found that lack of *Kdm5b* increased the amount of broad H3K4me3 domains and widened the H3K4me3 domains at the morula stage; furthermore, lack of *Kdm5b* caused unsuccessful lineage differentiation compared with the control and *Kdm5b*-overexpressing embryos (Extended Data Fig. 6f–i). These observations suggest that the *Kdm5b* deficiency can widen the H3K4me3 breadth in promoters and might further cause unfaithful lineage differentiation, probably through transcription regulation.

We then compared the transcriptome activity and length of the H3K4me3 domains in the promoters of individual genes. We found that genes with broad H3K4me3 domains mainly indicated a higher level of RNA expression and those genes with narrow H3K4me3 domains or without H3K4me3 signals in the promoters were almost silenced (Fig. 3d, e and Extended Data Fig. 7a). However, in cell lines, low expression levels were observed for the genes with narrow H3K4me3 domains in the promoters. The different pattern observed in the MII oocytes was possibly due to the interference of maternal RNAs (Extended Data Fig. 7a). Furthermore, we selected genes with stage-specific broad H3K4me3 domains (from the two-cell to the

morula stage) and found that they were enriched in the highly expressed genes in the corresponding stage (Fig. 3f). In the transition from the morula to the ICM/TE, the gain or loss of broad H3K4me3 in promoters was also associated with the expression dynamics in this process (Fig. 3d, e and Extended Data Fig. 7b, c, f). As observed, the average H3K4me3 domain length shrank, and the number of genes covered by broad H3K4me3 domains was substantially reduced during the ES cells and TSC derivation from the ICM and TE (Extended Data Figs 6a, b and 7d). In the ES cells and TSCs, over 90% of the broad H3K4me3 domains were inherited from the ICM and TE, which were enriched in key markers and central regulators for the corresponding cell types (Fig. 3e, g and Extended Data Fig. 7d, e, f). Some marker genes for the later development stages of the TE, such as *Elf5*, also appeared to inherit the H3K4me3 features in the TE and to spread further in the TSCs (Extended Data Fig. 7g). Our data indicate that broad H3K4me3 might be strongly associated with transcription regulation and developmental control. The dynamic nature and maintenance of broad H3K4me3 domains on promoters may play important roles in both cell differentiation *in vivo* and inheritance of lineage characteristics in cell line derivation.

Bivalent domains, which are co-occupied by the H3K4me3 and H3K27me3 marks, were first described in ES cells and were postulated to poise developmental regulator genes for later transcriptional activation upon differentiation[25,26]. Bivalency has been observed in other systems, such as primordial germ cells (PGCs) and mid-blastula transition (MBT) embryos of zebrafish, but the regulatory features might be distinct in different cases[27,28]. We analysed the dynamics of bivalent promoters and found that the total number of bivalent genes in pre-implantation embryos was much lower than that in ES cells or TSCs (Extended Data Fig. 8a, b). Moreover, approximately half of the bivalent promoters observed in early cleavage stages were not

Figure 4 | Bivalent chromatin domains in the early mouse embryos.
a, Sources of gained bivalent genes and directions of lost bivalent genes. The colours of new gained bivalent transcripts (top) indicate their status at the previous stage, and the colours of lost bivalent transcripts (bottom) indicate their present status (red, H3K4me3; blue, H3K27me3; grey, control). **b**, Broad H3K4me3 showed a higher mutually exclusive index for H3K27me3. **c**, Genome browser view of H3K4me3 and H3K27me3 density at the ES-cell-maintained bivalent gene *Hoxd9* in the ES cell and ICM

samples. **d**, Tissue specificity scores for ES-cell-activated, ES-cell-repressed and ES-cell-maintained bivalent genes. A randomly picked group of genes was used as the control, and the promoters covered by broad H3K4me3 domains in ES cells are also shown. **e**, Bar plot of the PRC2 complex binding conditions for the three types of bivalent genes defined in **d**. The PRC2-binding conditions were calculated based on *Ezh2* and *Suz12* ChIP–seq peaks derived from published mouse ES cell data.

maintained in the later stages (Extended Data Fig. 8a–c). We further found that the dynamic of bivalent promoters mostly depended on occupation or removal of H3K27me3 (Fig. 4a and Extended Data Fig. 8a). The major waves of bivalency establishment were observed during the transition from the morula to the ICM/TE and from the ICM/TE to the ES cells/TSCs, consistent with the decline of the mutually exclusive index between H3K4me3 and H3K27me3 (Figs 2e, 4a, Extended Data Fig. 8a–c and Supplementary Table 2). Additionally, bivalent genes were more likely to be associated with genes with medium and narrow H3K4me3 domains in promoters (Fig. 4b and Extended Data Fig. 8d), indicating the resistance of H3K27me3 occupation by broad H3K4me3 domains.

During differentiation, bivalent domains were newly established in the ICM and TE (Extended Data Fig. 8a, c and Supplementary Table 2). However, unlike the major decrease of H3K4me3 broad domains, more than 60% of the bivalent domains in ICM and TE were inherited during ES cell and TSC derivation (Extended Data Fig. 9a). We further analysed the ICM bivalent genes with different fates during ES cell derivation and found that nearly 72% of ICM bivalent genes maintained bivalency in the ES cells ('maintained'). Meanwhile, approximately 22% of genes lost H3K27me3 and returned to an activated state ('activated'). Only less than 1% of genes lost H3K4me3 and returned to a repressed state ('repressed') (Fig. 4c and Extended Data Fig. 9a–c). We calculated the tissue specificity scores (see Methods for details) of the indicated genes based on the expression profiles of 19 different tissues in mouse, in which a higher score indicated higher specificity and a lower score indicated pluripotency. We found that the ES-cell-activated bivalent genes had lower tissue specificity scores, were covered with higher H3K4me3 and lower H3K27me3 signals in the ICM stage and showed higher expression levels in the early embryos (Fig. 4d and Extended Data Fig. 9b, d, e). These data suggest that the ES-cell-activated bivalent genes may be related to pluripotency and a repressive mark generated at the ICM stage in these promoters could be removed in ES cell derivation. In contrast, the ES-cell-repressed bivalent genes possessed high tissue specificity scores and were covered with lower H3K4me3 and higher H3K27me3 signals in the ICM stage. They were also silenced in all stages of the early embryos (Fig. 4d and Extended Data Fig. 9c, d, e). Genes that maintained bivalency appeared to be in the 'medial' score (medial levels of H3K4me3 and H3K27me3) and had low levels of expression (Fig. 4d and Extended Data Fig. 9d, e). Moreover, the maintained bivalent genes, but not the activated or repressed bivalent genes, were mostly targeted by the two core components of polycomb repressive complex 2 (PRC2) in the ES cells (Fig. 4e, see Methods for details), suggesting that PRC2-mediated H3K27 methylation might be involved in the inheritance of H3K27me3 and in maintaining bivalency during ES cell derivation from ICM.

In summary, we present the first genome-wide map of the histone modification landscape of mouse pre-implantation embryos during ZGA and the first cell lineage differentiation. Consistent with the findings in *Xenopus* and zebrafish[29], zygotic H3K4me3 accumulated more rapidly than H3K27me3 in mouse, probably due to maternal proteins, which also suggests the potential role of maternal factors in setting a permissive mark for ZGA in mice. Our findings suggest that the dynamicity of H3K4me3 breadth may be a novel mechanism of epigenetic regulation in early cleavage embryos when most repressive H3K27me3 is absent from promoters. Because the breadth of broad H3K4me3 domains changed gradually in continuous developmental stages and was highly resistant to repressive H3K27me3, broad H3K4me3 seems to protect transcription from the occasional transient change on a promoter and to sustain the transcriptional stability of core factors in a defined cell type. As supporting evidence, a recent study demonstrated that in flies, worms and mammals, strong chromatin marking was associated with a stable RNA product, and an absence of canonical marks permits rapid on–off switching[30]. Additionally, the present data from our and other studies suggest that inherited broad H3K4me3 in derived cell lines and specific broad H3K4me3 in differentiated tissues preferentially mark key cell identity/function genes[21,22]. Therefore, broad H3K4me3 might be a convertible 'epigenetic barrier' on active transcripts during development and may possess additional functions in multiple biological processes.

1. Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat. Rev. Genet.* **3**, 662–673 (2002).
2. Bogliotti, Y. S. & Ross, P. J. Mechanisms of histone H3 lysine 27 trimethylation remodeling during early mammalian development. *Epigenetics* **7**, 976–981 (2012).
3. 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).
4. Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome regulation by polycomb and trithorax proteins. *Cell* **128**, 735–745 (2007).
5. Ringrose, L. & Paro, R. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu. Rev. Genet.* **38**, 413–443 (2004).
6. Erhardt, S. *et al.* Consequences of the depletion of zygotic and embryonic enhancer of zeste 2 during preimplantation mouse development. *Development* **130**, 4235–4248 (2003).
7. van der Heijden, G. W. *et al.* Asymmetry in histone H3 variants and lysine methylation between paternal and maternal chromatin of the early mouse zygote. *Mech. Dev.* **122**, 1008–1022 (2005).

8. Zhang, M., Wang, F., Kou, Z., Zhang, Y. & Gao, S. Defective chromatin structure in somatic cell cloned mouse embryos. *J. Biol. Chem.* **284,** 24981–24987 (2009).

9. Wang, H. & Dey, S. K. Roadmap to embryo implantation: clues from mouse models. *Nat. Rev. Genet.* **7,** 185–199 (2006).

10. Chang, G. *et al.* High-throughput sequencing reveals the disruption of methylation of imprinted gene in induced pluripotent stem cells. *Cell Res.* **24,** 293–306 (2014).

11. Soshnikova, N. & Duboule, D. Epigenetic temporal control of mouse *Hox* genes *in vivo. Science* **324,** 1320–1323 (2009).

12. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9,** 215–216 (2012).

13. Plath, K. *et al.* Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300,** 131–135 (2003).

14. Mak, W. *et al.* Mitotically stable association of polycomb group proteins Eed and Enx1 with the inactive X chromosome in trophoblast stem cells. *Curr. Biol.* **12,** 1016–1020 (2002).

15. Thomson, J. P. *et al.* CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464,** 1082–1086 (2010).

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

17. Brandeis, M. *et al.* Sp1 elements protect a CpG island from *de novo* methylation. *Nature* **371,** 435–438 (1994).

18. Hashimshony, T., Zhang, J., Keshet, I., Bustin, M. & Cedar, H. The role of DNA methylation in setting up chromatin structure during development. *Nat. Genet.* **34,** 187–192 (2003).

19. Lande-Diner, L. *et al.* Role of DNA methylation in stable gene repression. *J. Biol. Chem.* **282,** 12194–12200 (2007).

20. Wang, L. *et al.* Programming and inheritance of parental DNA methylomes in mammals. *Cell* **157,** 1735–1735 (2014).

21. Benayoun, B. A. *et al.* H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell* **158,** 673–688 (2014).

22. Chen, K. *et al.* Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nat. Genet.* **47,** 1149–1157 (2015).

23. Greer, E. L. & Shi, Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* **13,** 343–357 (2012).

24. Huang, J. *et al.* Impairment of preimplantation porcine embryo development by histone demethylase KDM5B knockdown through disturbance of bivalent H3K4me3-H3K27me3 modifications. *Biol. Reprod.* **92,** 72 (2015).

25. Azuara, V. *et al.* Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* **8,** 532–538 (2006).

26. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125,** 315–326 (2006).

27. Vastenhouw, N. L. *et al.* Chromatin signature of embryonic pluripotency is established during genome activation. *Nature* **464,** 922–926 (2010).

28. Sachs, M. *et al.* Bivalent chromatin marks developmental regulatory genes in the mouse embryonic germline *in vivo. Cell Reports* **3,** 1777–1784 (2013).

29. Paranjpe, S. S. & Veenstra, G. J. Establishing pluripotency in early development. *Biochim. Biophys. Acta* **1849,** 626–636 (2015).

30. Pérez-Lluch, S. *et al.* Absence of canonical marks of active chromatin in developmentally regulated genes. *Nat. Genet.* **47,** 1158–1167 (2015).

Author Contributions Y.G. and S.G. conceived and designed the experiments. X.L. performed the ChIP experiments. C.W. performed computational analysis. W.L. performed the micromanipulation of embryos. J.L., C.L., X.K., J.C., Y.Z., H.G. and H.W. assisted with the sample preparation. X.L., C.W., Y.Z. and Y.G. designed and performed the data analysis. X.L., C.W., Y.G. and S.G. wrote the manuscript.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. The box plots shown in Figs 2c, d, 3c, 4d and Extended Data Figs 2d, 3c–e, 4e, f, 5b, c, 6h, 7a, c, 9d, e consist of boxes, with the middle line of the box indicating the median, outer edges representing the first and the third quartiles, and the whiskers stand for $1.5 \times$ interquartile range below the lower quartile and above the upper quartile.

Animals and mouse embryo collection. Specific-pathogen-free (SPF) mice were housed in the animal facility at Tongji University, Shanghai, China. All animal maintenance and experimental procedures were performed according to Tongji University Guide for the use of laboratory animals.

B6D2F1 female mice (8–10 weeks old) were super-ovulated by injection with 7 IU of pregnant mare serum gonadotropin (PMSG), followed by injection of 5 IU of human chorionic gonadotropin (hCG) (San-Sheng Pharmaceutical Co. Ltd,) 48 h later. The super-ovulated female mice were mated with male mice. Then, the zygotes or 2-cell stage embryos were collected from the oviducts of female B6D2F1 mice. To obtain 4-cell, 8-cell, morula and blastocyst stage embryos, 2-cell stage embryos were cultured in Q2 medium to reach the corresponding stage.

Sample harvest for ChIP-seq. Samples of MII oocytes; 2-cell, 4-cell, and 8-cell stages; morula; inner cell mass (ICM) and trophectoderm (TE) of day 3.5 blastocysts; an ES cell line (R1, male) and a TSC line (RO-7, female) were harvested for ChIP sequencing. For the MII oocytes and cleavage-stage embryos, the zona pellucidae of the embryos were removed with 0.5% pronase E (Sigma), and the embryos were then incubated in Ca^{2+}-free Chatot-Ziomek-Bavister medium (CZB) for 5 min. Polar bodies were removed by gentle pipetting using a fire-polished glass needle with an inner diameter of $120\,\mu m$. For ICM and TE isolation, the zona pellucidae of blastocysts were removed with 0.5% pronase E (Sigma). The embryos were then incubated in Ca^{2+}-free CZB for 20 min, and the tight junctions of TE cells and ICM cells were separated by gently pipetting using a pipette with a diameter of $40–60\,\mu m$.

Cell culture. The R1 ES cells were purchased from the American Type Culture Collection (ATCC) and not further authenticated. The RO-7 TSCs were derived in our laboratory, verified by immunofluorescent staining and real-time PCR. All cell lines regularly tested negative for mycoplasm contamination. The R1 ES cells were cultured on mitomycin-C-treated MEFs in ES medium containing DMEM (Merk Millipore) supplemented with 15% (v/v) fetal bovine serum (Hyclone), 1 mM L-glutamine (Merk Millipore), 0.1 mM mercaptoethanol (Merk Millipore), 1% nonessential amino acid stock (Merk Millipore), penicillin/streptomycin ($100\times$, Merk Millipore), nucleosides ($100\times$, Merk Millipore) and 1,000 U/ml LIF (Merk Millipore). The RO-7 TSCs were cultured in in 70% FCM (feeder condition medium) + F4H (25 ng/ml fibroblast growth factor 4 (FGF4); Invitrogen and $1.0\,\mu g/ml$ heparin; Sigma) medium composed of 30% TS medium (RPMI1640 supplemented with 20% FBS, 1 mM sodium pyruvate, $100\,\mu M$ β-mercaptoethanol, 2 mM L-glutamine (Merk Millipore)), 70% FCM and F4H.

ULI-NChIP-seq. For ULI-NChIP-seq, 500 cells were used per reaction, and two or three replicates were performed for each stage. All isolated cells were washed three times in 0.5% bovine serum albumin in phosphate-buffered saline (BSA-PBS, Sigma) solution to avoid potential contamination. The ULI-NChIP procedure was performed as previously described[3]. One microgram of either histone H3K4me3 antibody (Cell signaling Technology, #9727) or histone H3K27me3 antibody (Diagnode, pAb-069-050) was used for each immunoprecipitation reaction.

The sequence libraries were generated using the KAPA Hyper Prep Kit for the Illumina platform (kk8504), following the manufacturer's instructions. Paired-end 125- or 100-bp sequencing was performed on a HiSeq 2500 or 2000 (Illumina) at the Berry Genomics Co. Ltd, National Institute of Biological Sciences (NIBS), and Peking University, and paired-end 150-bp sequencing was performed on a HiSeq X Ten (Illumina) at the Cloudhealth Medical Group Ltd.

Knock-down of H3K4me3 demethylases in early embryos. The siRNAs against mouse demethylases *Kdm2b*, *Kdm5a*, *Kdm5b*, *Kdm5c* and *Kdm5b* were diluted in nuclease-free water to $20\,\mu M$. The siRNAs for a given gene were mixed for knock-down, with a working concentration of $5\,\mu M$. Oocytes were injected with approximately 10 pl of siRNAs using a Piezo-driven micromanipulator. The injected oocytes were then incubated for at least 3 h for ICSI. To inject fresh spermatozoa, approximately $1\,\mu l$ of the incubated sperm suspension was mixed with a drop of HEPES–CZB containing 10% (w/v) polyvinylpyrrolidone (PVP; Irvine Scientific, Santa Ana, CA, USA). The sperm head was separated from the tail by application of several Piezo pulses, and the head was then injected into the oocyte according to the method described by Kimura and Yanagimachi[31]. The injected embryos were observed and summarized from the 2-cell stage to 3.5 and 4.5 days post-activation (dpa). The expanded blastocysts at 3.5 dpa and

the hatched blastocysts at 4.5 dpa were also calculated. HEPES-buffered CZB (HCZB)[31] medium was used for gamete handling and ICSI in air. CZB[32] medium was used for embryo culture in an atmosphere of 5% CO_2. For embryo incubation, CZB was overlaid with sterile mineral oil (Sigma).

siRNA sequences (Supplementary Table 4) were designed using GE Dharmacon siDESIGN Center (http://dharmacon.gelifesciences.com/design-center/) and were synthesized by the Biological Resource Center at NIBS.

ChIP-seq, RNA-seq and BS-seq data processing. ChIP–seq reads were aligned to the mouse genome build mm9 using the bwa (v0.7.12)[33] mem command. Signal tracks for each sample were generated using the MACS2 (v2.0.10.20131216)[34] pileup function and were normalized to 1 million reads for visualization. To examine the reproducibility of the ChIP–seq experiments, we calculated the correlation of the normalized signal intensity between biological replicates on all RefSeq gene promoters, defined as ±2 kb around the TSS. We pooled the biological replicates together for each stage and performed the downstream analysis. The quality information of ChIP data used in this study and the correlation of replicates in each stage are summarized in Supplementary Table 5. The RNA-seq reads were mapped to the mm9 reference genome using TopHat (v1.3.3)[35]. Expression levels for each sample were quantified to FPKM using Cufflinks (v1.2.0)[36], and FPKM values of replicates were averaged. We took advantage of the published mouse WGBS data at the 2- and 4-cell stages[20] and also generated RRBS data (which was published in another project[44]) from the 2-cell to blastocyst stage. All of the BS-seq reads were first processed using TrimGalore (v0.3.3) to trim adaptor and low-quality reads. Adaptor-trimmed reads were then mapped to a combined genome with mm9 and 48052 lambda sequence using bsmap (v2.89)[37]. The methylation level of each CpG site was estimated using mcall (v1.3.0)[38].

ChromHMM analysis and alluvial plotting. Chromatin states were identified and characterized using ChromHMM (v1.11)[12]. The alignment files of H3K4me3 and H3K27me3 modifications across 7 developmental stages and in ES cells and TSCs were binned into 200-bp bins using the BinarizeBam command, with the input alignment file as the control. We then applied a stringent threshold (fold enrichment greater than four using -f 4 option) for H3K4me3 samples to remove the low-enrichment domains. Next, we trained the model with 4 emission states using 200-bp resolution and default parameters using the LearnModel command. Finally, at each stage, the whole genome was classified into four states: H3K4me3 only (without H3K27me3) region, H3K27me3 only (without H3K4me3) region, bivalent region and non-marked region.

The segmentation file of each stage was further binned to 200-bp intervals to calculate the number of transitions between chromatin states during development or differentiation. Alluvial diagrams of developmental and differentiation lineages were plotted using the alluvial function in R to show the transitions. The total regions of each chromatin state were counted once the 200-bp intervals had been marked by the specific states during that lineage, and the percentages of the specified intervals in each stage were plotted to show the global trend of that specific chromatin state. The alluvial diagrams showed the percentage changes of chromatin states during each transition; the lines from the present stage to next stage cannot be traced, as they represent different genomic locations.

Transcript classification and heat map visualization. We classified the RefSeq transcripts based on their promoters overlapping with ChromHMM segments in each stage. Promoters were defined as ±2 kb around the TSS. Transcripts were defined as follows: H3K4me3-marked transcripts once they had >200-bp overlaps with the H3K4me3-only regions or bivalent regions; H3K27me3-marked transcripts once they had >200-bp overlaps with H3K27me3-only regions or bivalent regions; bivalent-marked transcripts once they had >200-bp overlaps with bivalent regions, or they overlapped with both H3K4me3-only and H3K27me3-only regions.

To visualize transcripts based on their chromatin states, we first transformed the transcript matrices of the developmental stages into 0,1 matrices based on their chromatin status, with 1 representing H3K4me3-marked transcripts and 0 representing non-marked transcripts, with the same criteria for the H3K27me3 matrix. For each transcript, we arranged the column based on the order of developmental stages and transformed the binary 0,1 value into a decimal score. All of the transcripts were then ranked based on their score to generate heat maps. A similar quaternary transition method was used for bivalent and H3K4me3 breadth visualization.

ChIP–seq signal intensity evaluation. We defined a distance scaled signal, p_i, to present the ChIP–seq signal intensity of a certain promoter.

$$\rho_i = \sum_k w_k s_k$$

p_i is the weighted sum of ChIP–seq reads s_k at genomic positions k, where their weights decrease with distance from the TSS of transcription i.

In this definition,

$$w_k = \frac{2e^{-\mu|k-t_i|}}{1 + e^{-\mu|k-t_i|}}$$

And t_i is the genomic position of the TSS for transcript i. The parameter μ determines the decay rate, which is a function of the distance from the TSS. For H3K4me3 and H3K27me3 markers, we set μ to be 2 kb from the TSS and the contribution of the corresponding signal decayed to 1/2 of that at the TSS.

Genomic annotation and local CpG ratio calculation. Local CpG ratio was calculated for 500-bp bins with 50-bp steps, as previously defined. The CpG ratio for each transcript was calculated as the max local CpG ratio around ±2 kb of the TSS. The transcripts were then separated into high-CpG-density promoters (HCPs), intermediate-CpG-density promoters (ICPs) and low-CpG-density promoters (LCPs) based on the CpG ratio and the GC content cut-off previously defined[39].

H3K4me3 marked transcript classification. We classified the H3K4me3-marked transcripts into four groups based on the segment lengths at their promoter regions. Transcripts overlapping with segments >5,000 bp were defined as broad H3K4me3 domain transcripts; those overlapping with segments ≤5,000 bp and >1,000 bp were defined as median H3K4me3 domain transcripts; those overlapping ≤1,000 bp and >200 bp were defined as narrow H3K4me3 domains; those overlapping ≤200 bp were defined as control H3K4me3 domains. We defined stage-specific broad domain transcripts as the broad domains that only occurred at that specific stage.

Mutually exclusive index. The global mutually exclusive index was calculated as the random coincidence of an H3K4me3 transcript and an H3K27me3 transcript versus real coincidence, $m_s = B_r/B_v$. Where

$$B_r = \frac{n_{K4me3} \times n_{K27me3}}{n_{total}}$$

$$B_v = n_{bivalent}$$

in which n_{K4me3} stands for the number of H3K4me3-marked transcripts in that stage, n_{K27me3} stands for the number of H3K27me3-marked transcripts, and $n_{bivalent}$ stands for the bivalent transcripts in that stage. For mutually exclusive indexes in each H3K4me3 transcript sub class, the real coincidence is the number of bivalent genes belonging to that specific sub-class.
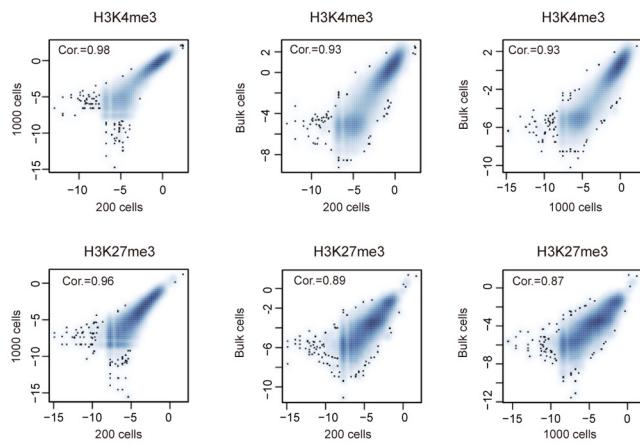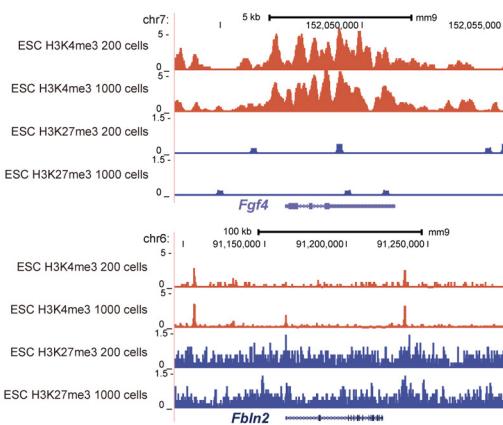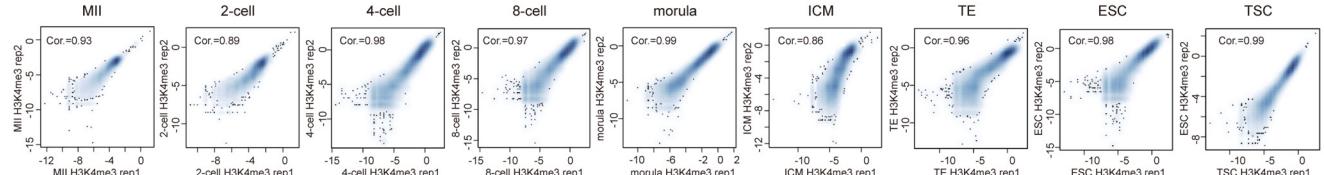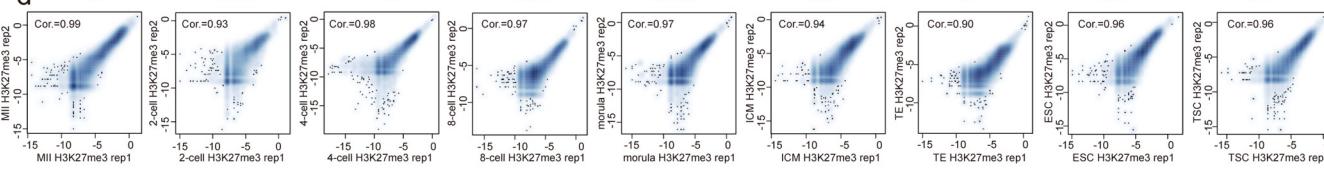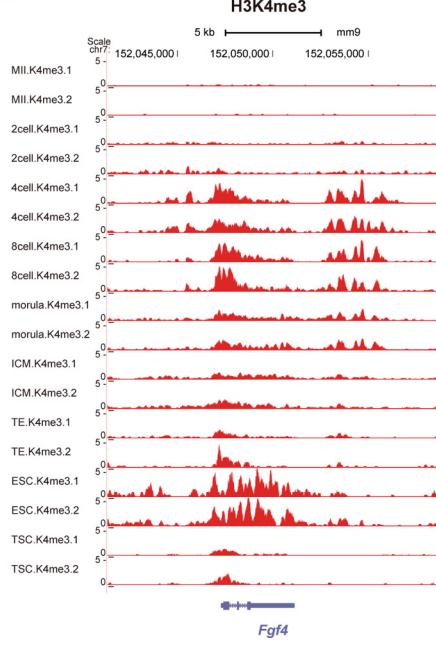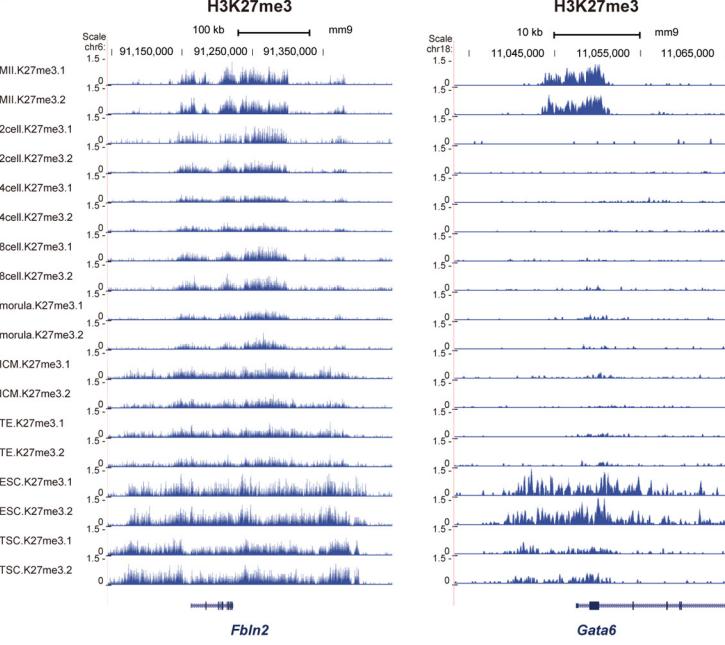
Gene ontology analysis. Functional annotation was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Resource[40]. Gene ontology terms for each functional cluster were summarized to a representative term, and P values were plotted to show the significance.

Tissue specificity score. We collected expression data from 19 mouse tissues to calculate tissue specificity scores[41]. We computed the Jensen–Shannon divergence between the relative expression levels across the tissue types. An extreme condition is defined as a transcript expressed in only one tissue type. The maximum tissue specificity score across all tissue types was defined as the specificity score for the transcript[42].

PRC2 complex targets. ChIP–seq data of Ezh2 and Suz12 in mouse embryonic stem cells were collected from the GEO database (www.ncbi.nlm.nih.gov/geo/)[43]. Ezh2 data were collected from GSE23943, GSE43231, GSE41314, GSE49178 and GSE55967; Suz12 data were collected from GSE11724, GSE13084, GSE19365, GSE39513 and GSE55967. All of the ChIP–seq data were processed as previously described. Peaks were called using MACS2 (v2.0.10.20131216) with the call peak function using the following parameters: macs2 callpeak –n Sample -g mm -B -q 0.05–nomodel–broad–shiftsize = 73–SPMR. For each data set, we removed the peaks with less than fivefold enrichment over control and merged the peaks from different data sets using bedtools to generate the final peak lists for Ezh2 and Suz12. Transcripts were defined as double-binding if both the Ezh2 and the Suz12 peaks overlapped with their promoter regions; as single-binding if only one of the two factors overlapped; and as having no PRC-related factor binding if none of the factor peaks overlapped with their promoter regions.
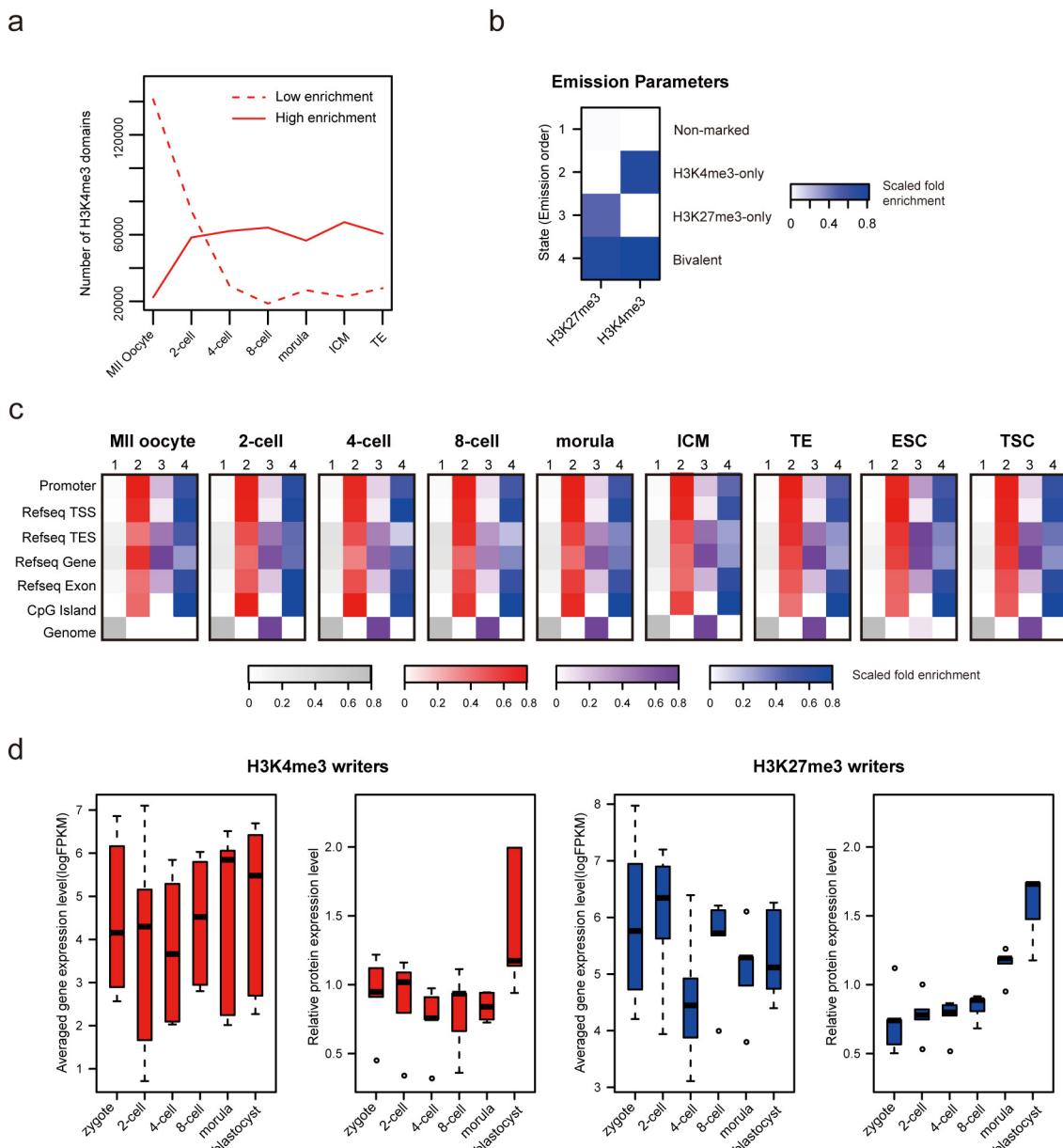
31. Kimura, Y. & Yanagimachi, R. Intracytoplasmic sperm injection in the mouse. *Biol. Reprod.* **52,** 709–720 (1995).
32. Chatot, C. L., Ziomek, C. A., Bavister, B. D., Lewis, J. L. & Torres, I. An improved culture medium supports development of random-bred 1-cell mouse embryos *in vitro. J. Reprod. Fertil.* **86,** 679–688 (1989).
33. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
34. Zhang, Y. *et al.* Model-based analysis of ChIP–seq (MACS). *Genome Biol.* **9,** R137 (2008).
35. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).
36. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28,** 511–515 (2010).
37. Xi, Y. & Li, W. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* **10,** 232 (2009).
38. Sun, D. *et al.* MOABS: model based analysis of bisulfite sequencing data. *Genome Biol.* **15,** R38 (2014).
39. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39,** 457–466 (2007).
40. Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocols* **4,** 44–57 (2009).
41. Shen, Y. *et al.* A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **488,** 116–120 (2012).
42. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25,** 1915–1927 (2011).
43. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.* **41,** D991–D995 (2013).
44. Liu, W. *et al.* Identification of key factors conquering developmental arrest of somatic cell cloned embryos by combining embryo biopsy and single-cell sequencing. *Cell Discovery* **2,** 16010 (2016).

Extended Data Figure 1 | See next page for caption.
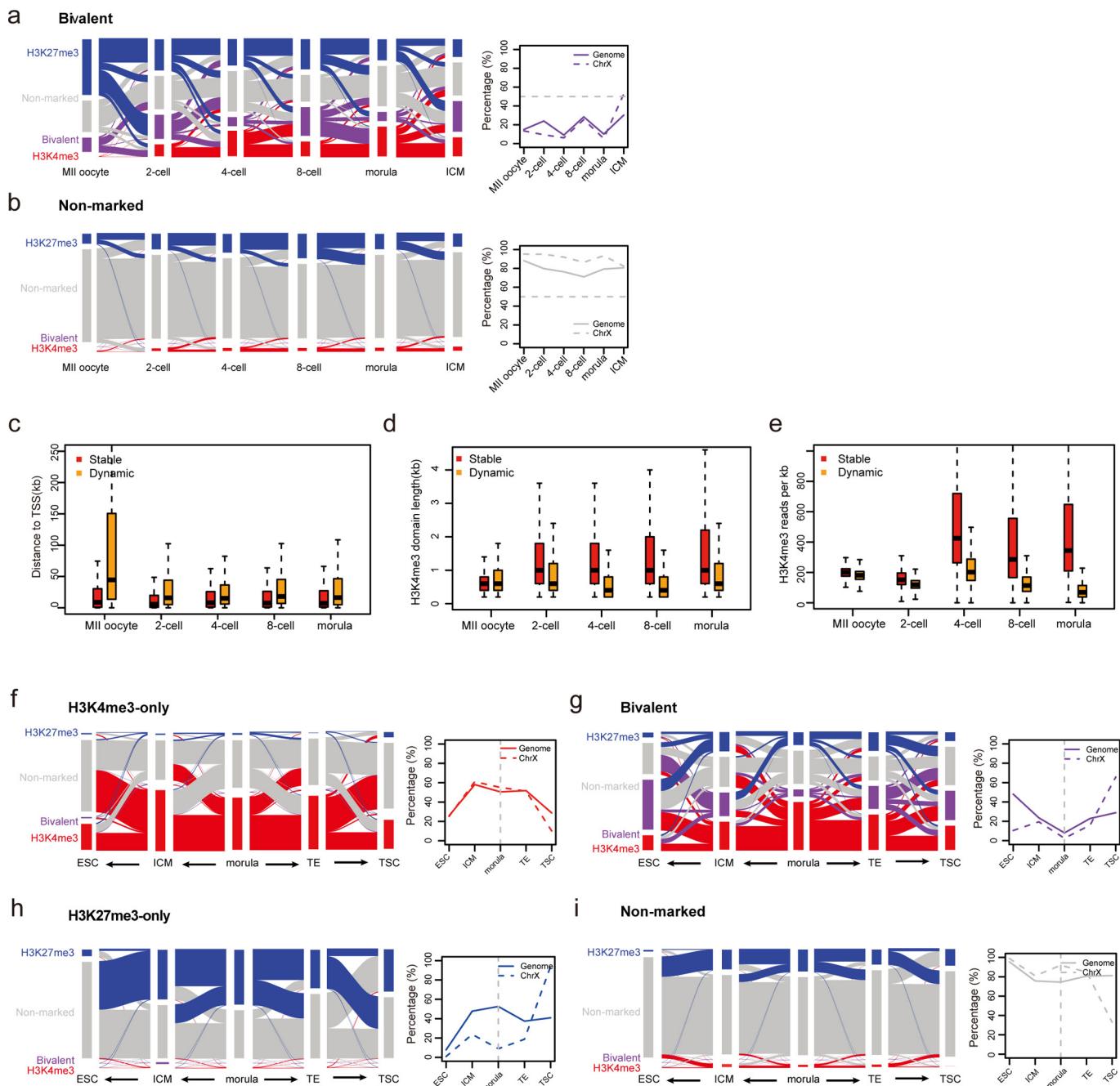
Extended Data Figure 1 | Ultra-low-input ChIP–seq in ES cells and pre-implantation embryos. a, Sample correlation for the ChIP–seq experiments using 200, 1,000 and bulk mES cells. Each point represents the normalized reads count of a RefSeq promoter for corresponding samples. Pearson's correlation coefficients of the two experiments were calculated using all the RefSeq promoters. **b**, Genome browser view of the H3K4me3 and H3K27me3 RPM (reads per million reads) at the *Fgf4* and *Fbln2* gene regions for the mES cells in duplicate. The ChIP–seq RPM (reads per million mapped reads) was smoothened by the mean of five pixels using the UCSC genome browser. **c**, Correlation of the H3K4me3 samples for embryo developmental stages and cell lines.

Each point represents the normalized reads count of a RefSeq promoter for the replicates. Pearson's correlation coefficients of the two replicates were calculated using all the RefSeq promoters. **d**, Correlation of the H3K27me3 samples for embryo developmental stages and cell lines. Each point represents the normalized reads count of a RefSeq promoter for the replicates. Pearson's correlation coefficients of the two replicates were calculated using all the RefSeq promoters. **e, f**, Genome browser view of the H3K4me3 RPM at the *Fgf4* (**e**) and the H3K27me3 RPM at the *Fbln2* and *Gata6* (**f**) gene regions for embryos and cell lines. The ChIP–seq RPM was smoothened by the mean of five pixels using the UCSC genome browser.
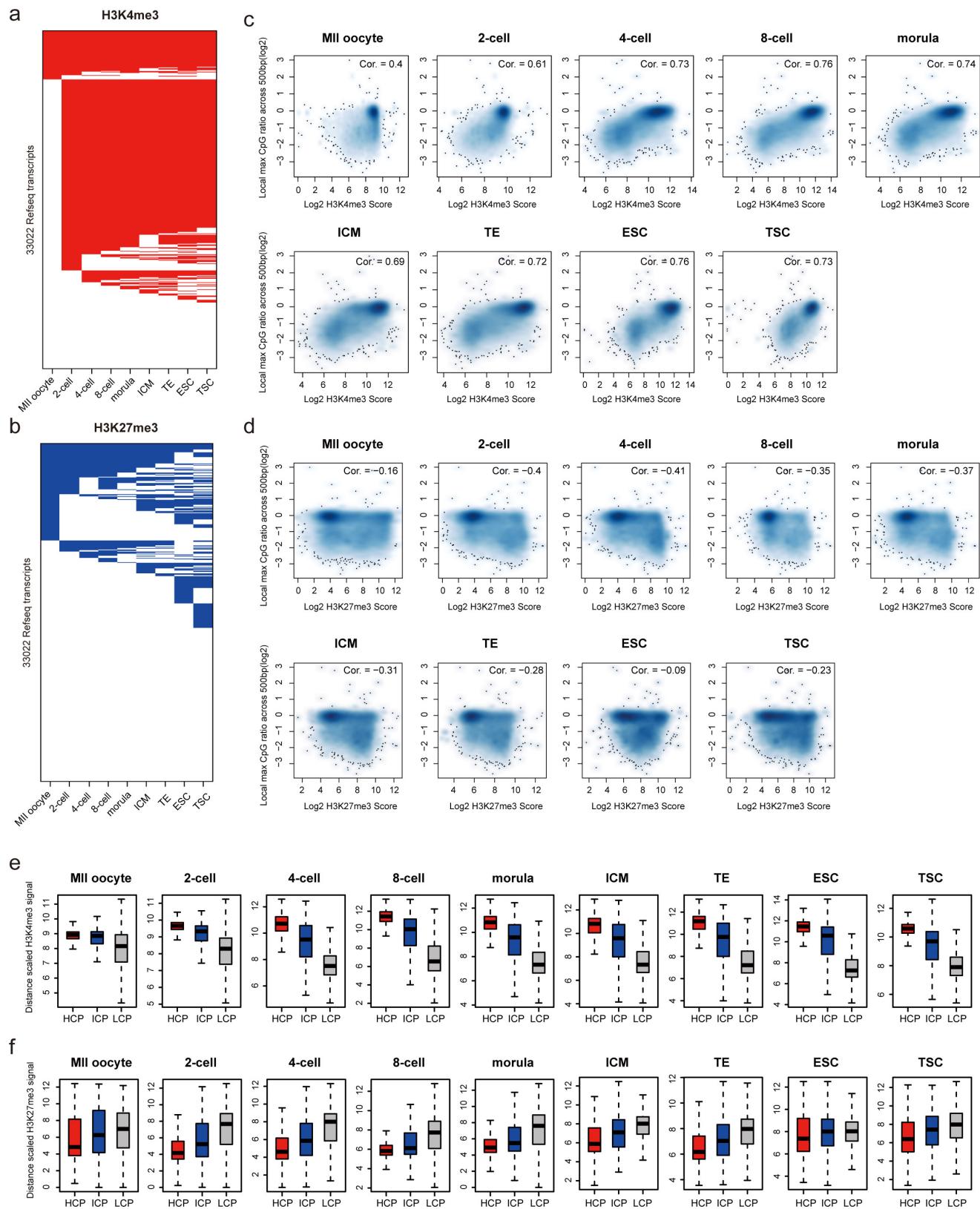
Extended Data Figure 2 | Features of the chromatin states in pre-implantation embryos. a. Number of high-enrichment and low-enrichment H3K4me3-only regions during embryogenesis. The high-enrichment H3K4me3-only domains were defined as segments retained using a stringent ChromHMM threshold (fourfold over input file) and the low-enrichment H3K4me3-only domains were the additional segments when the stringent threshold was not used. b, Emission parameters of the ChromHMM segmentations. Colour indicates the enrichment of chromatin modifications in each specific segmentation. The whole genome was classified into: 1, non-marked regions; 2, H3K4me3-only regions; 3, H3K27me3-only regions; and 4, bivalent regions. c, Genome distribution of the different classes of the ChromHMM segmentations across the mouse developmental stages and cell lines. The analysed regions are labelled to the left of the heat map, and segmentations are labelled at the top of the heat map. Colour represents relative percentage scaled by row. d, RNA expression level and relative protein level of the H3K4me3 and H3K27me3 writers. RNA-seq data and mass spectrometric data on early mouse embryos (unpublished) were used in this analysis. The expression level was evaluated using a log-scaled FPKM, and the protein level was evaluated using the relative ratio over all the stages. In this figure, H3K4me3 writers included *Rbbp5*, *Dpy30*, *Ash2l*, *Wdr5* and *Setd1a*. H3K27me3 writers included *Jarid2*, *Ezh2*, *Eed*, *Suz12* and *Rnf2*.
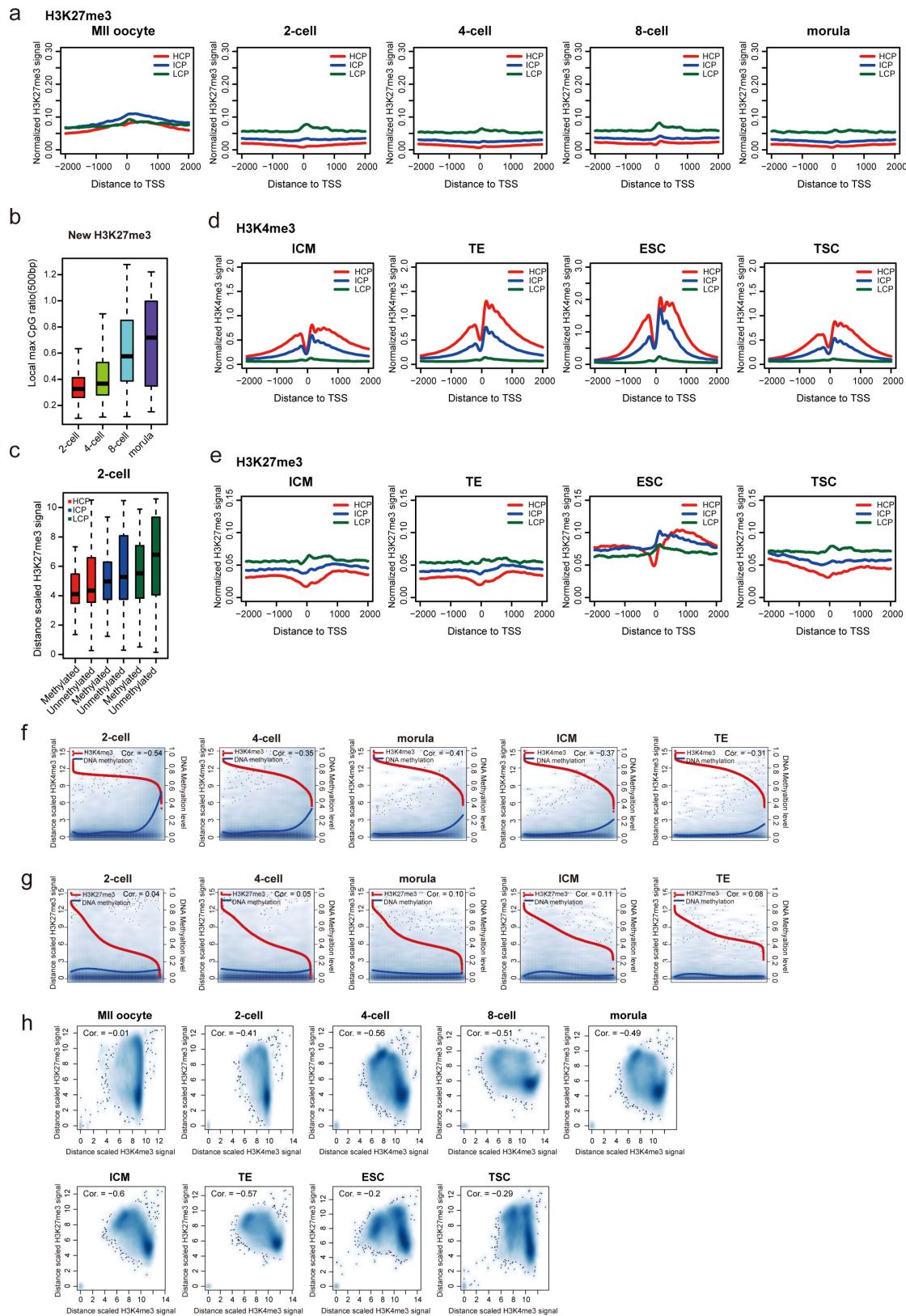
Extended Data Figure 3 | Dynamics of the histone-modification landscape in embryogenesis and during differentiation. a, b, Left panel, alluvial plots showing the global dynamics of the bivalent (**a**) and non-marked (**b**) regions during early embryo development. Each line represents a 200-bp bin defined based on the ChromHMM categories. For each panel, the total regions were those classified as this state in at least one analysed stage. Right panel, the percentage of the analysed states (that is, bivalent or non-marked) in each stage compared to the total regions. The global dynamics (solid line) and chromosome X dynamics (dashed line) are plotted separately. **c,** Averaged distance to the nearest TSS of the stable and dynamic H3K4me3-only marked regions. The stable H3K4me3-only regions were defined as those regions that retained the H3K4me3 mark in the next stage, while the dynamic H3K4me3-only regions were defined as those regions that lost the modification in the next

stage. **d,** Averaged domain length of the stable and dynamic H3K4me3-only marked regions. **e,** Averaged signal density of the stable and dynamic H3K4me3-only marked regions. The signal density was calculated as H3K4me3 ChIP–seq RPKM (reads per million reads per kb). **f–i,** Left panel, alluvial plots showing the global dynamics of H3K4me3-only (**f**), bivalent (**g**), H3K27me3-only (**h**) and non-marked (**i**) regions during the first cell lineage differentiation and cell line derivation. Each line represents a 200-bp bin defined according to the ChromHMM categories. For each state, the total regions were those classified as this state in at least one analysed stage. The morula stage was plotted in the middle, and the two differentiated lineages were plotted on either side. Right panel, the percentage of the analysed states in each stage compared to the total regions. The global dynamics (solid line) and chromosome X dynamics (dashed line) are plotted separately.

Extended Data Figure 4 | Establishment and dynamics of H3K4me3 and H3K27me3 on promoters. a, **b**, Heat map of the H3K4me3-marked (**a**) and H3K27me3-marked (**b**) transcript dynamic during early mouse embryo development and differentiation. The H3K4me3-marked transcripts were defined with an overlap of at least 200 bp with the ChromHMM H3K4me3 segmentations, including the H3K4me3-only and bivalent segmentations. H3K27me3-marked transcripts were defined as having an overlap of at least 200 bp with the ChromHMM H3K27me3 segmentations, including the H3K27me3-only and bivalent segmentations. The transcript clustering was based on the binary score; see the Methods section for the detailed clustering method. **c**, **d**, Correlation between the H3K4me3 (**c**) or H3K27me3 (**d**) signal intensity and local CpG ratio in all pre-implantation developmental stages and cell lines. ChIP–seq signal intensity was evaluated using a weighted ChIP–seq signal of ±2 kb around the TSS. Pearson correlation coefficients were calculated and are shown in the top right of each panel. **e**, **f**, Averaged H3K4me3 (**e**) and H3K27me3 (**f**) signal intensity on the HCPs, ICPs and LCPs throughout all pre-implantation developmental stages and cell lines. The definition of the three class promoters is described in the Methods.
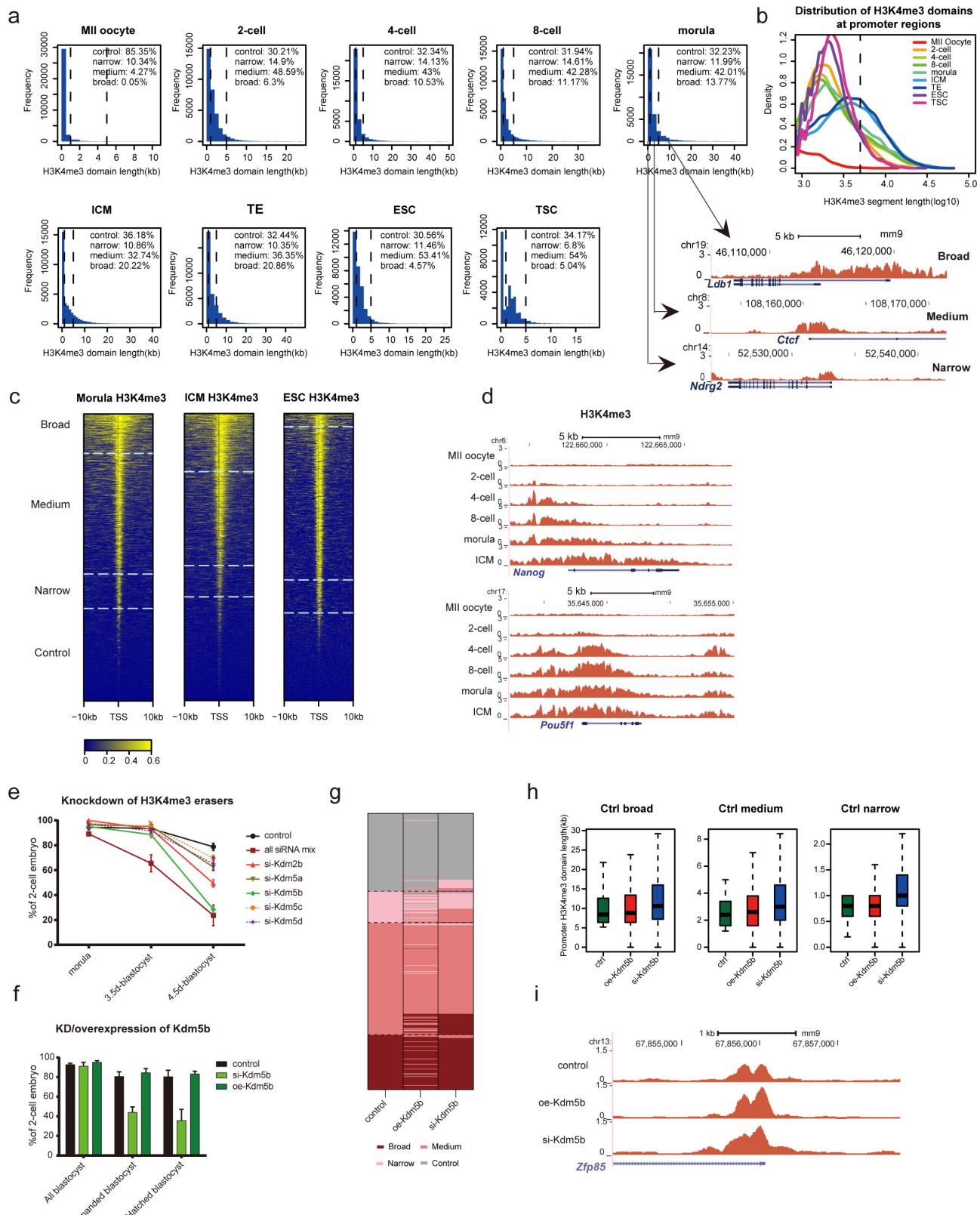
Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Distinct features of the establishment of H3K4me3 and H3K27me3 in embryogenesis and differentiation.
a, Profiles of the H3K27me3 signal density on the HCPs, ICPs and LCPs from MII to morula stage. The H3K27me3 signal density was calculated as H3K27me3 RPM with 50-bp resolution. **b**, Averaged local max CpG ratio of the new H3K27me3-marked transcripts. The newly established H3K27me3-marked transcripts were defined as transcripts belongs to non-marked state in the previous stage and marked by H3K27me3 in the present stage. **c**, Averaged H3K27me3 signals of the HCPs, ICPs and LCPs in the two-cell stage, classified by the methylation level. The published whole-genome bisulphite sequencing data was used in this analysis[20]. The methylated genes were defined as genes with an averaged promoter methylation level ≥ 0.6, and the unmethylated genes were defined as genes with an averaged promoter methylation level ≤ 0.25. **d**, **e**, Profiles of the H3K4me3 (**d**) and H3K27me3 (**e**) signal density on the HCPs, ICPs and LCPs in ICM, TE, ES cells and TSCs. The H3K4me3
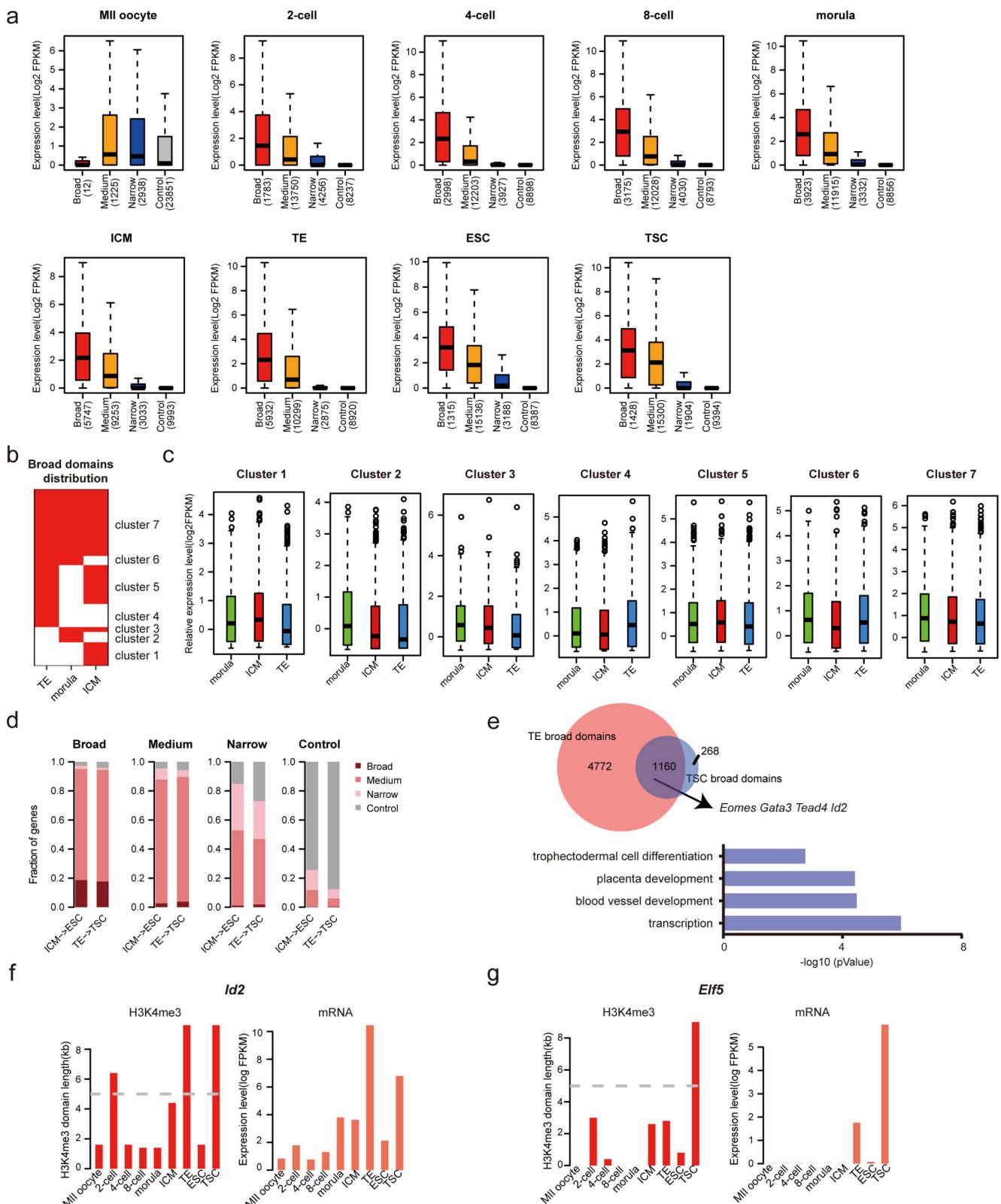
and H3K27me3 signal density was calculated as H3K27me3 RPM with 50-bp resolution. **f**, **g**, Scatter plot of the signal intensities of H3K4me3 (**f**) and H3K27me3 (**g**) with the promoter regions and DNA methylation levels. Our reduced representation bisulphite sequencing (RRBS) data published in another project[44] was used in this analysis. The red and blue fitting curves represent the ChIP–seq intensity and methylation level in the corresponding regions, respectively. The horizontal axis, from left to right, represents the transcripts ranked from high to low according to their promoter H3K4me3 or H3K27me3 intensities, and the y axis indicates the averaged methylation level of corresponding promoter Pearson correlations of the H3K4me3 intensities and DNA methylation levels were calculated and included in the panel **f** and Pearson correlations of the H3K27me3 intensities and DNA methylation levels in panel **g**.
h, Correlation of H3K4me3 and H3K27me3 signal intensity throughout early embryo developmental stages and cell lines.

Extended Data Figure 6 | See next page for caption.
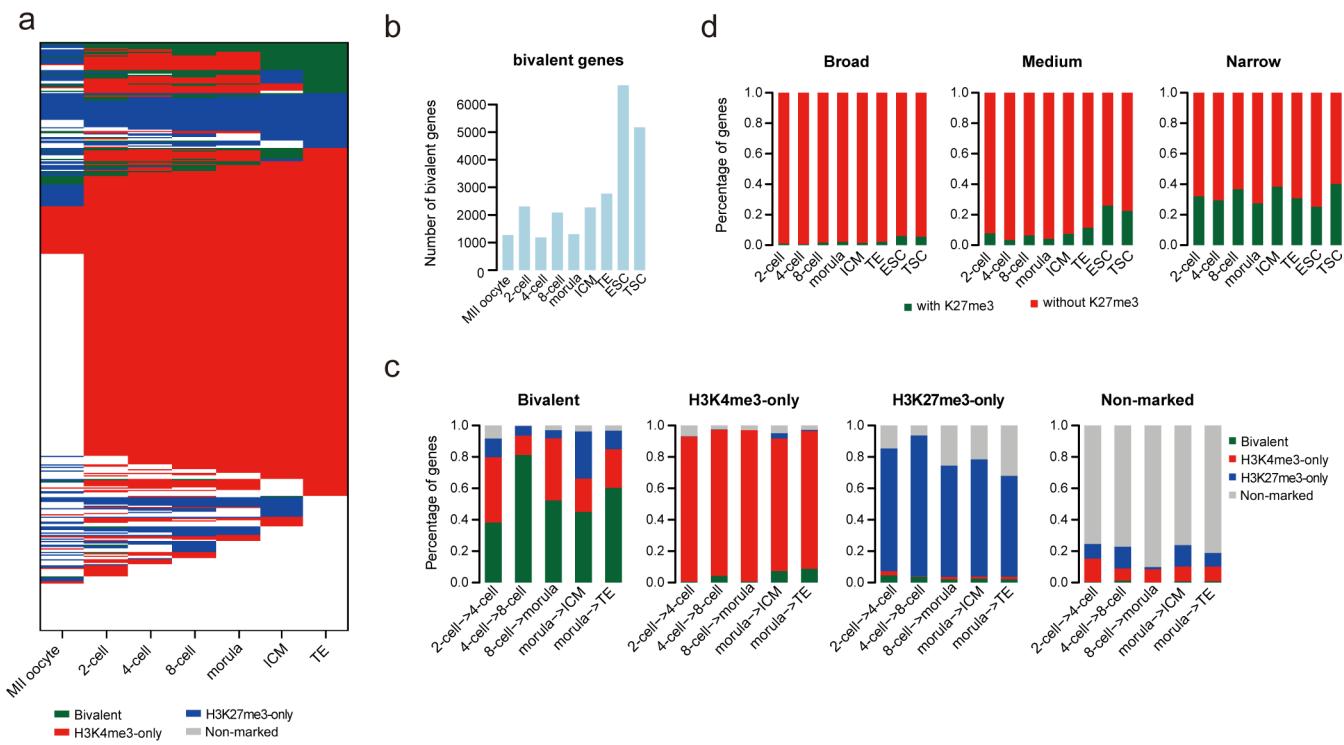
Extended Data Figure 6 | Identification and regulation of H3K4me3 broad domains on promoters. a, Bar plot showing the distribution of the H3K4me3 domain length at the promoter regions during mouse development. The two dashed lines represent the cut-offs between the plotted narrow, medium and broad domains. The percentages of transcripts classified into these three groups and control are listed in the top right of the panel. The representative genes *Ndrg2*, *Ctcf* and *Ldb1*, whose promoters were covered by narrow, medium and broad domains, respectively, are displayed. **b**, Distribution of the H3K4me3 domain length at the promoter regions in all pre-implantation developmental stages and cell lines. The dashed line represents the cut-off between the medium and broad domains. **c**, Heat map showing the H3K4me3 RPM on all the RefSeq transcripts of the four states of H3K4me3 domains for the morula, ICM and ES cell samples. Each row represents a promoter region with ± 10 kb around a TSS. **d**, Genome browser view of H3K4me3 broad domains establishment at *Nanog* and *Pou5f1* genes from the two-cell stage to the ICM stage. ChIP–seq RPM was smoothed by the mean of 5 pixels using the UCSC genome browser and visualized. **e**, Depletion of H3K4me3 demethylase inhibited the formation of blastocysts. The siRNA mixture of each predicated H3K4me3 demethylase was injected into MII oocytes separately or mixed together. The injected oocytes were then performed ICSI. The data are represented as the mean $\pm$ s.d. ($n = 3$). **f**, Knockdown of *Kdm5b* (si-*Kdm5b*) inhibited the formation of expanding and hatching of blastocysts. Injection of *Kdm5b* mRNA (oe-*Kdm5b*) as well as control RNA was also performed for comparison. The data are represented as the mean $\pm$ s.d. ($n = 3$). **g**, Heat map showing the breadth of H3K4me3 domains on promoters in control, oe-*Kdm5b* and si-*Kdm5b* samples. The H3K4me3 domains were classified into four groups: broad domains (dark red), medium domains (red), narrow domains (pink) and control domains (grey). The dashed lines represent the classification based on control samples. **h**, Box plot showing the averaged H3K4me3 domain breadth in different samples on promoter covered by broad (left), medium (middle) and control (right) H3K4me3 domains in control samples. **i**, Genome browser view of control, oe-*Kdm5b* and si-*Kdm5b* samples at *Zfp85* gene locus. ChIP–seq RPM was smoothed by the mean of 5 pixels using the UCSC genome browser and visualized.
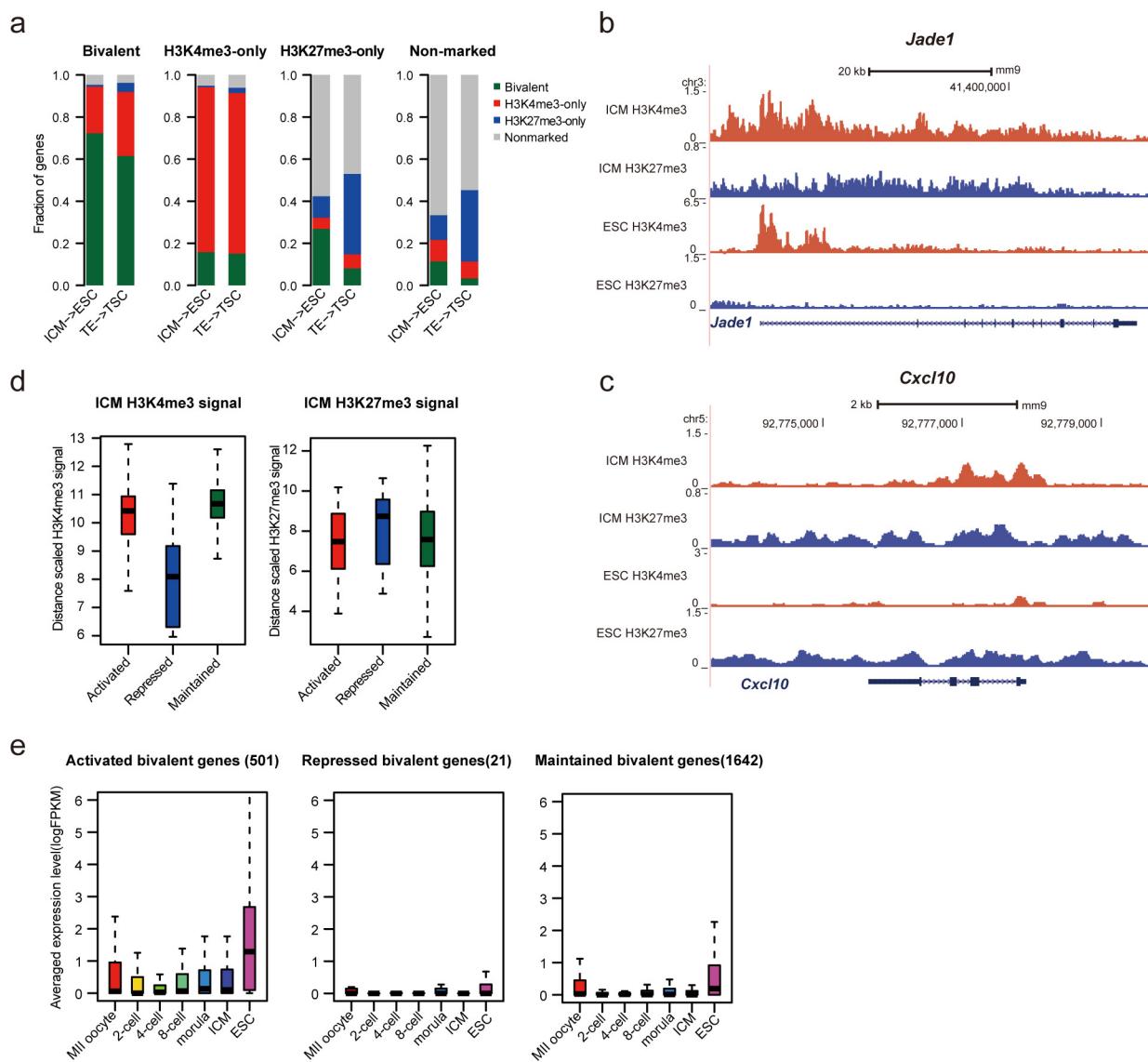
Extended Data Figure 7 | H3K4me3 breadth is highly associated with transcription activity and cell identity. a, Box plot showing the averaged expression level of the broad, medium, narrow and control H3K4me3 domains in all pre-implantation developmental stages and cell lines. The number of the transcript is listed with the group name. **b,** Heat map showing the gain and loss of the broad domains during the transition from the morula stage to the ICM and TE. The transcripts were clustered using a *k*-means function into seven groups. **c,** Averaged expression level of the transcript in the morula, ICM and TE samples for the clusters in **b**. The expression level was evaluated using the averaged log$_2$ FPKM.

d, Tendencies of the four types of H3K4me3 domains in the next stage during stable cell line derivation. Each panel represents a specific type of H3K4me3 domain. Each bar represents the types and percentages of the transcripts in the next stage. **e,** Broad domains inherited from the TE to TSCs. A gene ontology analysis was performed on the overlapped genes from the TE to TSCs, and representative genes were labelled. The *P* value was calculated using multiple Student's *t*-tests. **f, g,** Bar plot of the H3K4me3 domain length at the *Id2* (**f**) and *Elf5* (**g**) promoter regions (left panel) and the corresponding expression level (right panel) in embryo and cell lines. The dashed line represents the cut-off for broad domains.

Extended Data Figure 8 | Dynamics of the bivalent promoters in embryogenesis and differentiation. a, Heat map showing the dynamics of the bivalent genes during early mouse embryo development. Each row represents one RefSeq transcript. Red indicates a transcript that is only marked by H3K4me3, blue represents a transcript that is marked by H3K27me3, and dark green indicates a transcript in a bivalent state. **b**, Number of bivalent transcripts in MII oocytes, pre-implantation embryos and cell lines. **c**, Tendencies of the H3K4me3-only promoters,

H3K27me3-only promoters, bivalent promoters and non-marked promoters in the next stage during early mouse embryo development. Each bar represents the type and percentage of the transcript in the next stage. **d**, Bar plot showing the possibility of the H3K27me3-marked condition for transcripts with a broad, medium or narrow H3K4me3 domains. Red represents transcripts without the H3K27me3 mark, and dark green represents transcripts with the H3K27me3 mark, indicating a bivalent state.

Extended Data Figure 9 | Inheritance of the ICM-bivalent promoters in ES cell derivation. a, Tendencies of the H3K4me3-only promoters, H3K27me3-only promoters, bivalent promoters and non-marked promoters in the next stage during cell line derivation. Each bar represents the type and percentage of the transcript in the next stage. The ICM-bivalent genes, which are covered only by H3K4me3 signals, or only by H3K27me3 signals or maintain bivalent states in the ES cell samples, were defined as ES-cell-activated, ES-cell-repressed and ES-cell-maintained, respectively. **b, c**, Genome browser view of the H3K4me3 and H3K27me3 density at activated bivalent gene *Jade1* (**b**) and repressed bivalent gene *Cxcl10* (**c**) in ES cell and ICM samples. The ChIP–seq RPM was smoothed by the mean of 5 pixels using the UCSC genome browser. **d**, H3K4me3 and H3K27me3 signal intensities of the three types of ICM-bivalent transcripts, ES-cell-activated, ES-cell-repressed and ES-cell-maintained, defined in **a**. **e**, Expression level during early mouse embryo development of the three types of ICM bivalent transcripts, ES-cell-activated, ES-cell-repressed and ES-cell-maintained, defined in **a**.