

# Epigenetic regulator function through mouse gastrulation

<https://doi.org/10.1038/s41586-020-2552-x>

Received: 11 November 2018

Accepted: 6 May 2020

Published online: 29 July 2020

 Check for updates

Stefanie Grosswendt<sup>1,8</sup>, Helene Kretzmer<sup>1,8</sup>, Zachary D. Smith<sup>2,3,4,8</sup>, Abhishek Sampath Kumar<sup>1</sup>, Sara Hetzel<sup>1</sup>, Lars Wittler<sup>5</sup>, Sven Klages<sup>6</sup>, Bernd Timmermann<sup>6</sup>, Shankar Mukherji<sup>7</sup> & Alexander Meissner<sup>1,2,3,8</sup>✉

During ontogeny, proliferating cells become restricted in their fate through the combined action of cell-type-specific transcription factors and ubiquitous epigenetic machinery, which recognizes universally available histone residues or nucleotides in a context-dependent manner<sup>1,2</sup>. The molecular functions of these regulators are generally well understood, but assigning direct developmental roles to them is hampered by complex mutant phenotypes that often emerge after gastrulation<sup>3,4</sup>. Single-cell RNA sequencing and analytical approaches have explored this highly conserved, dynamic period across numerous model organisms<sup>5–8</sup>, including mouse<sup>9–18</sup>. Here we advance these strategies using a combined zygotic perturbation and single-cell RNA-sequencing platform in which many mutant mouse embryos can be assayed simultaneously, recovering robust morphological and transcriptional information across a panel of ten essential regulators. Deeper analysis of central Polycomb repressive complex (PRC) 1 and 2 components indicates substantial cooperativity, but distinguishes a dominant role for PRC2 in restricting the germline. Moreover, PRC mutant phenotypes emerge after gross epigenetic and transcriptional changes within the initial conceptus prior to gastrulation. Our experimental framework may eventually lead to a fully quantitative view of how cellular diversity emerges using an identical genetic template and from a single totipotent cell.

Gastrulation represents a process of embryogenesis that begins with the induction of the primitive streak and proceeds through the generation of distinct germ layers and initial body axes<sup>4</sup>. To comprehensively assess this period in mouse development, we generated single-cell RNA sequencing (scRNA-seq) data from pools of sibling embryos with a B6/CAST F<sub>1</sub> father, allowing us to computationally distinguish single replicates by their randomly inherited CAST genotype, and sex according to the expression of chromosome X- and Y-linked genes (Extended Data Fig. 1a–c, Supplementary Tables 1, 2). We sampled 9–11 wild-type embryos per time point, beginning with the pluripotent epiblast and proceeding to early organogenesis (embryonic day (E) 6.5 to E8.5) (Fig. 1a). In total, our wild-type compendium comprises 88,779 high-quality single-cell transcriptomes from 50 embryos (median of 16,898 transcripts and 3,854 genes per cell and approximately 2–49% of each embryo, depending on developmental stage) (Extended Data Fig. 1d, Supplementary Table 1).

To build a reference of transcriptional states, we iteratively clustered our data across replicates and time points. We also adjusted the number of informative ‘marker’ genes per state, leading to a set of 712 that reliably assigns individual cells to one of 42 reproducible states (Fig. 1b, Extended Data Fig. 1e–g). We then assembled a complete lineage tree using state emergence time and gene expression, with additional support from RNA velocity analysis to indicate transcriptome dynamics

(Fig. 1c, Extended Data Fig. 1h, i, Supplementary Tables 3, 4). In our tree, all embryonic lineages stem from the pluripotent epiblast, which gives rise to early ectoderm followed by neural and non-neural sub-lineages, as well as to products of the primitive streak, including extraembryonic and embryonic mesoderm, neuromesodermal progenitors, the embryonic endoderm, and primordial germ cells (PGCs). We provide a detailed and referenced explanation for our nomenclature and tree placement in the Methods and Supplementary Note.

## Disrupting epigenetic regulators

With our wild-type reference established, we proceeded to disrupt one of several epigenetic regulators zygotically, prioritizing key enzymes with known viability issues during early and mid-gestation<sup>3</sup> (Extended Data Fig. 2a). We included the three major DNA methyltransferases, the maintenance enzyme DNMT1 and the de novo enzymes DNMT3A and DNMT3B, as well as the repressive histone 3 lysine 9 (H3K9) methyltransferase G9A. Mutations in DNMT1, G9A and DNMT3B are lethal after gastrulation, whereas DNMT3A mutants die postnatally with noted neuronal abnormalities<sup>19</sup>. We also selected both canonical and noncanonical Polycomb complex subunits, which repress developmental genes in a temporal and cell-type-specific manner: RNF2 (also known as RING1B) and EED are essential to PRC1 and PRC2, whereas

<sup>1</sup>Department of Genome Regulation, Max Planck Institute for Molecular Genetics, Berlin, Germany. <sup>2</sup>Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>3</sup>Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. <sup>4</sup>Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, USA. <sup>5</sup>Department of Developmental Genetics, Max Planck Institute for Molecular Genetics, Berlin, Germany. <sup>6</sup>Sequencing Core Facility, Max Planck Institute for Molecular Genetics, Berlin, Germany. <sup>7</sup>Department of Physics, Washington University in St Louis, St Louis, MO, USA. <sup>8</sup>These authors contributed equally: Stefanie Grosswendt, Helene Kretzmer, Zachary D. Smith. ✉e-mail: meissner@molgen.mpg.de

KDM2B and L3MBTL2 are non-canonical PRC1.1 and PRC1.6 complex subunits, respectively<sup>20</sup>. Finally, we included the H3K4 methyltransferases KMT2A and KMT2B, Trithorax group orthologues that promote developmental gene expression in opposition to Polycomb<sup>21</sup>.

Unlike many transcription factors, these genes are expressed across most lineages and cell states, although the *de novo* DNMT enzymes are particularly abundant before lineage commitment (Extended Data Fig. 2b, c). To disrupt these genes, we injected B6/CAST-fertilized zygotes with the endonuclease Cas9 and 3–4 single-guide RNAs (sgRNAs) targeted to exons common to all isoforms, transferred E3.5 embryos into pseudopregnant females and recovered scRNA-seq data for 8–12 E8.5 embryos comprising 7,548–25,408 cells (Extended Data Fig. 2a). We confirmed gene disruption by inspecting read alignments over their respective target sites (Extended Data Fig. 2d).

### Detecting emerging morphological defects

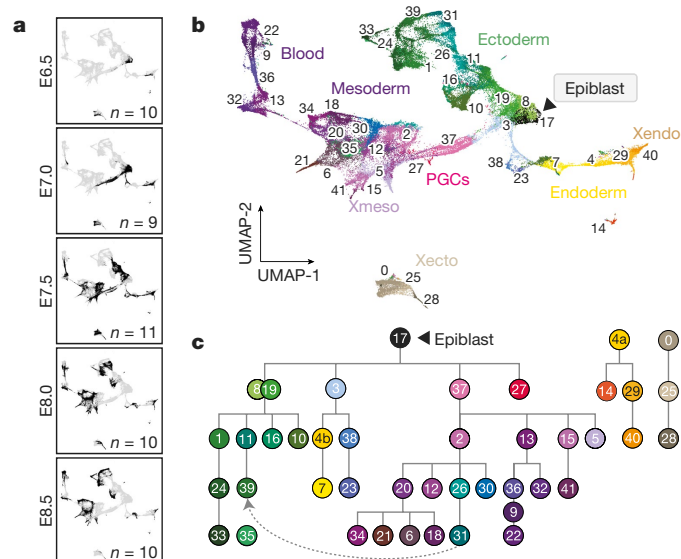
Our data allow us to explore mutant embryos both anatomically (developmental progression and morphology) and molecularly (transcriptional state). First, we assigned mutant cells by marker expression to their closest wild-type state and examined embryo composition across replicates (Extended Data Fig. 2e, f). Certain regulator mutants clearly influence the number and kinds of states produced, but generally do not perturb their gross transcriptional identity (Fig. 2a, Extended Data Fig. 3a). Instead, most mutants appear to occupy earlier cell states of the overall lineage hierarchy, suggesting developmental delays. We therefore developed a stage-matching metric that considers and weighs the types of states and their relative proportions within an embryo compared to our wild-type data (Extended Data Fig. 3b–d, Supplementary Table 5).

Our approach confirmed many historical observations, including increased severity for DNMT1 mutations compared with DNMT3A or DNMT3B, as well as of KMT2B compared with KMT2A, consistent with its primary role in orchestrating early differentiation<sup>22</sup>. Of our E8.5-isolated embryos, L3MBTL2 showed the greatest delay, primarily arresting after the onset of gastrulation at around E7.0, followed by EED and RNF2, which gastrulate but largely fail to progress beyond around E7.5 (Fig. 2b). Notably, we observe developmental defects beyond progression or growth. For example, EED and RNF2 clearly gastrulate, but are unable to produce neural ectoderm and bias the primitive streak towards posterior lineages such as the extraembryonic mesoderm and PGCs (Fig. 2a, Extended Data Fig. 3a). By contrast, L3MBTL2-mutant embryos form some tissues of the early primitive streak that do not mature, but continue to produce abundant extraembryonic tissues (Extended Data Fig. 3a).

### PRC1 and PRC2 converge to a common gene set

Assigning each cell from our mutant embryos to pre-defined wild-type states allowed us to measure within-state expression changes in addition to over- or underproduction of certain lineages. To compare our ten mutant cohorts, we initially identified differentially expressed genes for each mutant cell state against wild-type embryos as ‘up’ or ‘down’ regulated (Supplementary Table 6). We then calculated the fraction of cell states in which each gene is deregulated, doing so separately for the embryonic and extraembryonic lineages because of their independent origins and distinct epigenetic regulation<sup>23,24</sup>. As expected, L3MBTL2 remains a global outgroup, with the largest number of recurrently deregulated genes, though this may be affected by lower overall embryonic complexity. We were surprised to see that the non-canonical PRC1 subunit KDM2B clusters with EED and RNF2, even though KDM2B-mutant embryos generally produce more mature cell types (Fig. 2c, Extended Data Fig. 3a, e).

All three Polycomb-associated mutants also converge towards functional ontologies associated with developmental processes and cell cycle regulation (Extended Data Fig. 3e). For example, the tumour



**Fig. 1 | Single-cell profiling of early post-implantation development.**

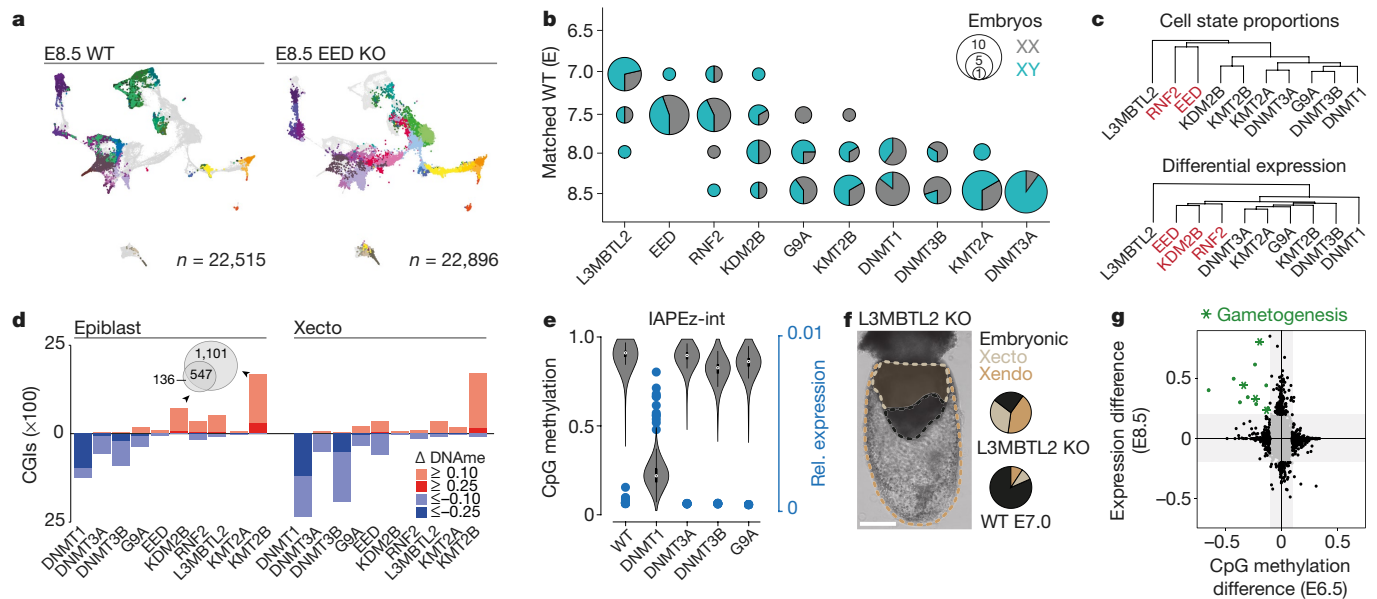
**a**, Uniform manifold approximation and projection (UMAP) plots of 88,779 wild-type single-cell transcriptomes (black dots), separated by time point. *n* denotes replicate embryos. **b**, UMAP of our wild-type scRNA-seq time series from **a**. Numbers denote the 42 cell states, colours indicate states of the same major lineages. **c**, Curated lineage tree of cell states. States were annotated and connected according to their emergence, marker gene expression, the literature and scRNA velocity (Extended Data Fig. 1h, i, Methods, Supplementary Information). Dashed arrow represents neuromesodermal progenitor (NMP)-containing states that contribute to neural ectoderm<sup>39</sup>. Extraembryonic ectoderm and endoderm (Xecto and Xendo, respectively) are disconnected to reflect their preimplantation origins<sup>23</sup>. 0: extraembryonic ectoderm 1; 1: neural ectoderm anterior; 2: primitive streak late; 3: streak pre-specified/anterior; 4a and 4b: endoderm primitive and definitive, respectively; 5: allantois; 6: secondary heart field/splanchnic lateral plate; 7: gut; 8: ectoderm early 1; 9: primitive blood early; 10: preplacodal ectoderm; 11: neural ectoderm posterior; 12: posterior lateral plate mesoderm; 13: haematopoietic/endothelial progenitor; 14: parietal endoderm; 15: amnion mesoderm early; 16: surface ectoderm; 17: epiblast; 18: somites; 19: ectoderm early 2; 20: splanchnic-lateral/anterior-paraxial mesoderm; 21: primitive heart tube; 22: primitive blood late; 23: notochord; 24: fore/midbrain; 25: extraembryonic ectoderm 2; 26: NMPs early; 27: primordial germ cells (PGCs); 28: differentiated trophoblasts; 29: visceral endoderm early; 30: presomitic mesoderm; 31: NMPs late; 32: angioblasts; 33: neural crest; 34: pharyngeal arch mesoderm; 35: similar to neural crest; 36: primitive blood progenitor; 37: primitive streak early; 38: node; 39: future spinal cord; 40: visceral endoderm late; 41: amnion mesoderm late. Note, ‘state 35: similar to neural crest’ is not enriched for specific markers but most closely resembles ‘33: neural crest’. It is disconnected from the tree to reflect this ambiguity.

suppressor *Cdkn2a*, a known PRC-regulated locus<sup>25</sup>, is constitutively targeted by both PRC1 and PRC2. By contrast, L3MBTL2-mutant embryos show limited overlap with EED, RNF2 or KDM2B, supporting a predominantly PRC1-independent role for L3MBTL2 in early development that cannot be compensated for by PRC2 or other PRC1 complexes (Supplementary Table 7). In general, our PRC-associated mutant data provide the most compelling results in terms of transcriptional and morphological defects to the gastrulation process itself. However, DNMT1- and G9A-mutant embryos also exhibit functional ontologies associated with loss of imprinting and other previously described targets that will warrant further investigation (Supplementary Tables 6, 7).

### Epigenetic deregulation at E6.5

To provide greater clarity into the onset of epigenetic disruption as it may relate to phenotype, we generated whole-genome bisulfite





**Fig. 2 | Morphological and molecular consequences of epigenetic regulator mutation.** **a**, Example of single-cell data for one of our epigenetic regulator mutants at E8.5. Mutant (KO) cells were assigned to wild-type (WT) cell states (colours) and projected onto our gastrulation UMAP (see Extended Data Fig. 3a for all regulator mutants). **b**, Developmental staging according to cell-state composition. Circle size denotes embryo number assigned to a given wild-type stage (y axis). Colours indicate sex. **c**, Hierarchical clustering of mutant embryos based on composition (top) and transcriptional deregulation (bottom) of cell states compared to matching wild-type stages. **d**, CGI methylation across our mutants at E6.5. Hyper- or hypomethylation in comparison to wild-type embryos for absolute changes  $\geq 0.1$  and  $\geq 0.25$ . DNMT1 shows the greatest loss in the epiblast, and both DNMT1 and DNMT3B show large effects in Xecto. KMT2B and KDM2B show substantial and overlapping CGI methylation in epiblast (Venn diagram), whereas Xecto is only affected by KMT2B. **e**, DNA methylation (violin plots) of IAPs in E6.5 epiblast and per

sequencing (WGBS) data of E6.5 epiblast and extraembryonic ectoderm (Xecto) for each regulator mutant, as these tissues represent the latest homogeneous progenitors before the actions of gastrulation (Extended Data Fig. 4). We see clear global DNA methylation differences for the DNMT1 mutant, as well as more subtly for DNMT3B and DNMT3A (Extended Data Fig. 4a–d).

We examined changes at CpG islands (CGIs), which represent a major focal point of Polycomb and Trithorax group-based regulation and are usually free of DNA methylation in the epiblast. We observe notable CGI methylation in response to KDM2B and KMT2B mutation, which both proceed through gastrulation with some developmental delay, but not for KMT2A, which advances normally up to E8.5<sup>26</sup> (Fig. 2d, Extended Data Fig. 4d). Functionally, KMT2B appears to protect a larger number of CGIs than KDM2B and also operates within Xecto, which suggests either a broader utility or earlier preimplantation activity (Fig. 2d). Using publicly available data, we find that KDM2B and KMT2B protected promoter CGIs are also approximately 2.5-fold enriched for H3K27me3-based regulation in the epiblast. However, in general these CGI-associated genes are lowly expressed in our data, suggesting that the influence of both regulator mutants may be too subtle to pinpoint with the current scRNA-seq strategy (Extended Data Fig. 4e). By contrast, core PRC subunits EED and RNF2 do not appear to be required for protection against CGI methylation in the epiblast, but do influence the methylation status of surrounding regions (Extended Data Fig. 4f).

We also applied our scRNA-seq and WGBS data to explore retrotransposons within our DNMT or G9A mutants, which otherwise exhibited limited gene expression differences (Extended Data Fig. 5a). Here, the

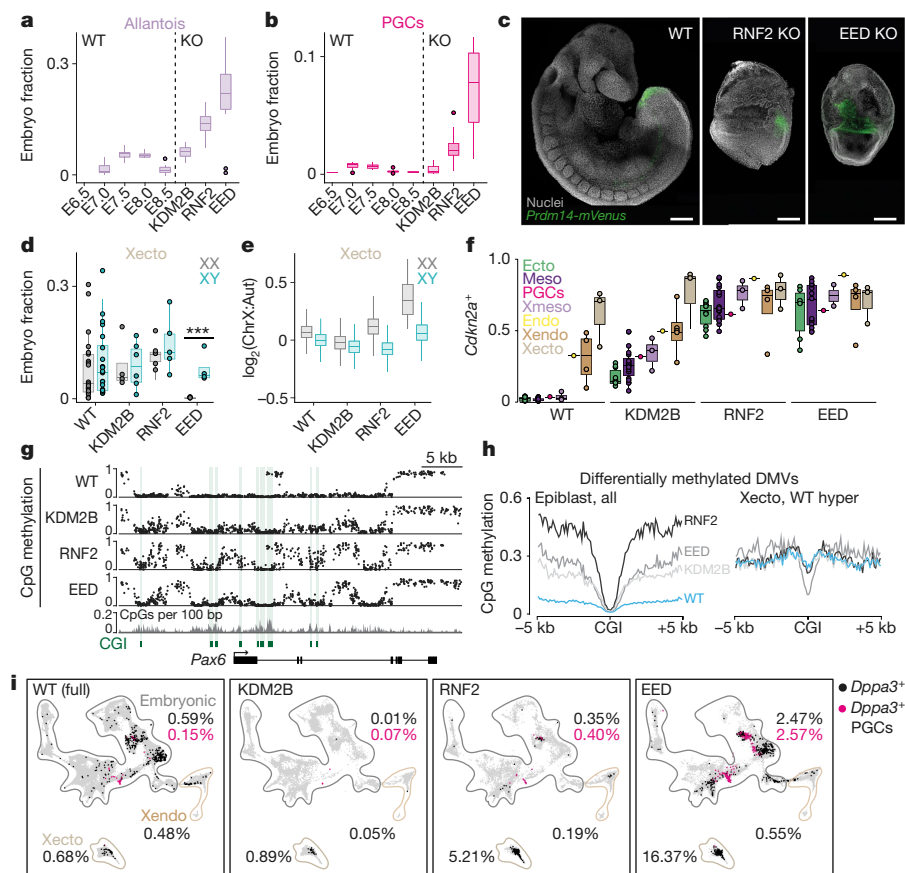
embryo IAP expression (blue dots) as a fraction of reads in E8.5 embryonic lineages. White dots denote median; edges denote the interquartile range (IQR); and whiskers denote  $1.5 \times \text{IQR}$ . **f**, Representative E8.5-isolated L3MBTL2-mutant embryo (of 10 total collected for scRNA-seq, injections were replicated three independent times with similar morphological results). Dashed lines demarcate lineages and pie charts show median proportions compared to stage-matched wild-type embryos as calculated by scRNA-seq. Xecto and Xendo are overabundant, while embryonic lineages are substantially impeded. Scale bar, 200  $\mu\text{m}$ . **g**, Scatterplot of changes in E6.5 epiblast promoter methylation (x axis) and E8.5 embryonic expression (y axis) for L3MBTL2 compared with wild-type embryos. Green indicates genes that lose promoter methylation ( $\geq 0.1$ ) and increase expression ( $\geq 0.2$  fraction of positive cells). Asterisks denote genes that function in gametogenesis.  $N = 12$  gene promoters.

ERV subfamily of long terminal repeats (LTRs) shows strongly coupled demethylation and transcription within DNMT1 mutants, specifically methylation-sensitive intracisternal A-type particles (IAPs)<sup>27</sup> (Fig. 2e, Extended Data Fig. 5b). This may explain the impeded progression and death of DNMT1-mutant embryos within around 1–2 days after E8.5. By contrast, continued IAP repression in DNMT3B or G9A mutants suggests that these embryos maintain a sufficient threshold to preserve epigenetic silencing (Fig. 2e).

Finally, we re-examined our L3MBTL2 mutants to better understand their severe phenotype, including near total embryonic arrest and continued extraembryonic growth (Fig. 2f). We analysed cells from either the embryonic or Xecto lineage separately for expression changes as they relate to promoter DNA methylation. We identify 12 genes that are both highly overexpressed and show lower than expected promoter methylation (Fig. 2g, Extended Data Fig. 5c, d), including several previously reported non-canonical PRC1.6 targets associated with gametogenesis<sup>28</sup>. Examined over early wild-type development, we find that L3MBTL2-sensitive genes are specifically unmethylated in both gametes and become methylated following implantation (Extended Data Fig. 5e). Thus, L3MBTL2 functions to actively silence select gamete-specific genes, the aberrant and exogenous expression of which may otherwise be detrimental, particularly within the epiblast.

## PRC2 dominates early lineage restriction

We next compared EED, RNF2 and KDM2B mutant phenotypes, which converge to a similar gene set but differ morphologically. Both EED



**Fig. 3 | Phenotypic and molecular abnormalities of PRC regulator mutants.**

**a, b**, Fraction of cells assigned to the allantois (**a**) and PGC (**b**) state per embryo. Dots denote outliers;  $n = 10; 9; 11; 10; 10; 11$  and  $10$  embryos (left to right). **c**, *Prdm14* reporter<sup>31</sup> activity in representative E8.5-isolated wild-type, RNF2 and EED mutant embryos (from total of 4, 7 and 5 embryos, respectively). Scale bars, 200  $\mu$ m. **d**, Per embryo fractions assigned to Xecto cell states, separated by sex. \*\*\* $P \leq 0.001$ , two-sided Wilcoxon test;  $n = 25; 25; 4; 6; 6; 5$  and  $5$  embryos and  $P$  value = 0.1276; 0.9118; 0.7618 and 0.0001 (left to right). **e**, Ratio of chromosome X (ChrX) to autosome (Aut) transcripts per Xecto cell, separated by sex. Outliers omitted;  $n = 1,769; 3,685; 755; 1,372; 1,465; 1,220; 19$  and  $1,594$  cells. **f**, Box plots of the PRC target *Cdkn2a*, shown as the fraction of transcript positive cells for each state (dots) grouped by lineage (colours). In wild-type embryos, expression is limited to Xecto and Xendo, with mixed signal in endoderm (Endo) potentially reflecting its heterogeneous origins<sup>16</sup>.  $n = 10; 19$ ;

1; 3; 3 and 4 cell states. For all box plots, lines denote median values, edges denote quantiles, and whiskers denote  $1.5 \times \text{IQR}$ . **g**, Developmental gene-associated DMVs gain methylation in PRC mutants in E6.5 epiblast. CpG resolution genome browser tracks of WGBS data for the neural ectodermal regulator *Pax6*. CGIs and CpG density are provided. **h**, Median CpG methylation centred on CGIs within differentially methylated DMVs for E6.5 epiblast (left) and Xecto (right). CGIs that are normally methylated in the Xecto (wild type) do not acquire de novo methylation in EED (Supplementary Table 8). **i**, *Dppa3*-transcript positive cells over our wild-type time series (E6.5–E8.5) and in E8.5-isolated PRC mutants, with the PGC-assigned subset highlighted in pink. Percentages are indicated per major lineage. In the embryo proper, black and pink dots sum to the total fraction of *Dppa3*<sup>+</sup> cells. Embryonic/Xecto/Xendo cells: WT  $n = 77,298/5,454/6,027$ , KDM2B  $n = 14,624/2,127/2,192$ , RNF2  $n = 9,696/2,685/3,208$ , EED  $n = 18,723/1,613/2,560$ .

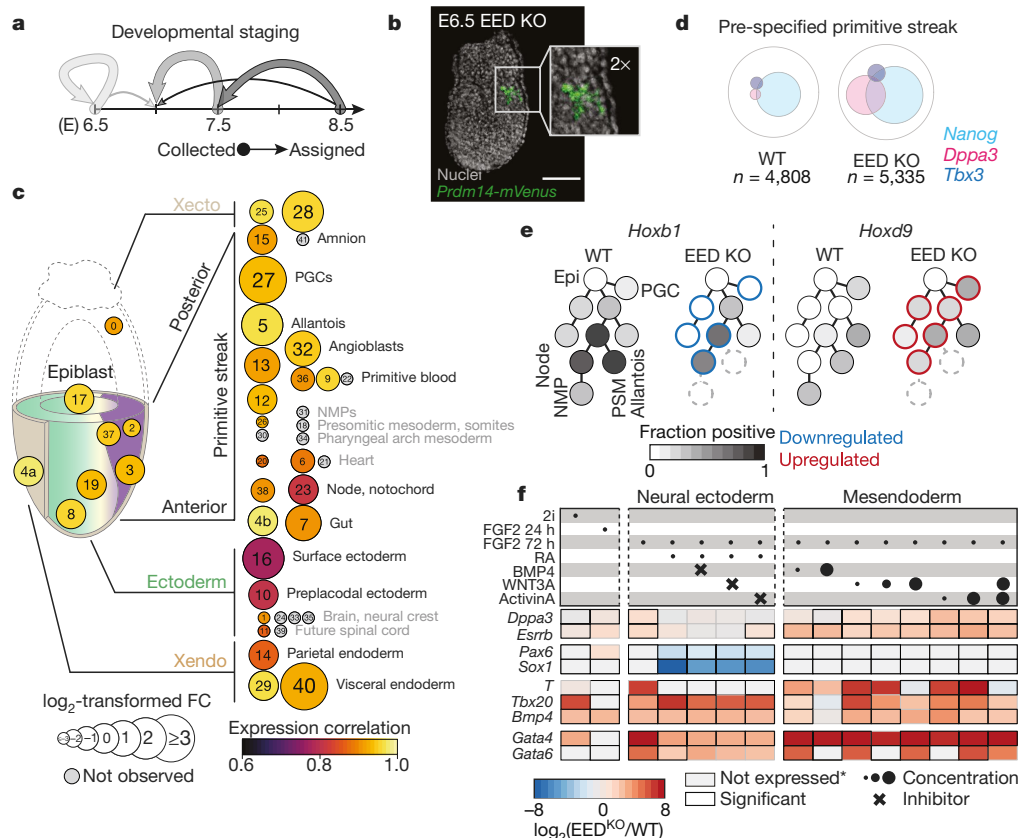
and RNF2 overproduce posterior-proximal structures, such as the allantois, without advancing the embryo proper<sup>29,30</sup> (Fig. 3a). Notably, EED mutants also have substantially more PGC state cells, a result we confirmed by generating embryos carrying a reporter driven by the promoter of *Prdm14*, a master regulator for this lineage<sup>31</sup> (Fig. 3b, c, Extended Data Figs. 3a, 6a).

PRC complexes also contribute to imprinted X chromosome inactivation (XCI) in extraembryonic and random XCI in embryonic cells<sup>32,33</sup>. Separated by sex, we find that EED-mutant females consistently fail to maintain the Xecto lineage and derepress chromosome X-linked genes, whereas RNF2 more subtly deviates and KDM2B appears normal (Fig. 3d, e). Notably, extraembryonic endoderm (Xendo) also undergoes imprinted XCI but does not exhibit either of these phenotypes. Furthermore, we observe largely normal ratios of chromosome X transcripts within embryonic lineages, although PRC2 participates in random XCI as well (Extended Data Fig. 6b).

In contrast to core PRC1 and PRC2 components, KDM2B largely appears to have a secondary role on the same overall gene set. For example, *Cdkn2a* is upregulated in all embryonic lineages for all three

mutants, but to a substantially higher degree in EED and RNF2 (Fig. 3f). Despite a canonical role as a tumour suppressor, expression of *Cdkn2a* does not explain the overall gastrulation defect observed in EED-mutant embryos, as co-injecting sgRNAs to *Cdkn2a* and *Eed* does not produce appreciable differences compared to targeting *Eed* alone (Extended Data Fig. 6c–e).

In our three PRC regulator mutants, differentially methylated CpGs collect within multi-kilobase territories, termed DNA methylation valleys (DMVs), that are normally maintained in an unmethylated state<sup>34</sup>. After PRC disruption, a subset of DMVs becomes hypermethylated, although internal CGIs remain protected (Fig. 3g, h, Extended Data Fig. 7a). These PRC-sensitive DMVs are enriched for H3K27me3 and cover larger genetic territories than those that show no change in DNA methylation. They also preferentially contain promoters of lineage-specific marker genes, although we see no straightforward correlation between epigenetically deregulated genes and morphology (Extended Data Fig. 7b, c, Supplementary Table 8). We see the same overall change in methylation pattern in EED-, RNF2- and KDM2B-mutant epiblast, but with lower net values for KDM2B (Fig. 3g, h,



**Fig. 4 | The EED mutant phenotype extends from disrupted pluripotency exit.** **a**, Developmental stage assignment of our EED mutant time series as in Fig. 2b. Arrow strength corresponds to the fraction of embryos matching a wild-type reference stage. Subsequent analyses compare EED-mutant embryos to their closest wild-type stage. **b**, Representative E6.5 EED-mutant embryo (out of four collected and analysed from one experiment) showing *Prdm14-mVenus* signal within the posterior–proximal epiblast. Scale bar, 100  $\mu$ m. **c**, Composition and expression changes in EED-mutant embryos. The mutant state proportions were compared to matched wild-type stages (circle sizes). Gene expression correlation (purple to yellow) was determined using our defined marker genes. States are organized according to lineage. The earliest states (within conceptus) are less affected compared to later states. Many later states are not observed (grey), but whether some could be produced before lethality remains unclear. Proportion changes for outermost tissues

(Xendo and Xecto) may be sensitive to technical variability during isolation. State annotations as in Fig. 1c. FC, fold change. **d**, Venn diagram of pre-specified and anterior primitive streak (state 3) cells in wild-type and EED-mutant embryos that express key transcription factors with shared functions in pluripotency and the germline<sup>3</sup>. **e**, Select mesodermal lineages and PGCs as they stem from the pluripotent epiblast. Grey-scale indicates fraction of *Hoxb1*<sup>+</sup> and *Hoxd9*<sup>+</sup> cells. EED-mutant embryos induce *Hoxd9* prematurely, leading to a profile that resembles posterior lineages. Differential up- or downregulation is highlighted in red or blue, respectively. States that are not produced are dashed. **f**, The log<sub>2</sub>-transformed fold change between EED-knockout (EED<sup>KO</sup>) and wild-type mES cells for select genes, taken from a total of 44 profiled using Nanostring (*n* = 3 experimental replicates) (Extended Data Fig. 11). Top rows: circles, morphogen concentrations; crosses, inhibitors. Grey boxes denote expression below threshold in both samples. RA, retinoic acid.

Extended Data Fig. 8a, b). Both epigenetically and transcriptionally, KDM2B shows some deregulation consistent with core PRC subunits, but does not seem to pass a critical threshold to substantially alter early embryo patterning.

More detailed analysis of the extraembryonic ectoderm also supports the heightened severity of the EED mutant phenotype. In wild type, DMVs are preferentially unmethylated within the epiblast, but are de novo methylated within the Xecto lineage, including embedded CGIs (Extended Data Fig. 7a, b). However, these CGIs do not gain methylation specifically within the EED-mutant Xecto, including promoters of crucial early regulators of the epiblast and germline, such as OTX2 and PRDM14, respectively (Fig. 3h, Extended Data Fig. 8b, c). Disruption of EED therefore affects developmental gene promoters in both embryonic and extraembryonic compartments, producing an overall similar epigenetic pattern (methylated except at CGIs). *Dppa3*, another key PGC marker, is more broadly expressed within EED-mutant embryos and particularly abundant within the Xecto lineage (Fig. 3i). These data further support an early, PRC2-dominant role in restricting germline-relevant genes that extends into the trophoctoderm.

## The PRC2 phenotype precedes gastrulation

We generated additional EED-mutant scRNA-seq data from E6.5 and E7.5 to determine whether these phenotypes otherwise obey general principles of wild-type development, including stepwise induction of committed progenitors by exogenous signals (Extended Data Figs. 9, 10). Morphologically, we find minimal changes at E6.5, but confirm diminished embryo complexity and delays by E7.5 (Fig. 4a, Extended Data Fig. 10b, c). Notably, *Prdm14* reporter activity indicates that PGCs remain positionally specified in EED mutants, with signal limited to a few cells within the posterior–proximal epiblast at E6.5 (Fig. 4b). More generally, the relative proportions and transcriptional stability of early extraembryonic and embryonic products resemble wild-type embryos, but subsequently become either abnormal or inconsistent (Fig. 4c). By E7.5, products of the primitive streak skew posteriorly towards extraembryonic mesoderm and PGCs, whereas the axial mesoderm (node, notochord) is highly abnormal and more advanced embryonic mesoderm or neural ectoderm does not develop (Fig. 4c, Extended Data Fig. 10c, d).

We sought to identify when transcriptional biases first emerge that may determine the ultimate partitioning of EED-mutant embryos.



Many changes detected at E8.5 are already apparent before gastrulation: *Cdkn2a* is already induced within the embryo proper and chromosome X is aberrantly transcribed within the female Xecto lineage (Extended Data Fig. 10e, f). Moreover, PGC-associated marker genes tend to be abundant and co-expressed within the same cells of several early states, including the epiblast and pre-specified primitive streak (Fig. 4d, Extended Data Fig. 10g). These genes also function during naive pluripotency and are generally silenced during implantation<sup>35</sup>. Furthermore, we observe failed stepwise induction of homeotic (*Hox*) genes throughout the primitive streak, which generally matures from a HOXB1-positive state into HOXB1 and HOXD9 double-positive caudal mesodermal tissues<sup>36</sup>. By contrast, EED-mutant embryos express *Hoxd9* prematurely and destabilize *Hoxb1* induction, mirroring the eventual posteriorized phenotype (Fig. 4e, Supplementary Table 9).

Currently our pipeline cannot easily address the influence of non-autonomous factors. For example, deregulation of extraembryonic tissues may alter the initial morphogen gradients that set the primitive streak, which could lead to underdevelopment or promote biases. To examine how PRC2 interacts with these parameters, we generated a knockout mouse embryonic stem-cell line (EED-knockout mES cells) to induce alternative fates in vitro (Extended Data Fig. 11a, b, Supplementary Fig. 1). Specifically, we directed EED-knockout mES cells into a formative epiblast-like state using FGF, followed by exposure to different concentrations of signalling components for an additional 48 h. Across many conditions, EED-knockout mES cells less reliably silenced pluripotency-associated genes, such as *Dppa3* and *Esrrb*, and broadly expressed posterior-proximal mesodermal genes, such as *Bmp4* and *Bmp8b*, which also support PGC production in vivo<sup>37</sup>. We were unable to induce neural ectodermal genes, even when using small molecule inhibitors to impede competing mesendodermal or surface ectodermal pathways (Fig. 4f, Extended Data Fig. 11c, d). Thus, the EED mutant phenotype seems to reflect a failure to adequately demarcate exit from pluripotency with the independent and exogenous priming of neural ectodermal and mesendodermal lineages.

## Discussion

We present a combined genetic perturbation, scRNA-seq strategy to functionally dissect mammalian embryogenesis. Our platform is designed to understand complex mutant phenotypes comprehensively, both anatomically and molecularly, and to account for natural variation that may be fundamental to a given developmental process. We investigated several key epigenetic regulators that produce lethal phenotypes after gastrulation but have been difficult to characterize in full because they are presumed to buffer differentiation across many contexts. Using this approach, we confirm that core PRCs largely function cooperatively to counteract an otherwise innate mesodermal bias and safeguard neural regulator induction within the early ectoderm. However, PRC2 mutants exhibit greater severity that includes the overproduction of PGCs, broad destabilization of a shared PGC and pluripotency subnetwork, and failure to establish several key epigenetic features within the extraembryonic ectoderm. Further work is needed to fully resolve the interactions between these and other regulators as they coordinate morphogenesis.

We believe that our approach is highly tractable and may be expanded to address these and other questions, including the simultaneous disruption of multiple genes to explore epistasis or redundancy and conditional strategies to infer temporal, lineage, maternal or non-autonomous effects<sup>38</sup>. Integrating molecular lineage recording will contribute additional information by tracking progenitor field dynamics in the absence of epigenetic supervision<sup>16</sup>. Finally, the ability to measure several parameters across replicates will provide insight into the robustness of developmental encoding: how these indeterminate processes reliably reproduce an identical body plan. Cumulatively, these strategies may ultimately yield a complete description of the

interactions between genetic and epigenetic mechanisms that govern ontogeny.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2552-x>.

- Hemberger, M., Dean, W. & Reik, W. Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. *Nat. Rev. Cell Biol.* **10**, 526–537 (2009).
- Meissner, A. Epigenetic modifications in pluripotent and differentiated cells. *Nat. Biotechnol.* **28**, 1079–1088 (2010).
- Surani, M. A., Hayashi, K. & Hajkova, P. Genetic and epigenetic regulators of pluripotency. *Cell* **128**, 747–762 (2007).
- Rivera-Pérez, J. A. & Hadjantonakis, A.-K. The dynamics of morphogenesis in the early mouse embryo. *Cold Spring Harb. Perspect. Biol.* **7**, a015867 (2014).
- Plass, M. et al. Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science* **360**, eaq1723 (2018).
- Briggs, J. A. et al. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* **360**, eaar5780 (2018).
- Wagner, D. E. et al. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018).
- Farrell, J. A. et al. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, eaar3131 (2018).
- Scialdone, A. et al. Resolving early mesoderm diversification through single-cell expression profiling. *Nature* **535**, 289–293 (2016).
- Peng, G. et al. Spatial transcriptome for the molecular annotation of lineage fates and cell identity in mid-gastrula mouse embryo. *Dev. Cell* **36**, 681–697 (2016).
- Mohammed, H. et al. Single-cell landscape of transcriptional heterogeneity and cell fate decisions during mouse early gastrulation. *Cell Rep.* **20**, 1215–1228 (2017).
- Ibarra-Soria, X. et al. Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation. *Nat. Cell Biol.* **20**, 127–134 (2018).
- Peng, G. et al. Molecular architecture of lineage allocation and tissue organization in early mouse embryo. *Nature* **572**, 528–532 (2019).
- Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).
- Nowotschin, S. et al. The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* **569**, 361–367 (2019).
- Chan, M. M. et al. Molecular recording of mammalian embryogenesis. *Nature* **570**, 77–82 (2019).
- Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
- Argelaguet, R. et al. Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* **576**, 487–491 (2019).
- Nguyen, S., Meletis, K., Fu, D., Jhaveri, S. & Jaenisch, R. Ablation of de novo DNA methyltransferase Dnmt3a in the nervous system leads to neuromuscular defects and shortened lifespan. *Dev. Dyn.* **236**, 1663–1676 (2007).
- Laugesen, A. & Helin, K. Chromatin repressive complexes in stem cells, development, and cancer. *Cell Stem Cell* **14**, 735–751 (2014).
- Piunti, A. & Shilatifard, A. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science* **352**, aad9780 (2016).
- Glaser, S. et al. Multiple epigenetic maintenance factors implicated by the loss of ML2 in mouse development. *Development* **133**, 1423–1432 (2006).
- Rossant, J., Chazaud, C. & Yamanaka, Y. Lineage allocation and asymmetries in the early mouse embryo. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **358**, 1341–1349 (2003).
- Smith, Z. D. et al. Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature* **549**, 5434–547 (2017).
- Gil, J. & Peters, G. Regulation of the INK4b-ARF-INK4a tumour suppressor locus: all for one or one for all. *Nat. Rev. Mol. Cell Biol.* **7**, 667–677 (2006).
- Boulard, M., Edwards, J. R. & Bestor, T. H. FBXL10 protects Polycomb-bound genes from hypermethylation. *Nat. Genet.* **47**, 479–485 (2015).
- Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* **20**, 116–117 (1998).
- Qin, J. et al. The polycomb group protein L3mbtl2 assembles an atypical PRC1-family complex that is essential in pluripotent stem cells and early development. *Cell Stem Cell* **11**, 319–332 (2012).
- Faust, C., Schumacher, A., Holdener, B. & Magnuson, T. The eed mutation disrupts anterior mesoderm production in mice. *Development* **121**, 273–285 (1995).
- Voncken, J. W. et al. Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc. Natl Acad. Sci. USA* **100**, 2468–2473 (2003).
- Yamaji, M. et al. Critical function of *Prdm14* for the establishment of the germ cell lineage in mice. *Nat. Genet.* **40**, 1016–1022 (2008).
- Żylicz, J. J. et al. The Implication of early chromatin changes in X chromosome inactivation. *Cell* **176**, 182–197.e23 (2019).
- Wang, J. et al. Imprinted X inactivation maintained by a mouse Polycomb group gene. *Nat. Genet.* **28**, 371–375 (2001).
- Li, Y. et al. Genome-wide analyses reveal a role of Polycomb in promoting hypomethylation of DNA methylation valleys. *Genome Biol.* **19**, 18–16 (2018).

- 
35. Leitch, H. G. & Smith, A. The mammalian germline as a pluripotency cycle. *Development* **140**, 2495–2501 (2013).
36. Forlani, S., Lawson, K. A. & Deschamps, J. Acquisition of Hox codes during gastrulation and axial elongation in the mouse embryo. *Development* **130**, 3807–3819 (2003).
37. Saitou, M. Specification of the germ cell lineage in mice. *Front. Biosci.* **14**, 1068–1087 (2009).
38. Nicetto, D. et al. H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. *Science* **363**, 294–297 (2019).
39. Tzouanacou, E., Wegener, A., Wymeersch, F. J., Wilson, V. & Nicolas, J.-F. Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis. *Dev. Cell* **17**, 365–376 (2009).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2020

## Methods

### Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

### Embryo generation

Protocols are adapted from those previously described<sup>40</sup>. In brief, B6D2F1 strain female mice (age 6–8 weeks, Jackson Labs) were superovulated by serial injection of Pregnant Mare Serum Gonadotropin (5 IU per mouse, Prospec Protein Specialists) followed by human chorionic gonadotropin (5 IU, Millipore) 46 h later. Then 12–14 h after priming, MII stage oocytes were isolated in M2 media supplemented with hyaluronidase (Millipore) and stored in 25  $\mu$ l drops of pre-gassed KSOM with half-strength concentration of amino acids (Millipore) under mineral oil (Irvine Scientific). Zygotes were generated by piezo-actuated intracytoplasmic sperm injection (ICSI) as previously described<sup>41</sup> using thawed B6/CAST F<sub>1</sub> strain sperm in batches of 30–50 oocytes and standard micromanipulation equipment, including a Hamilton Thorne XY Infrared laser, Eppendorf Transferrman NK2 and Patchman NP2 micromanipulators, and a Nikon Ti-U inverted microscope. Alternatively, for material subjected to WGBS, which does not require single nucleotide polymorphism (SNP)-based analysis, hormone-primed females were mated overnight with B6D2F1 males (age 2–12 months, Jackson Labs) and zygotes were isolated as described above for oocytes.

For zygotic disruption, pronuclear stage 3 (PN3) zygotes were recovered after around 6 h of incubation and injected with a cocktail consisting of 200 ng  $\mu$ l<sup>-1</sup> Cas9 mRNA and a 100 ng  $\mu$ l<sup>-1</sup> equimolar ratio of 3–4 sgRNAs targeting different exons of an epigenetic regulator gene locus (designed using ChopChop<sup>42</sup> and the IDT CRISPR–Cas9 guide RNA checker, as previously described<sup>24</sup>) (Supplementary Table 10). Preferentially, targeted exons were chosen to be located towards the 5' end and to be shared across isoforms. At around 84 h after fertilization, cavitated blastocysts were transferred into the uterine horns of pseudopregnant CD-1 strain females (25–35g, Charles River) generated by mating with vasectomized SW strain males (Taconic), which results in a 24 h offset in gestational time to accommodate implantation.

B6/CAST F<sub>1</sub> mice were generated in house by breeding C57BL/6J strain female mice with CAST/EiJ strain males. Swimming sperm were isolated from the caudal epididymis for males (>2 months of age) in M2 media (Millipore), decapitated by brief pulse sonication in a Branson Sonifier with double stepped tip (Branson), and stored at –40 °C in 25  $\mu$ l aliquots for use within 6 months to a year of collection. Cas9 mRNA and sgRNAs were in vitro transcribed using the mMESSAGE mACHINE T7 Ultra or MEGAshortscript Kits (Thermo Fisher), purified using the RNA clean and concentrator kit (Zymogen), and resuspended in injection buffer (5 mM Tris buffer, 0.1 mM EDTA, pH 7.4).

All procedures have been performed in our specialized facility, followed all relevant animal welfare guidelines and regulations, and were approved by Harvard University IACUC protocol (28-21) and the Max Planck Institute for Molecular Genetics (G0247/13-SGr1).

### Single-cell transcriptome profiling of embryos

Wild-type and mutant embryos were isolated from surrogate mice between E6.5 and E8.5 at 12-h intervals for wild-type and at gestational day E8.5 for epigenetic regulator mutant experiments. The emergence of the EED-mutant phenotype was profiled in more detail by additionally sampling at E6.5 and E7.5. Outermost extraembryonic tissues (yolk sac, trophoblast-derived tissues) were preserved if possible. Microscope images recorded embryo number and morphology. Embryos were serially washed through several droplets of 1 $\times$  PBS and 0.4% BSA, pooled without any morphology-based pre-selection and subjected to tissue dissociation in 200  $\mu$ l TryPLE Express (Gibco) for 40–60 min at 37 °C, with pipetting in 5-min intervals. The cell suspension was filtered using

Scienceware Flowmi Cell Strainers, 40  $\mu$ m. Cells were washed twice with 1 ml 1 $\times$  PBS and 0.4% BSA, and centrifuged for 5 min at 1,200 rpm. The cell concentration was determined using a haemocytometer and cells were subjected to single-cell RNA sequencing (10x Genomics, Chromium Single Cell 3' v2 or v3) aiming for a target cell recovery of up to 13,000 sequenced cells per sequencing library. Single-cell libraries were generated following the manual instructions, with the exception of fewer PCR cycles than recommended during cDNA amplification and library generation/sample indexing to increase library complexity. Libraries were sequenced with a minimum of 230 million paired end reads according to parameters as described in the manual. For details see Supplementary Tables 1, 2.

### Imaging embryos for morphology and size measurement

E7.5 and E8.5 wild type and EED-mutant embryos acquired in experiments that were performed independently of the single-cell sequencing experiments were imaged using ZEISS AxioZoom.V16 microscope and ZENBLUE imaging software at 10 $\times$  and 7 $\times$ , respectively, with z-stacks of 12–17- $\mu$ m intervals. To obtain a higher resolution of morphology, individual E7.5 embryo images were acquired at 50 $\times$  and E8.5 embryos at 40 $\times$ . The E6.5 embryos, which were subjected to single-cell sequencing, were imaged using an Olympus IX71 inverted microscope and MetaMorph software. Images of E6.5 wild-type embryos were acquired at 4 $\times$  and mutant embryos at 10 $\times$ . Wild-type E7.5 and E8.5 embryos shown in Extended Data Fig. 10 to provide a size and morphological comparison to EED-mutant embryos were generated by natural mating. Surface area (in  $\mu$ m<sup>2</sup>) of embryos was measured using the 'region' tool by drawing a polygon contour around each embryo in ZENBLUE.

### *Prdm14*-mVenus reporter experiments: in vitro fertilization, electroporation and embryo imaging

In vitro fertilization of B6D2F1 oocytes was performed with reporter sperm from heterozygous males with mVenus under the control of *Prdm14*, as previously described<sup>43</sup>. The reporter strain was generated by the laboratory of M. Saitou<sup>31</sup> and the mVenus *Prdm14* promoter sperm (B6.Cg-Tg(*Prdm14*-Venus)1Sait/SaitRbr; BRC No. RBRC05384) were provided by the RIKEN BRC through the National Bio-Resource Project of the MEXT/AMED, Japan (Acc. No. CDB0461T; <http://www.cdb.riken.jp/arg/mutant%20mice%20list.html>)<sup>44</sup>. PN3 zygotes were washed in M2 medium and prepared for electroporation. Electroporation reactions were set up according to the Alt-R CRISPR–Cas9 ribonucleoprotein (RNP) complex protocol from Integrated DNA Technologies (IDT). RNP complexes were assembled just before electroporation. In brief, 2  $\mu$ l of 200  $\mu$ M tracrRNA and 0.67  $\mu$ l of each 200  $\mu$ M crRNA were mixed, heated to 95 °C for 5 min and allowed to anneal at room temperature for 10 min. Three microlitres of crRNA-tracrRNA mix, 1  $\mu$ l of 61  $\mu$ M Alt-R Hi-Fi Cas9 Nuclease 3NLS was diluted in 46  $\mu$ l of Opti-MEM medium and incubated at room temperature for 20 min.

The NEPA21 electroporator was used with the following settings. Impedance values were maintained between 120 and 160 k $\Omega$ . Four poring pulses of 30 V and 2.5 ms was used with an interval of 50 ms, voltage decay of 10% and (+) polarity. Transfer pulse was applied at 5 V for 50 ms with an interval of 50 ms, voltage decay of 40% and alternating polarity (+) and (–).

Zygotes that developed to blastocyst stage were screened for mVenus fluorescence in the inner cell mass as only half of the embryos are expected to carry the reporter. mVenus positive embryos were re-transferred to pseudopregnant CD-1 fosters and isolated after in vivo development at E6.5, E7.5 and E8.5. Isolated embryos were washed in cold 1 $\times$  PBS with 0.4% BSA and fixed overnight in 4% paraformaldehyde (PFA) at 4 °C followed by three washes in cold 1 $\times$  PBS. Nuclei were stained with 0.24  $\mu$ g ml<sup>-1</sup> DAPI for 40 min at 4 °C. Images were acquired using Zeiss LSM880 at 10 $\times$  magnification and z-stacks of 5- $\mu$ m intervals. Images were processed and maximum intensity projections of the z-stacks were generated using the 3D-project tool of the ImageJ software



bundled with Java 1.8.0. Four, six and seven EED-mutant embryos carrying a *Prdm14* reporter were isolated at E6.5, E7.6 and E8.5, respectively, and demonstrated similar results per developmental stage.

### Immunofluorescence

Embryos were dissected from deciduae at specific stages in cold 1× HBBS and then fixed overnight in 4% PFA at 4 °C. Embryos were rinsed three times in 1× PBS and permeabilized with PBT0.5 (0.5% Triton X-100 in 1× PBS) for 2 h followed by blocking for 1 h with blocking solution (10% FBS in PBT0.5). Embryos were incubated for 72 h at 4 °C with the primary antibody anti-H3K27me3 (Abcam ab6002), diluted in blocking solution to 1:200. The next day, embryos were washed with PBT0.5 four times (30 min per wash), and blocked overnight at 4 °C in blocking solution. The following day, embryos were incubated overnight at 4 °C with donkey anti-mouse Alexa Fluor 488 (Invitrogen A21202), diluted in blocking solution at 1:400. Embryos were subsequently washed with PBT0.5 four times (30 min per wash), and nuclei were counterstained with DAPI (0.24 µg ml<sup>-1</sup>) for 40 min at 4 °C. Embryos were washed with PBT0.5 four times (30 min per wash), and post-fixed with 4% PFA for 20 min. Final washes (three times, 15 min) were performed with 0.02 M phosphate buffer (0.025 M NaH<sub>2</sub>PO<sub>4</sub>; 0.075 M Na<sub>2</sub>HPO<sub>4</sub>, pH 7.4) followed by optical clearing at 4 °C for 24–48 h with 1.62 M RIMS clearing agent (Histodenz in 0.02 M phosphate buffer). Images were acquired using Zeiss LSM710 at 63× magnification (oil immersion) and z-stacks of 2.13-µm intervals were generated. Two independent experiments were performed with similar results. A representative z-stack is shown in Extended Data Fig. 9c.

### EED-knockout mouse ES cell line generation and fate induction experiments

Wild-type V6.5 mES cells (provided by the laboratory of K. Hochedlinger, tested negative for mycoplasma, authenticated by Nanostring for mouse pluripotency markers) were simultaneously transfected with two plasmids encoding Cas9 alongside one of two sgRNAs targeting sequences flanking the *Eed* gene locus (Supplementary Table 10). Sub-clones were expanded and homozygous *Eed* disruption was confirmed by target site amplification and Sanger sequencing.

Our selected EED-knockout cell line was expanded for 16 passages in serum/LIF to ensure complete depletion of H3K27 methylation before western blotting for H3K27me3 on histone extracts using the tri-methylhistone H3(Lys27) antibody (Cell Signalling, C36B11, at 1:500 dilution). Histone 4 was detected by anti-H4 antibody as a loading control (Millipore, 07-108, at 1:1,000 dilution). Tricine gels were used with tricine buffer and SeeBlue Plus2 Pre-stained Protein Standard (Invitrogen, LC5925). Two independent histone isolations and western blots have been performed with similar results.

Wild-type and EED-knockout mES cells were maintained in serum/LIF and cultured for at least two weeks in N2B27-containing 2i/LIF media on gelatin-coated plates to ensure their full conversion to naïve pluripotency before induction with exogenous factors<sup>45</sup>. For signalling experiments, 10,000 cells were plated per well into N2B27 media containing 12 ng ml<sup>-1</sup> bFGF. For these experiments, we used human plasma fibronectin (purified protein, Millipore) coated eight-well chamber slides (µ-Slide 8 Well, ibidi). After 24 h, media was exchanged with N2B27 containing 12 ng ml<sup>-1</sup> bFGF and select concentrations of signalling compounds and/or small molecule inhibitors for an additional 48 h. Final concentrations of growth factors or small molecule inhibitors were as follows: 12 ng ml<sup>-1</sup> recombinant human bFGF (R&D Systems); 0.25 µM retinoic acid (Sigma); 5 and 500 ng ml<sup>-1</sup> recombinant human BMP4 protein (R&D Systems); 10, 100 and 1,000 ng ml<sup>-1</sup> recombinant human WNT-3A protein (R&D Systems); 10 and 1,000 ng ml<sup>-1</sup> recombinant human/murine/Rat Activin A (Peprotech); 0.5 µM MALK2/3 inhibitor LDN-193189 (Stemgent, 10 mM solution) to inhibit the BMP4 pathway; 3.3 µM tankyrase1/2 inhibitor XAV939 (Tocris) to inhibit the Wnt pathway; and 10 µM TGFβ RI kinase inhibitor VI SB431542 (Millipore) to inhibit

Activin/Nodal signalling. Total RNA was isolated by washing wells twice with PBS followed by adding 350 µl RLT buffer as part of the RNeasy Plus Micro Kit protocol (Qiagen). Additional samples for each experiment include N2B27 containing 2i/LIF at day 0 and N2B27 containing 12 ng ml<sup>-1</sup> bFGF after 24 and 72 h, respectively. Mouse ES cell experiments and RNA isolation were done as three independent experiments.

### Expression profiling of lineage-specific genes using NanoString

To profile the expression of 44 genes and 4 housekeeping genes (*Polr1b*, *Hprt*, *Abcf1* and *Gusb*), 400 ng total RNA per sample were used in a NanoString nCounter PlexSet assay to profile 88 RNA samples of the mES cell experiments described above (triplicates for all but one condition, duplicate for 100 ng ml<sup>-1</sup> WNT3A) (Supplementary Table 11). Probe hybridization was set up according to manufacturer's instructions and performed for 24 h (MAN-10040-05). Reactions were pooled per column, generating 12 pools and run on the NanoString nCounter SPRINT Instrument. False negative probes detected up to 14 counts, which informed the magnitude of potential false negative signal. Thus, 20 counts were conservatively removed from all measurements. To provide reliable estimates on expression differences, fold changes between transcript counts in wild-type and EED knockouts were only calculated if, for a given experimental condition, the gene was detected with at least 50 counts (after background subtraction) in at least one of the two cell lines. Significance of expression differences was tested for all genes (*t*-test, R function *t.test*).

### Bioinformatics

Unless stated otherwise, all statistics and plots are generated using R version 3.5.1 'Feather Spray'. Boxes indicate the median and quartiles, with whiskers reaching up to 1.5 times the interquartile range. The violin plot outlines illustrate kernel probability density, such that the width of the shaded area represents the proportion of the data located there. For violin plots, boxes indicate the median, with quartiles and whiskers reaching up to 1.5 times the interquartile range. Heat maps were plotted using the Complex Heatmap package<sup>46</sup> and browser track figures using the Gviz package<sup>47</sup>.

### Pre-processing

The Cell Ranger pipeline version 3 (10x Genomics Inc.) was used for each scRNA-seq data set to de-multiplex the raw base call files, generate the fastq files, perform the alignment against the mouse reference genome mm10, filter the alignment and count barcodes and UMIs. Outputs from multiple sequencing runs were also combined using Cell Ranger functions.

### Genotyping: alignment

For each experiment, the scRNA-seq data were aligned against an mm10 hybrid mouse genome assembly using STAR<sup>48</sup> with default settings and '-outSAMattributes NH HI NM MD'. The hybrid genome was prepared using SNPsplite<sup>49</sup> to mask SNPs between the mouse version mm10 (GRCm38) and the CAST/Eij strain genomes with the ambiguity base (N). Subsequently, SNPsplite was used to sort reads that cover SNPs by origin (reference genome). Unambiguous and unique alignments of wild-type samples were used to create a list of SNPs that were covered by reads originating from both reference genomes. Finally, reads covering these SNPs were used to determine the allele composition for each cell as the fraction of CAST/Eij specific SNPs.

### Genotyping: cell-to-embryo assignment, doublet removal, and sex determination

Single cells were assigned to embryos according to the autosomal fraction of CAST SNPs, a 19-dimension vector that allowed us to estimate the number of embryos per experiment. A minimum number of 1,000 covered SNPs and SNP information for each autosome was required. *k*-means clustering for multiple *k* (*k*-means function in R,

$k=2-15$ , default parameters) was performed on cells that fulfilled this criterion and evaluated by calculating the AIC for each model. The  $k$  with the minimal AIC defined the number of detected embryos, and the kernel averages represent the SNP profile for each embryo in the pool. Cells were then assigned to the embryo based on minimum distance in their SNP profile.

We found that poor cell to embryo assignments were often the result of either low or very high UMI counts, probably representing cell multiplets. To eliminate these, we performed 100 iterations of our embryo assignment strategy using a randomly sampled 20% of each cell's SNPs and discarded cells that changed their assignments (Extended Data Fig. 1b). Stably assigned cells were consistently assigned to the same embryos based on the  $k$ -means clustering (Supplementary Table 2).

Embryo sex was determined based on the expression of the following genes: *Xist* (ENSMUSG00000086503) to count XX contexts and *Erd1* (ENSMUSG00000096768), *Ddx3y* (ENSMUSG00000069045) and *Eif2s3y* (ENSMUSG00000069049) to reliably detect transcription from the Y chromosome. The Cell Ranger gene barcode matrices were used to obtain per cell expression counts for these 4 genes and determine the fraction of positive cells per embryo. Embryos with a high percentage of *Xist* expressing cells were determined to be female while embryos with higher fractions of *Erd1*, *Ddx3y* or *Eif2s3y* were determined to be male (Supplementary Table 1).

## Genotyping: cluster determination

Cluster determination was split into four main parts and was largely done using the R package Seurat with default settings<sup>50</sup>. The establishment of the wild-type reference cell states was published previously using Cell Ranger version 2 processed data<sup>16</sup>. In brief, (1) a preliminary set of clusters were generated by clustering wild-type embryos of the same stage as a pool without taking replicate identity into account, followed by generating per replicate clusters according to this assignment. Then, (2) replicate embryo clusters from step 1 were used to generate median expression vectors and clustered across time points to obtain preliminary cell states. Next, (3) all wild-type cells were assigned to their most similar cluster by Euclidean distance according to a reduced set of 712 marker genes to determine the specific cell-state kernel. Finally, (4) all wild-type and mutant embryo cells were assigned to their most similar cluster by Euclidean distance according to a reduced set of 706 marker genes to determine their specific cell-state identity after reprocessing with Cell Ranger version 3.

(1) Embryo-specific centres (wild-type): All de-convoluted wild type single cells of the same developmental stage were processed together after discarding cells that were not confidently assigned to a genotype/embryo. Parameters were adopted from the Seurat manual. The expression data were log-normalized, scaled to 10,000 and UMI biases were removed (vars.to.regress = "nUMI"), followed by calling of variable genes (parameters: mean.function = ExpMean, dispersion.function = logVMR, x.low.cutoff = 0.0125, x.high.cutoff = 3, y.cutoff = 0.5). Next, the variable genes were used to run the PCA and the first 20 PC's were used for cluster detection. The average expression for each embryo and cluster was calculated, which we refer to as 'embryo-specific centres'. This allowed us to detect even rare cell states while preserving embryo-specific variability.

(2) Cell clusters (wild-type): The embryo-specific centres of all wild-type stages were combined into one analysis to determine variable genes. A PCA was run based on the variable genes and the first 20 PCs were used to cluster the embryo specific centres (parameters adjusted for low 'cell' number: k.param = 8, k.scale = 50, prune.SNN = 1/10). This resulted in 42 clusters of embryo-specific centres and the median expression profile of each cluster was calculated to form preliminary cell states. Then, as a temporary step, all wild-type cells from all stages were simultaneously assigned to their closest preliminary cell state based on expression similarity (Euclidean distance of log-expression

values of variable genes calculated above) to calculate a gene expression average (kernel).

At this stage, we observed that the number of variable genes was unevenly distributed across preliminary cell states, which created biases when comparing single cells across them (clusters defined by a greater number of variable genes have more opportunities to match sparse single-cell measurements, while those defined by fewer variable genes accumulate more noise by including them). We therefore sought to normalize the number of state-specific genes that contribute to each cluster by using the top 30 marker genes (highest difference in fraction of positive cells within the cluster versus other clusters) from each of the 42 cell states. We found that this reduced gene set provides a more stable, lower-noise assignment without biasing the information to describe each cell state ( $n=712$  unique genes, Extended Data Fig. 1e) and used this set of genes in (3).

(3) Refinement of wild-type reference cell states: wild-type cells were assigned to cell-state expression profiles (kernels) based on their Euclidean distance log-expression values for the 712 marker genes. Single-cell distances are significantly smaller to their matched cell states than to next-best matches. Cell Ranger version 3.0 was released by 10x Genomics over the course of the generation of this manuscript. Thus, raw data were reprocessed and the cell-state kernels were adjusted by again assigning the wild-type cells to the kernels.

(4) Cell states of single cells: The wild-type and mutant embryo cells were assigned to the cell states based on their Euclidean distance log-expression values for the now 706 marker genes (Cell Ranger v3 adjusted). Single-cell distances are significantly smaller to their matched cell states than to next-best matches (Extended Data Figs. 1f, 2e, 6a). Cell states with an insufficient number of cells from mutant embryos ( $\leq 30$  cells) were discarded from further analysis (except PGC state).

We believe our experimental strategy should largely account for differences in embryo genotype by sampling multiple siblings: each allele will only be heterozygous for the *castaneus* background in 50% of embryos, our trends are generally observed across all replicates, and the processes of gastrulation are highly conserved. Nonetheless, we cross referenced our 712 marker genes against those with reported *castaneus* expression biases across 23 adult and embryonic tissues, including those from all three germ layers, the extraembryonic ectoderm, and the extraembryonic endoderm<sup>51</sup>. Of the 1,530 genes that show biased expression in at least 10% of these tissues within an  $F_1$  context (with reciprocal crosses to control for potential imprinting), only 53 were also marker genes (0–7 per cell state, median 2). Furthermore, we saw that all cell states consisted of several embryos and never resulted from a single embryo.

## Cell-state prevalence

Prevalence of cell states with respect to embryo stage (Extended Data Fig. 1h) was evaluated by normalizing each state across the recovered time points (row).

## Cell-state proportions

Cell-state proportions per embryo were calculated as the number of cells assigned to a cell state divided by the total number of cells comprising an embryo. The stage specific median embryo was calculated as the median proportion of cell-state fractions of all embryos from the same developmental stage (applied after our delay adjusted assignment, see below). Proportion changes in Fig. 2c were calculated as the  $\log_2$ -transformed fold change between the mean proportions of developmentally stage-matched mutant and wild-type embryos.

## Correlation of gene expression

Gene expression profiles were compared between wild-type and EED-mutant states by correlating the average gene expression profiles of the marker genes (R function cor, Pearson correlation).

## Differential expression

We called differentially expressed genes between wild-type and mutant experiments for every detectable cell state. To account for changes in 10x Genomics chemistry versions and possible batch effects, we ran the `removeBatchEffect` function of the `limma` package per cell state across all samples<sup>52</sup>. For comparisons across all embryos, we normalized our percent positive cells data with the same function for each state individually. The resulting normalized read count data were used for differential gene expression of the mutant vs wild-type cells. A gene was called differentially expressed within a cell state between wild-type and mutant if it fulfilled the following criteria: (1) adjusted *P* value of  $< 0.05$ , (2) minimum detectable fraction of 0.05 within at least one condition (WT or KO) and (3) a minimum difference of 10% transcript positive cells and a minimum absolute  $\log_2$ -transformed fold change of 0.2. Sex chromosomal genes were excluded from further analysis, as well as the PGC cell state as it is not highly observed across many of the mutants that proceed to later developmental stages.

We assigned genes as recurrently deregulated if they were differentially expressed in at least two cell states within the extraembryonic derived lineages (Xendo, Xecto) or the embryonically derived lineages (epiblast, extraembryonic mesoderm, embryonic endoderm, embryonic ectoderm, embryonic mesoderm) and was prevalently up- or downregulated (Supplementary Table 6).

Pathway enrichment for the recurrently differentially expressed genes was performed by a hypergeometric test using the GSEA online tool. The *P* value was adjusted for multiple testing according to Benjamini and Hochberg, with 0.05 as a cut off (Supplementary Table 7).

## Stage matching metric to assign 'developmental stage'

The gestational age of all mutant embryos was adjusted for developmental delay by comparing cell-state data to the median of the wild-type embryos from each time point. Because some states may be more informative about developmental stage than others, we performed two distinct PCAs using the wild-type replicate data: (1) using the cell-state proportions and (2) using the binary information on presence and absence of a cell state. For the cell-state proportion assignments, only the embryonic cell states were used (embryonic mesoderm, embryonic endoderm, embryonic ectoderm, epiblast, PGCs and extraembryonic mesoderm), because Xendo and Xecto cell-state proportions are more sensitive to technical variability during embryo dissection and single cell dissociation. The R function `prcomp` (parameters: `retx = TRUE`, `centre = TRUE`, `scale = TRUE`) was used to calculate PCs for wild-type embryos and wild-type medians and the predict function transformed mutant embryo data according to the wild-type loadings (Extended Data Fig. 3b–d, Supplementary Tables 1, 5). The first principal components of both PCAs were used to assign each mutant embryo to its closest median wild type by Euclidean distance.

## H3K27me3 ChIP-seq data

Publicly available H3K27me3 ChIP-seq data of E6.5 epiblast and extraembryonic ectoderm<sup>53</sup> were used to calculate the average H3K27me3 occupancy of the promoter region of each gene (calculated as the region 1,500 bp upstream to 500 bp downstream of the TSS). Only the first two replicates were used for each, because these two replicates showed a similar trend when compared to wild-type gene expression of our epiblast or Xecto cell states, whereas the third replicate did not show any linear relation to gene expression. For both datasets, a cutoff of 400 (average H3K27me3 peak level) showed the strongest drop in gene expression and thus most likely represents functional repression by H3K27me3. The binary assignment of having a promoter H3K27me3 peak was set at this threshold.

## PGC number estimation

The total number of PGCs per wild-type or EED-mutant embryo was estimated using the fraction of recovered state 27 (PGC) cells within

that embryo multiplied by its total estimated cell number. The total estimated cell number was calculated by multiplying the fraction of the embryo within the pool to the total number of cells in the single-cell suspension (as measured using a haemocytometer, see above). We then applied a correction to account for potential technical biasing of embryonic versus extraembryonic sampling during isolation, though this did not change our estimates substantially. The enrichment was tested using the Wilcoxon test (R function `wilcox.test`, two-sided). All state 27 counts are given in Supplementary Table 5.

## UMAP projection

UMAP was used as a dimension reduced visualization of single-cell marker gene expression profiles<sup>54</sup>. Transformation of the wild-type data was performed using the R function `umap` and subsequently applied to all mutant embryo data to project it onto the same manifold as produced for the wild type.

## RNA velocity

RNA velocity was calculated using the `velocyto` tool<sup>55</sup> and visualized using `scanpy`<sup>56</sup>. The previously calculated UMAP was used for velocity projection.

## Cut site analysis

Single reads covering the targeted genes were extracted from the initial alignment and were realigned against the intron-free cDNA sequence of the respective gene using STAR<sup>48</sup> with default settings and '`-alignEndsType EndToEnd -outSAMattributes NH HI NM MD`'. The aligned reads were next classified with respect to the target site of the sgRNA as: (1) 'spliced/deleted' if they did not match any nucleotide but were spanning across the entire target site, (2) 'mismatched' if any of the nucleotides were aligned as a mismatch/deletion/insertion to the reference, (3) 'complete' if all nucleotides matched the target site, or (4) 'insufficient' if the reads did not span the full target site.

## Retrotransposon detection

To quantify retrotransposon expression, only reads that do not overlap with gene annotations were considered. In addition, split reads as well as reads containing an extensive poly-A stretch were excluded. A read was defined as covering a poly-A region if (1) the last 70% of bases were mainly A (A stretch with maximal 10 bases C, G or T) or (2) the first 70% were mainly T (T stretch with maximal 10 bases A, C or G). The remaining reads were overlapped with annotated repetitive elements (repeat masker file downloaded from UCSC) and reads with a minimum overlap of 90% were considered for further analysis. Reads that mapped uniquely or multiple times to the same repeat family were counted once per family, reads that mapped to different repeat families were excluded. Subsequently, reads were counted per repeat family, embryo, and cell state and then normalized to full number of considered reads (number of repeat reads plus number of UMIs sequenced).

## WGBS library generation and data processing

E6.5 epiblast and Xecto were isolated from at least 7 embryos, pooled and processed into WGBS libraries using the Accel-NGS Methyl-seq kit as previously described using  $\leq 9$  final PCR cycles<sup>24</sup>. Reads were aligned to the mouse mm10 reference genome using BSMAP with flags '`-v 0.1 -s 16 -w 100 -S 1 -q 20 -u -R`'. To determine the methylation state of all CpGs captured and assess the bisulfite conversion rate, we used the `mcall` module in the MOABS software suite with standard parameter settings<sup>57</sup>. Finally, we converted the resulting CpG level files to bigwig files, filtering out all CpGs that were covered with less than ten reads.

For all downstream analysis, replicates were averaged after having applied the coverage cutoff and differentially methylated CpGs/genomic regions were defined by having a minimum difference of 0.1 to the respective wild-type tissue.



# Article

CGIs were downloaded from the UCSC genome browser, gene annotations were obtained from the build-in Cell Ranger version gtf file, and promoter regions were defined as 2.5 kb upstream to 500 bp downstream of annotated TSS. Xecto hypermethylated CGIs were previously defined<sup>24</sup>.

The CpG density of a genomic region was calculated as the fraction of CpG dinucleotides within a 100 bp window (sliding window with 20 bp offset).

DMVs were detected using a 2-kb sliding window (500 bp offset). Regions with an average methylation rate below 0.15 in wild-type (excluding CGI methylation) were merged given a maximum distance of 1 kb.

## Reporting summary

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

## Data availability

All datasets have been deposited in the Gene Expression Omnibus (GEO) and are accessible under GSE137337 (or published previously under GSE122187). Source data for Figs. 1a, b, 2, 3a, b, d–f, h, i, 4a, c–f, Extended Data Figs. 1b, c, e–i, 2b–f, 3, 4b–f, 5a–c, e, 6, 7, 8a, 9b, 10b–g, 11c, d are available at <https://oc-molgen.gnz.mpg.de/owncloud/s/F8g3y5F79JZRYof>. Previously published data used in this study include H3K27me ChIP–seq data (GSE98149), WGBS data for sperm and oocyte (GSE112320), preimplantation samples, including 8 cell stage embryos and the inner cell mass and trophectoderm of the E3.5 blastocyst (GSE84236), and late stage samples including an average of somatic tissues and the E14.5 placenta (GSE42836).

## Code availability

Code is available at [https://github.com/HeleneKretzmer/EpigeneticRegulators\\_MouseGastrulation](https://github.com/HeleneKretzmer/EpigeneticRegulators_MouseGastrulation).

40. Wang, H. et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918 (2013).
41. Platt, R. J. et al. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440–455 (2014).
42. Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **42**, W401–W407 (2014).
43. Nakagata, N. Cryopreservation of mouse spermatozoa and in vitro fertilization. *Methods Mol. Biol.* **693**, 57–73 (2011).
44. Ohinata, Y. et al. A comprehensive, non-invasive visualization of primordial germ cell development in mice by the Prdm1-mVenus and Dppa3-ECFP double transgenic reporter. *Reproduction* **136**, 504–514 (2008).
45. Ying, Q.-L. & Smith, A. G. Defined conditions for neural commitment and differentiation. *Methods Enzymol.* **365**, 327–341 (2003).
46. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
47. Hahne, F. & Ivanek, R. Visualizing genomic data using Gviz and Bioconductor. *Methods Mol. Biol.* **1418**, 335–351 (2016).
48. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
49. Krueger, F. & Andrews, S. R. SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes. *Fluorescence* **5**, 1479 (2016).
50. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
51. Andergassen, D. et al. Mapping the mouse Allelome reveals tissue-specific regulation of allelic expression. *eLife* **6**, e146 (2017).
52. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
53. Wang, C. et al. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat. Cell Biol.* **20**, 620–631 (2018).
54. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for dimension reduction. Preprint at <https://arxiv.org/abs/1802.03426> (2018).
55. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).

56. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15–5 (2018).
57. Sun, D. et al. MOABS: model based analysis of bisulfite sequencing data. *Genome Biol.* **15**, R38–R12 (2014).
58. Keane, T. M. et al. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289–294 (2011).
59. Lei, H. et al. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* **122**, 3195–3205 (1996).
60. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247–257 (1999).
61. Tachibana, M. et al. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.* **16**, 1779–1791 (2002).
62. Yu, B. D., Hess, J. L., Horning, S. E., Brown, G. A. & Korsmeyer, S. J. Altered Hox expression and segmental identity in *Mll*-mutant mice. *Nature* **378**, 505–508 (1995).
63. Hammoud, S. S. et al. Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis. *Cell Stem Cell* **15**, 239–253 (2014).
64. Smallwood, S. A. et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* **11**, 817–820 (2014).
65. Nashun, B. et al. Continuous histone replacement by Hira is essential for normal transcriptional regulation and de novo DNA methylation during mouse oogenesis. *Mol. Cell* **60**, 611–625 (2015).
66. Hon, G. C. et al. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat. Genet.* **45**, 1198–1206 (2013).
67. Auclair, G., Guibert, S., Bender, A. & Weber, M. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biol.* **15**, 545 (2014).
68. Wymeersch, F. J. et al. Position-dependent plasticity of distinct progenitor types in the primitive streak. *eLife* **5**, e10042 (2016).
69. Niswander, L., Yee, D., Rinchik, E. M., Russell, L. B. & Magnuson, T. The albino deletion complex and early postimplantation survival in the mouse. *Development* **102**, 45–53 (1988).
70. Faust, C., Lawson, K. A., Schork, N. J., Thiel, B. & Magnuson, T. The Polycomb-group gene *eed* is required for normal morphogenetic movements during gastrulation in the mouse embryo. *Development* **125**, 4495–4506 (1998).
71. Kalantry, S. & Magnuson, T. The Polycomb group protein EED is dispensable for the initiation of random X-chromosome inactivation. *PLoS Genet.* **2**, e66 (2006).
72. Niswander, L., Yee, D., Rinchik, E. M., Russell, L. B. & Magnuson, T. The albino-deletion complex in the mouse defines genes necessary for development of embryonic and extraembryonic ectoderm. *Development* **105**, 175–182 (1989).
73. Han, J. et al. Tbx3 improves the germ-line competency of induced pluripotent stem cells. *Nature* **463**, 1096–1100 (2010).
74. Magnúsdóttir, E. & Surani, M. A. How to make a primordial germ cell. *Development* **141**, 245–252 (2014).
75. Arnold, S. J. & Robertson, E. J. Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat. Rev. Mol. Cell Biol.* **10**, 91–103 (2009).
76. Semrau, S. et al. Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. *Nat. Commun.* **8**, 1096 (2017).

**Acknowledgements** We thank A. Bolondi, R. Weigert and other members of the Meissner laboratory, M. Chan and D. Hnisz for discussions and advice, S. Otto for experimental support characterizing the EED-knockout mES cell line, M. Walter for support with embryo isolations, T. Ahsendorf for help with initial efforts to optimize our genotyping pipeline, and D. Andergassen for discussions on SNP typing. We are also grateful to F. Koch and the transgenic facility, including M. Peetz, and C. Giesecke-Thiel of the Flow Cytometry Facility for their feedback and support. We thank M. Saitou for the mVenus *Prdm14* promoter sperm that were provided by the RIKEN BRC through the National Bio-Resource Project of the MEXT/AMED, Japan (Acc. No. CDB0461T). This work was funded by the National Institutes of Health (NIH; 1P50HG006193, P01GM099117, 1R01HD078679 and 1DP3K111898) and the Max Planck Society.

**Author contributions** S.G., Z.D.S. and A.M. designed and conceived the study. S.G., H.K., Z.D.S. and A.M. prepared the manuscript with the assistance of the other authors. Z.D.S. performed the mouse experiments for scRNA-seq and WGBS. S.G. and Z.D.S. performed all scRNA-seq experiments. L.W. and A.S.K. performed in vitro fertilization and electroporations for reporter experiments. A.S.K. conducted the staining and imaging of the embryos supervised by S.G. S.G. and Z.D.S. made the scRNA-seq libraries with raw data processing by S.K., B.T. and H.K. Downstream data processing was performed by H.K. Analysis of repetitive elements from scRNA-seq data was performed by S.H. Data analysis was performed by S.G., H.K. and Z.D.S. and assisted by S.M. S.G. conducted in vitro experiments and analysed Nanostring data. A.M. supervised the work.

**Competing interests** The authors declare no competing interests.

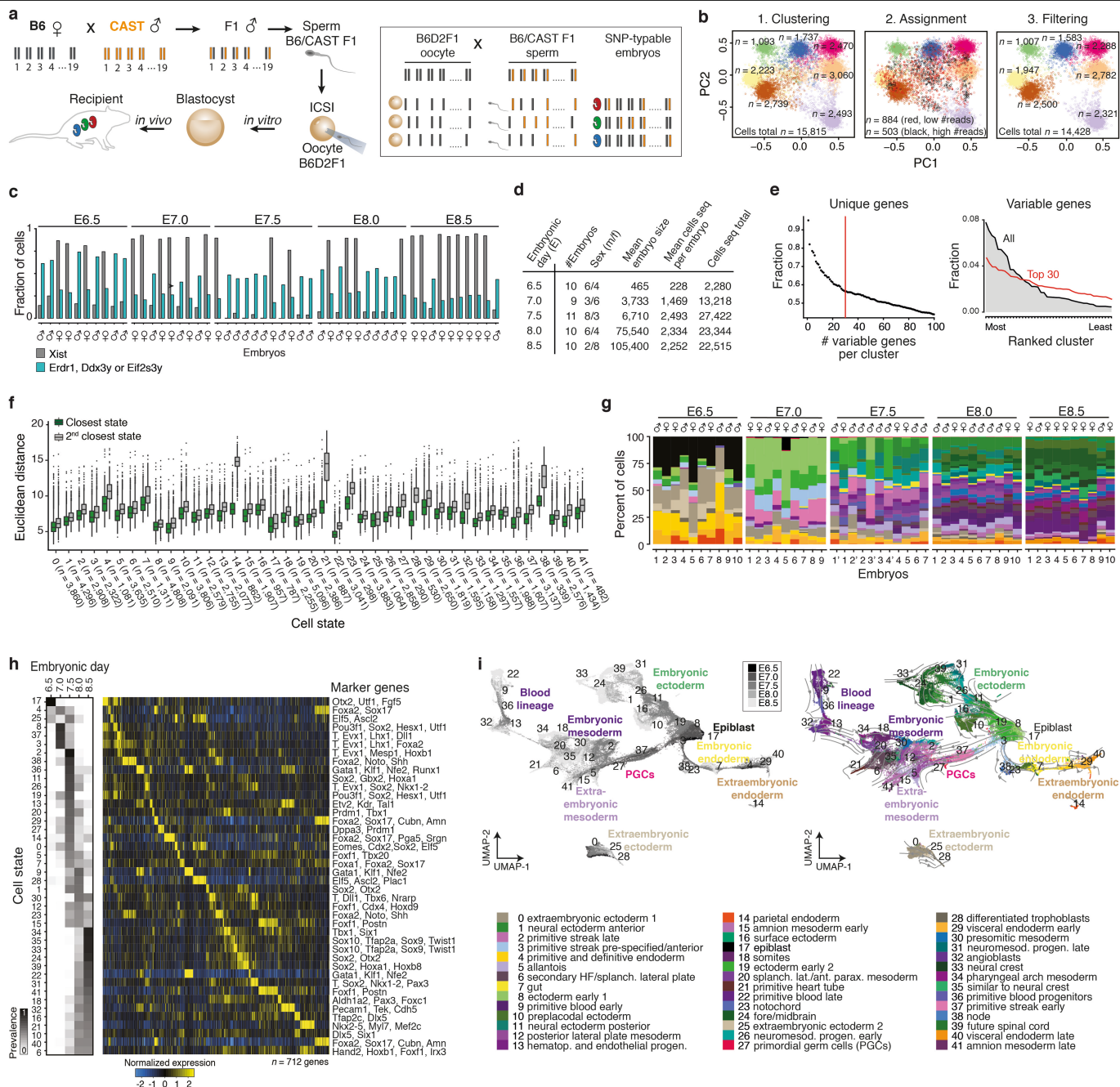
## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2552-x>.

**Correspondence and requests for materials** should be addressed to A.M.

**Peer review information** *Nature* thanks Haruhiko Koseki, Terry Magnuson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



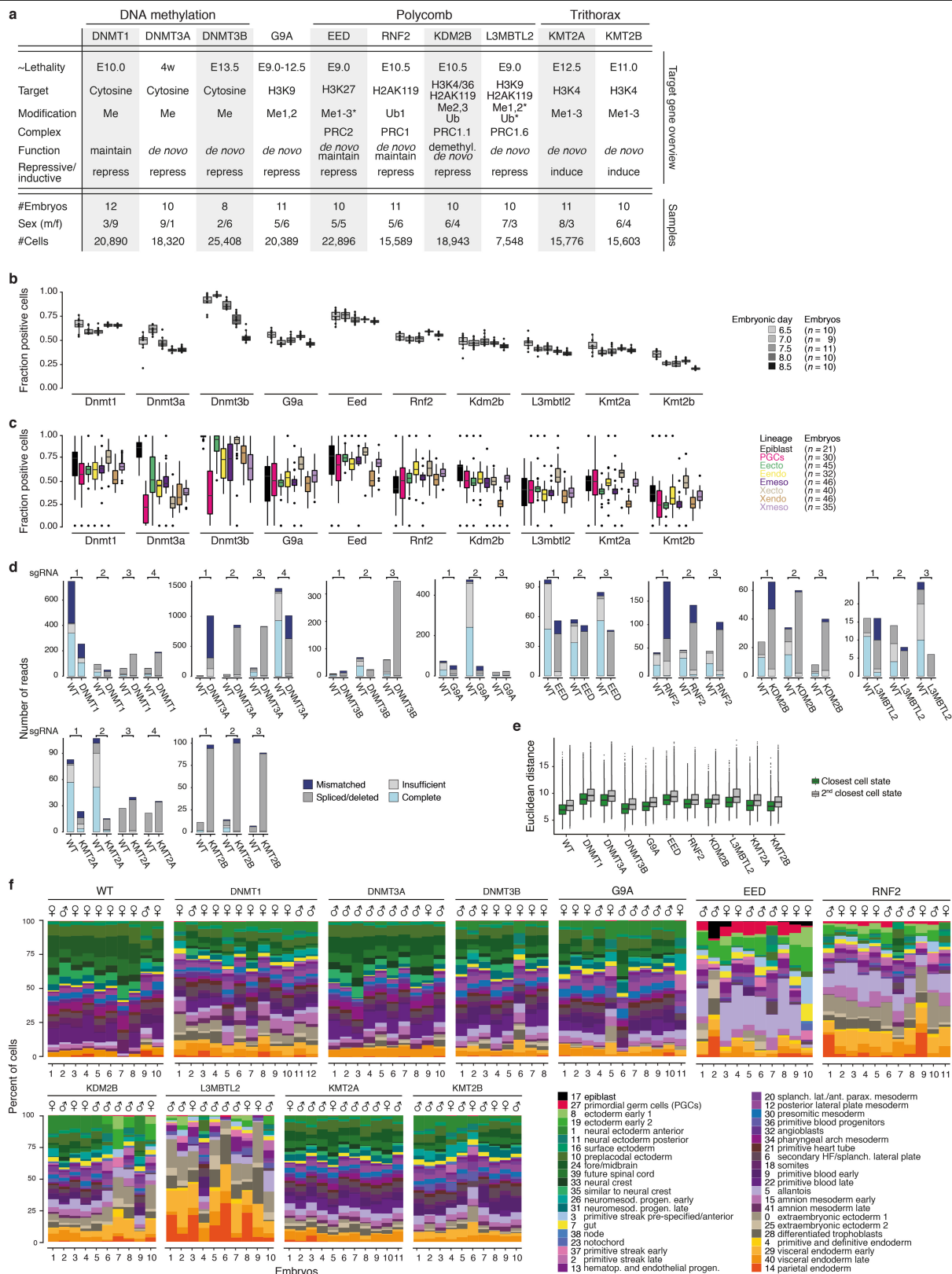
**Extended Data Fig. 1** | See next page for caption.

# Article

**Extended Data Fig. 1 | SNP-based genotyping and assignment of single cells into 42 discrete cell states.** **a**, SNP-based cell-to-embryo assignment strategy. Embryos were generated by intracytoplasmic sperm injection using sperm from hybrid males (C57BL/6j × CAST/Eij) to confer a randomly inherited CAST/Eij haplotype. Siblings (individually coloured embryos) are pooled before scRNA-seq and computationally deconvoluted based on their embryo-specific SNP profiles. In brief, the ratios of CAST-specific SNPs (orange) are scored per chromosome to cluster cells into distinct embryos. We use B6D2F1 (C57BL/6j × DBA) oocytes, whose genotypes differ by only approximately 4.5 million SNPs compared with 17.7 million for CAST/Eij<sup>58</sup>. **b**, SNP-based deconvolution of seven pooled E7.5 wild-type embryos. Left, principal component analysis (PCA) projection of autosomal CAST SNP ratios for all sequenced cells with ≥1,000 covered SNPs. Cells are coloured by cluster assignment, indicating individual genotypes (embryos). Centre, iterative sampling of 20% covered SNPs per cell flags cells with unstable embryo assignments. Flagged cells with lower than median SNP counts represent low quality cells, whereas those with higher counts collect between clusters and probably reflect doublets. Cells with unstable genotype assignments were excluded from further analysis. Right, PCA projection of all cells that were stably assigned to an embryo. **c**, Per embryo fraction of cells with *Xist* (grey) and three Y-chromosome linked gene transcripts (*Erd1*, *Ddx3y* or *Eif2s3y*; blue) used for sex-typing. For cell numbers, see Supplementary Tables 1 and 2. **d**, Summary statistics of profiled wild-type embryos from E6.5–E8.5 ( $n = 50$  total). **e**, Left, fraction of variable genes that are

uniquely assigned to a single state when taking the top  $N$ -most differentially expressed genes per cluster. We selected the top 30 most unique genes per cluster ( $n = 712$  genes) because it maximizes the information per cluster under the constraint that the number of marker genes be as similar across states as possible. Right, ranked order distribution for the fraction of all variable or of the top 30 marker genes expressed in each of our 42 states. Our top 30 marker criterion reduces the range of variable genes that are used to assign single cells to each state. **f**, Single-cell Euclidean distances to their closest (green) or second closest (grey) state. The distribution of differences between first and second closest cluster are all significant ( $P < 2 \times 10^{-16}$ , Wilcoxon test, two tailed, paired test). **g**, Per embryo bar plots show the percentage of cells (y axis) assigned to each cell state ( $n = 42$  states, 50 embryos total). For absolute cell counts, see Supplementary Tables 1 and 5. **h**, Left, heat map of the prevalence of cell states across profiled embryonic stages. The median state proportions are calculated across embryos for each time point, and then row-normalized across time points to show their dynamics. Right, expression heat map of our 712 marker genes, with key markers for each state highlighted (see Supplementary Information). Mean state expression for each marker gene is normalized over the column and arranged by maximal expression value across states. **i**, Left, UMAP of wild-type cells ( $n = 88,779$ ) coloured by time point from dark to light grey. Right, wild-type UMAP overlaid with RNA velocity<sup>55</sup> information as an indicator of transcriptome dynamics between different cell states.



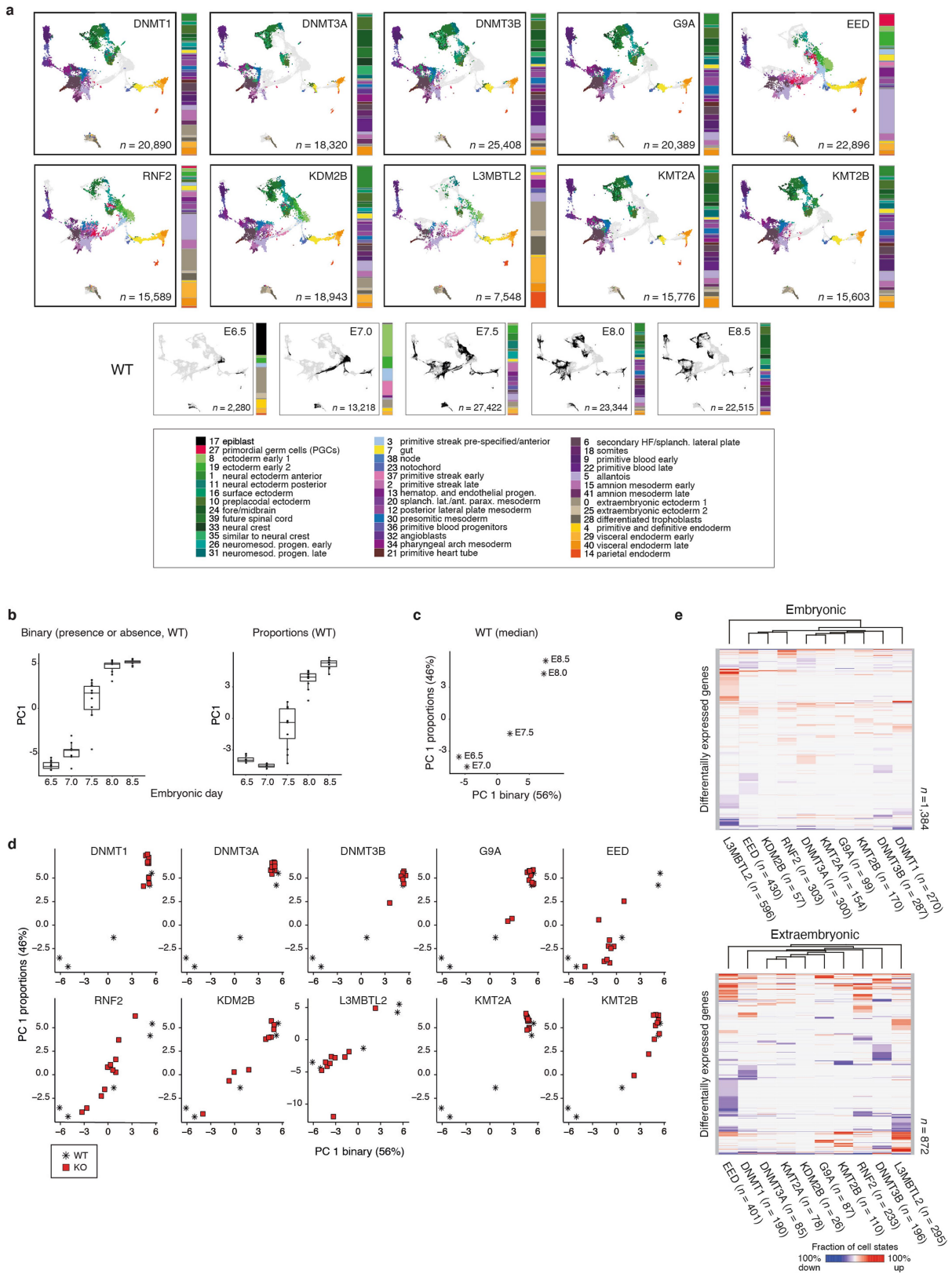


**Extended Data Fig. 2** | See next page for caption.

# Article

**Extended Data Fig. 2 | Efficient genetic perturbation of epigenetic regulators and cell-state characteristics across embryo replicates.** **a**, Top, epigenetic regulators investigated here with information about their target residues, function and grouped into three key pathways: regulation by DNA methylation, Polycomb or Trithorax. Most lethal phenotypes occur soon after our last experimental collection time point (E8.5)<sup>22,26,28–30,59–62</sup>. L3MBTL2 is a methyl-histone binding protein that participates in the regulation of PRC1 as part of non-canonical PRC1.6. L3MBTL2 and EED do not possess denoted enzymatic activities (asterisks) but are involved in the functionality of a multicomponent complex. DNMT3A mutants die postnatally (w, weeks), with signs of defective neural development that may initiate in utero. Bottom, summary statistics of scRNA-seq data generated for E8.5-isolated embryos with mutations in one of ten target epigenetic regulators ( $n=103$  embryos total). **b**, Fraction of cells positive for selected epigenetic regulator genes in wild-type embryos ordered by developmental stage (E6.5–E8.5). The de novo DNA methyltransferases *Dnmt3b* and to some degree *Dnmt3a* become less expressed as the embryo develops, congruent with their early role in remethylating the genome shortly after implantation. The  $n$  values reflect the number of embryos collected at each time point. **c**, Fraction of cells positive for selected epigenetic regulator genes in wild-type embryos for eight major developmental lineages. The  $n$  values reflect the number of embryos from which each lineage was recovered. **d**, Reads spanning the sgRNA protospacer sequences confirm highly efficient disruption of epigenetic regulator loci. Reads are grouped into the following categories: mismatched, at least one base

is a mismatch, deletion or insertion; spliced/deleted, split read spans discontinuously over the protospacer sequence; insufficient, reads do not span the entire cut site; complete, reads map without any mismatches to the cut site. Mapping distribution of scRNA-seq reads from wild-type E8.5 embryos is shown in comparison for each target site. A more comprehensive analysis of zygotic disruption is presented for EED-mutant embryos in Extended Data Fig. 9. **e**, Mutant embryo cells can be described using wild-type-defined states. Box plots show the single cell Euclidean distances to their closest (green) and second closest (grey) states per experiment. The differences between first and second closest cluster are all significant ( $P < 2 \times 10^{-16}$ , Wilcoxon test, two tailed, paired test). We observe similar differences between first and second state assignment between mutant cells as we do for the wild-type cells from which our state kernels were derived.  $n=88,779; 20,890; 18,320; 25,408; 20,389; 22,896; 15,589; 18,943; 7,548; 15,776$ ; and  $15,603$  (left to right). **f**, Bar plots showing the percentage of cells per embryo (x axis) that were assigned to each of our 42 cell states (colours, y axis) with E8.5 wild-type embryos provided for comparison. Notably, mutant embryos frequently match earlier developmental stages (Fig. 2a, b, Extended Data Fig. 1g, for comparison). Aberrant cell-state proportions indicate morphological abnormalities beyond developmental delay. For example, L3MBTL2 mutants underproduce early embryonic states, whereas EED and RNF2 mutants initially progress through gastrulation but substantially overproduce posterior products, such as allantois and amnion (states 5, 15 and 41, respectively). For absolute cell counts, see Supplementary Table 5.



**Extended Data Fig. 3** | See next page for caption.

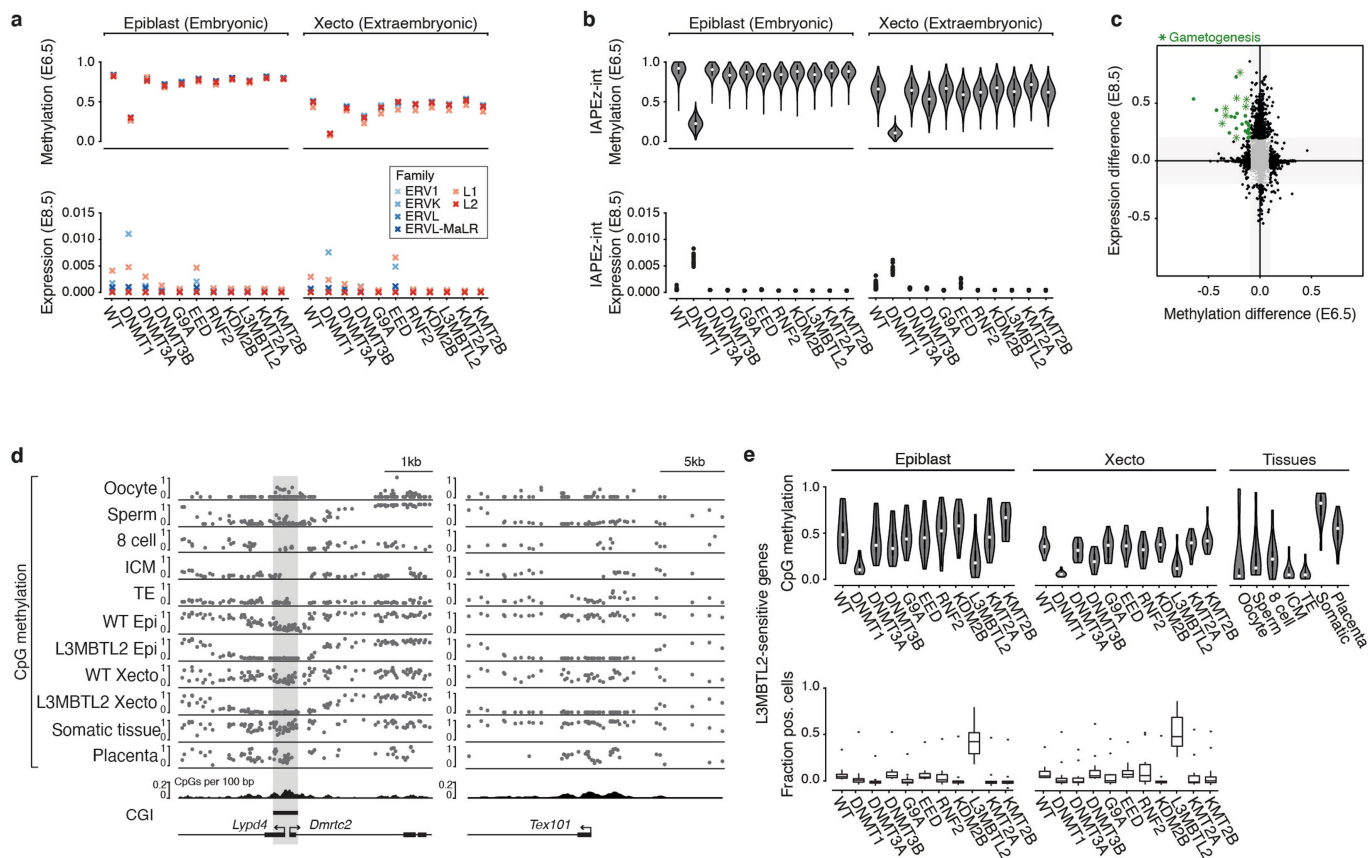
## Extended Data Fig. 3 | Quantifying developmental delay of mutant embryos by cell-state composition.

**a.** Cell-state composition of epigenetic regulator mutants. Cells were assigned to one of 42 wild-type cell states and projected onto our wild-type-defined gastrulation UMAP. That mutant cells fall within wild-type states cannot confirm equivalent functionality or potential, but does suggest that cell states are largely constrained even without key epigenetic regulators. Instead, many mutant embryos differ from wild-type embryos by cell-state composition. Adjacent bar plots reflect the median embryo composition. A reference key for our wild-type time series is provided. *n* denotes number of cells. **b.** Distribution across principal component 1 (PC1) for wild-type embryos (dots *n* = 10, 9, 11, 10 and 10) per time point using two data resolutions: a thresholded, binarized score of state presence (left), or the exact proportion (right). In PC1 space, embryos from early developmental stages (that is, E6.5–E7.5) are better resolved according to the presence or absence of key states associated with the primitive streak, whereas later time points (that is, E8.0–E8.5) share many of the same states, but at different proportions. Tissues prone to technical recovery biases during embryo isolation (Xecto and Xendo) were excluded from this analysis (Supplementary Table 5). **c.** PC1 values for median wild-type embryos (*n* = 5 time points, asterisks). PCAs were based on the binary presence or absence of cell states (*x* axis) and on cell-state proportions (*y* axis). **d.** Developmental staging of single mutant embryo replicates (squares) by projecting them onto the wild-type-defined PCA space described in **c**. Mutants of the DNA methyltransferases and the histone

methyltransferases KMT2A and G9A show no or mild developmental delays. The Polycomb components EED, RNF2, KDM2B and L3MBTL2 exhibit stronger setbacks in developmental progression, with greater variability. For staging information, see Supplementary Table 1. *n* = 12, 10, 8, 11, 10, 11, 10, 11 and 10 embryos. **e.** Clustering of epigenetic regulator mutants based on genes that are recurrently differentially expressed across cell states. Expression changes were determined from scRNA-seq data by comparing each mutant to wild-type cell state, split by embryonic (top) or extraembryonic (bottom) origin (Supplementary Table 6). Differentially up- or downregulated genes found in at least two states are shown in red and blue, with colour intensity reflecting the fraction of cell states that change in a given direction (calculated as an average of +1 and –1 states). Within the embryonic lineage, the KDM2B mutant clusters with canonical PRC subunits, even though it progresses further in development. Other regulators show expression differences in fewer cells states and many correspond to within-lineage transitions (Supplementary Tables 6, 7). In these contexts, we cannot distinguish whether lineage-specific regulation has been impeded or whether these differences are merely a consequence of subtly offset development. Additional Gene Ontology (GO) term analysis: upregulated genes in mutants of the three DNMT enzymes significantly overlap with imprinted genes ( $Q = 1.3 \times 10^{-12}$  for DNMT1 mutant embryonic, upregulated) and our G9A mutant is statistically enriched for genes upregulated in a transgenic G9A-knockout model ( $Q = 0.002$  for G9A-mutant embryonic, upregulated).



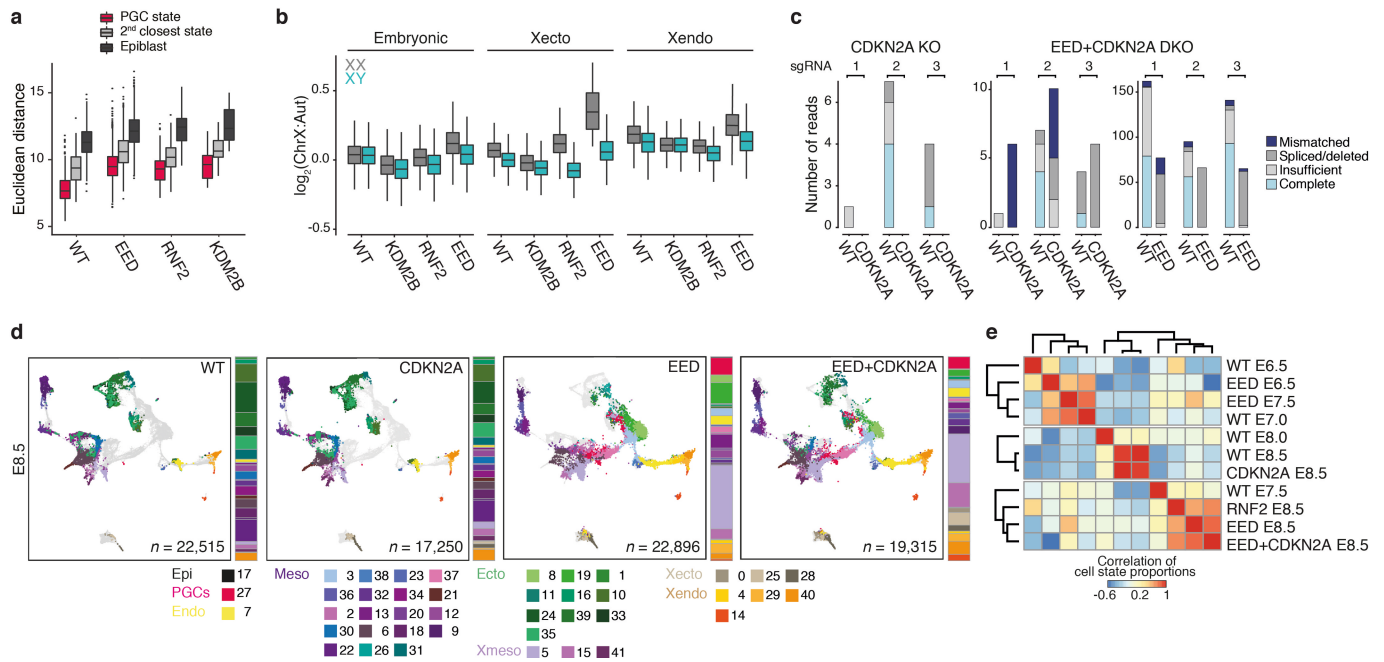




**Extended Data Fig. 5 | DNA methylation-dependent changes in gene and retrotransposon expression.** **a**, Average E6.5 DNA methylation (top) and E8.5 expression (bottom) for retrotransposon families. Expression was calculated as the normalized fraction of reads recovered from scRNA-seq data for each subfamily. DNMT1 mutants show the strongest reduction in methylation across retrotransposons in the epiblast and Xecto. The ERVK family of LTRs shows the strongest corresponding increase in expression, which is higher in the embryonic lineage than in Xecto. **b**, IAP expression as detected by scRNA-seq depends on DNA methylation. Top, DNA methylation levels as profiled by WGBS. The largest drop in global and IAP-specific methylation is observed for DNMT1 mutants. Bottom, mean expression within the embryonic and Xecto lineages of E8.5 mutant embryos, shown as the fraction of total reads. For the DNA methylation data, Epiblast IAPeZ-int:  $n = 5,585; 5,579; 5,510; 5,440$ ; and  $5,210$ , Xecto IAPeZ-int:  $n = 5,576; 5,577; 5,498; 5,421; 5,367; 5,575; 5,529; 5,518; 5,500; 5,411$ ; and  $5,543$ . **c**, Scatterplot of E6.5 promoter DNA methylation and E8.5 expression differences in the Xecto lineage of L3MBTL2 mutant compared to wild-type embryos, as shown for the embryonic lineage in Fig. 2g. Differentially hypomethylated ( $\geq 0.1$ ) and derepressed genes ( $\geq 0.2$  fraction positive cells) in L3MBTL2-mutants (green) were strongly enriched in GO terms related to gametogenesis (green asterisks,  $P < 0.05$ ), in line with previous reports on ncPRC1.6 targets. These genes contain key members of the Piwi-interacting RNA biogenesis pathway, including the dead-box helicase *Ddx4* (VASA homologue) and *Mael* as well as other genes with known functions or expression during gametogenesis. Extraembryonic lineages naturally express certain gametogenesis-associated regulators, which may explain their

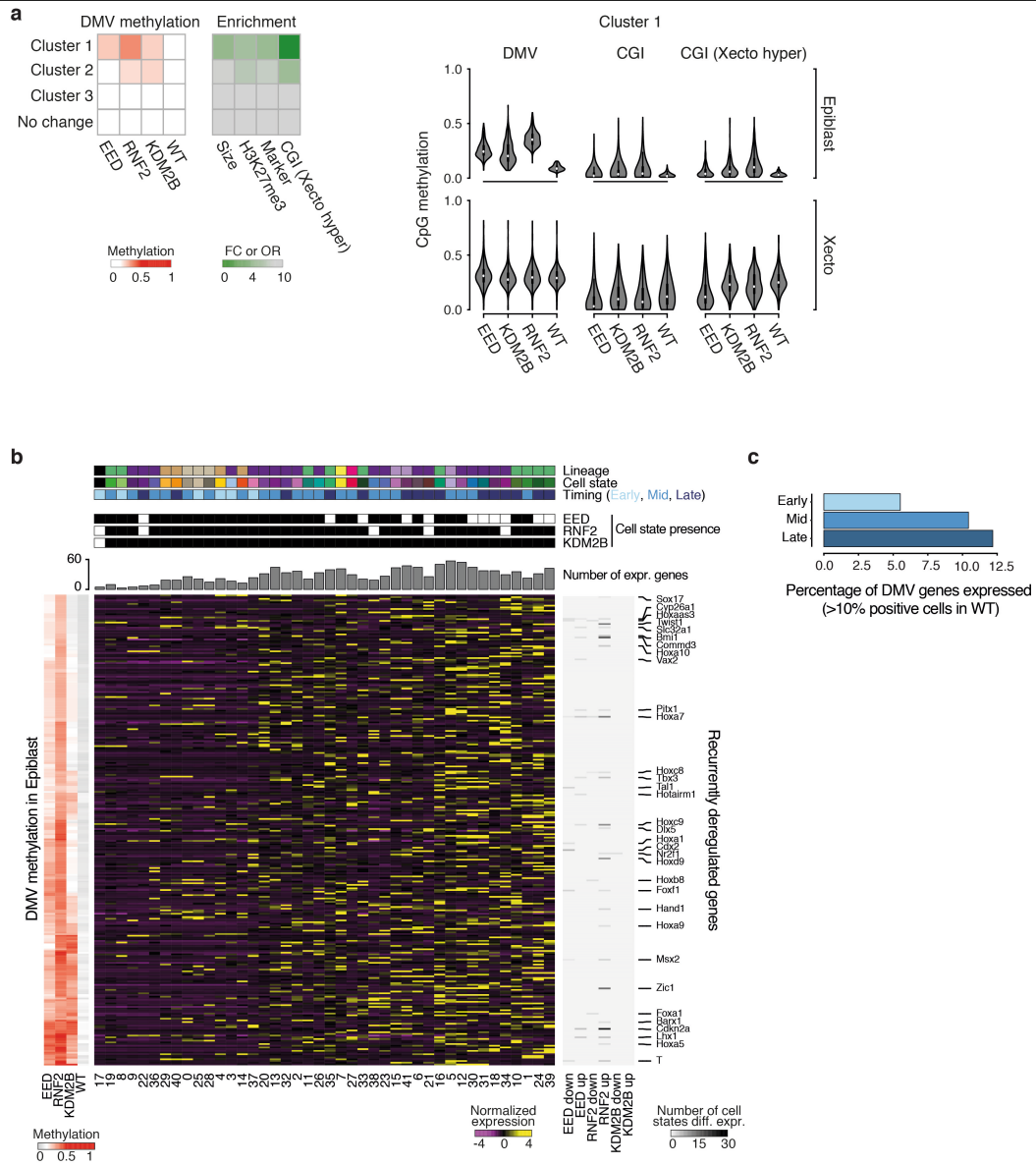
ability to proliferate in L3MBTL2-mutant embryos, whereas embryonic lineages arrest shortly after gastrulation onset. **d**, Genome browser tracks of WGBS methylation data for three aberrantly regulated loci in L3MBTL2-mutant embryos. The bidirectional genes *Lypd4* and *Dmrtc2* initiate from the same CGI, whereas *Tex101* does not have a CGI, but does have a higher than genomic average CpG density (see density track). These promoters are specifically hypomethylated in gametes and throughout preimplantation, followed by de novo methylation by E6.5 that increases over subsequent development. De novo methylation does not occur in L3MBTL2 mutants and corresponds with sharp increases in gene expression. Wild-type data from gametes, preimplantation embryos and late stage samples such as somatic tissues and the E14.5 placenta are from refs.<sup>63–66</sup>. **e**, Promoter DNA methylation (top) and E8.5 expression (bottom, shown as fraction of positive cells per embryo replicate) box plots of L3MBTL2-sensitive genes ( $n = 12$  genes taken from Fig. 2g, green). Many gametogenesis genes are regulated by ‘weak’ CGI-containing promoters that become methylated during development<sup>67</sup>. In line with this, the promoters of L3MBTL2 sensitive genes are hypomethylated in gametes and over preimplantation, but become de novo methylated shortly afterwards. Derepression is specific to L3MBTL2-mutant embryos, and does not occur in DNMT1 or DNMT3B mutants, in which methylation levels drop globally. Expression changes are also not substantial for RNF2 or G9A mutants although these regulators are also expected to participate in ncPRC1.6 complex-directed repression. A single outlier gene, *Ttr*, is expressed in all mutant and wild-type embryos, but is still upregulated in L3MBTL2 mutants. Additional data taken from previous studies<sup>63–65,68</sup>.





**Extended Data Fig. 6 | Effect of derepressed Polycomb group regulator targets.** **a**, Euclidean distances of PGC-assigned cells from our wild-type, EED-, RNF2- and KDM2B-mutant embryos to the mean marker gene expression of our PGC (state 27, magenta), their second closest (light grey) or the epiblast (state 17, dark grey) cell states. PGC-assigned mutant cells are transcriptionally distinct from the next closest or epiblast state, supporting our observation that this state is specifically overproduced in EED mutants. We include the epiblast state as it shares some master regulators with PGCs and because some cells of this state are still present in EED-mutant embryos. The differences between first and second closest or the epiblast state are all significant ( $P < 0.05$  for all tests, Wilcoxon test, two tailed). For each box plot, centre line denotes the median; edges denote the IQR; whiskers denote  $1.5 \times$  the IQR; outliers are individually plotted. Number of recovered PGC-state assigned cells is  $n = 290, 1,564, 250$  and  $44$  for wild-type, EED, RNF2 and KDM2B, respectively.  $P = 2.644257 \times 10^{-49}, 3.733801 \times 10^{-257}, 9.3103 \times 10^{-43}$  and  $1.136868 \times 10^{-13}$ , respectively, for PGC versus second closest state. **b**, Per cell ratio of chromosome X to autosome transcripts for PRC regulator mutant and wild-type cells, separated by sex. In our breeding system, X chromosomes are exclusively the C57BL/6J genotype, which impedes us from evaluating mono-versus biallelic transcription. However, these internally normalized measurements reveal increased transcription of chromosome X-linked genes within certain lineages of female embryos. The EED- mutant Xecto is most extreme and is strongly diminished at E8.5 (see Fig. 3d). Within EED-mutants, female-specific chromosome X deregulation is more subtly observed for the Xendo and embryonic lineages. This may indicate either higher redundancy

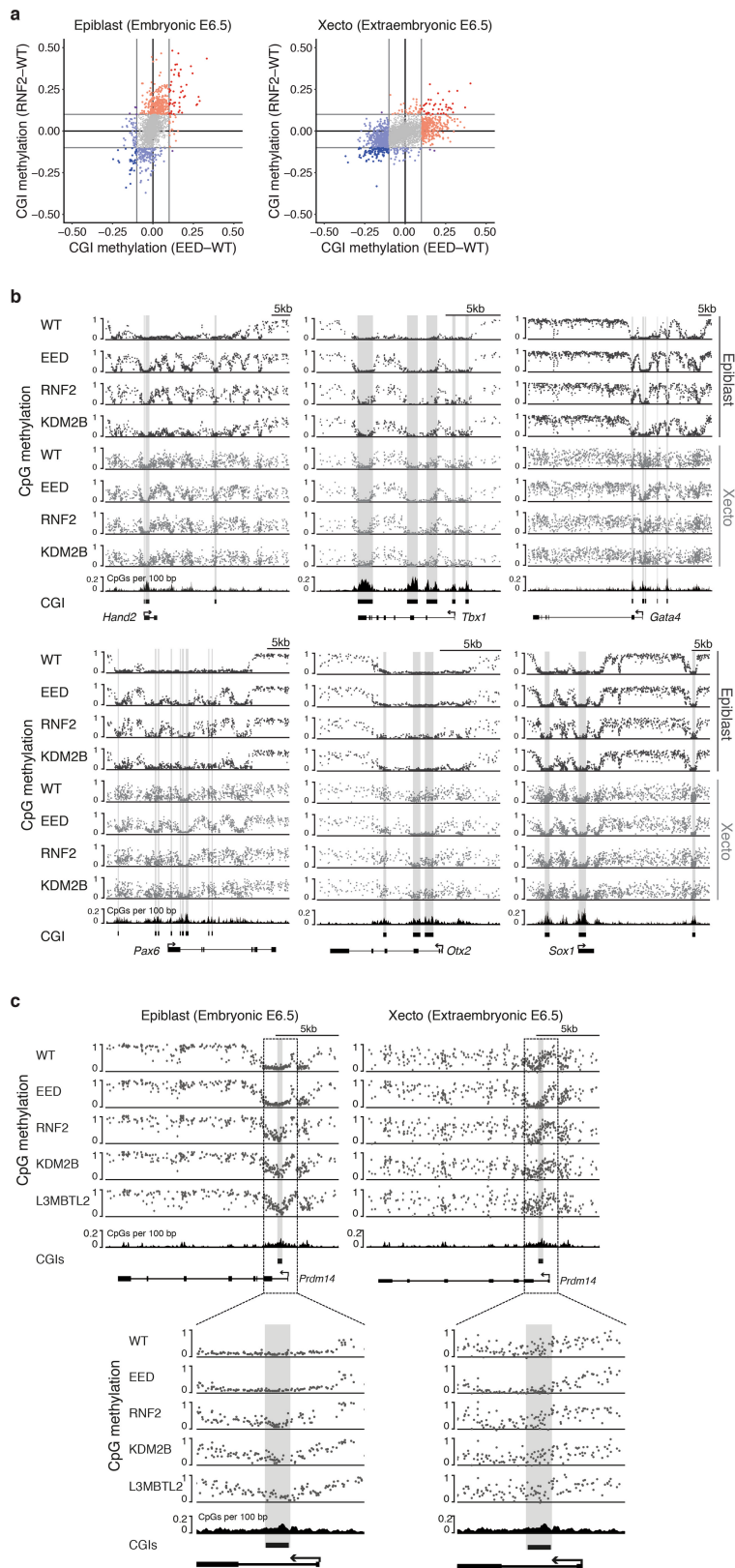
between PRC1 and PRC2 after the allocation of the trophectoderm or a lineage-specific failure to renormalize the transcriptional output of chromosome X within Xecto. In RNF2 mutants, the effects generally follow a similar trend but are more muted. Embryonic cells:  $n = 39,411; 37,887; 5,391; 9,233; 4,448; 5,248; 7,459$ ; and  $11,264$ , Xecto cells:  $n = 1,769; 3,685; 755; 1,372; 1,465; 1,220; 19$ ; and  $1,594$ , Xendo cells:  $n = 2,509; 3,518; 773; 1,419; 1,745; 1,463; 1,013$ ; and  $1,547$ . **c**, Reads spanning the sgRNA protospacer sequences confirms high efficiency disruption of *Eed* and *Cdkn2a* loci in single (CDKN2A) and double (EED+CDKN2A) sgRNA-injected embryos. Figure as in Extended Data Fig. 2d. **d**, Single cells from CDKN2A single and EED+CDKN2A double mutant embryos were assigned to one of our 42 wild-type cell states, projected onto our wild-type gastrulation UMAP and compared to E8.5 EED-mutant and wild-type embryos. Bar plot shows the median embryo composition. In general, our double mutant resembles EED, demonstrating that the derepression of the *Cdkn2a* locus in EED mutants is not responsible for the overall phenotype. The CDKN2A mutant is highly similar to wild-type embryos.  $n$  denotes number of cells. **e**, Correlation heatmap of average cell-state composition for our CDKN2A single and EED+CDKN2A double mutant embryos compared to wild-type stages and other core PRC component mutants, including a 24 h resolution EED-mutant time series described below (Fig. 4, Extended Data Fig. 10). CDKN2A-mutant embryos cluster with wild-type E8.0 and E8.5 embryos, whereas the Eed+CDKN2A double mutant clusters with wild-type E7.5 as well as our EED- and RNF2-mutant embryos.  $n = 42$  cell states, Pearson correlation.



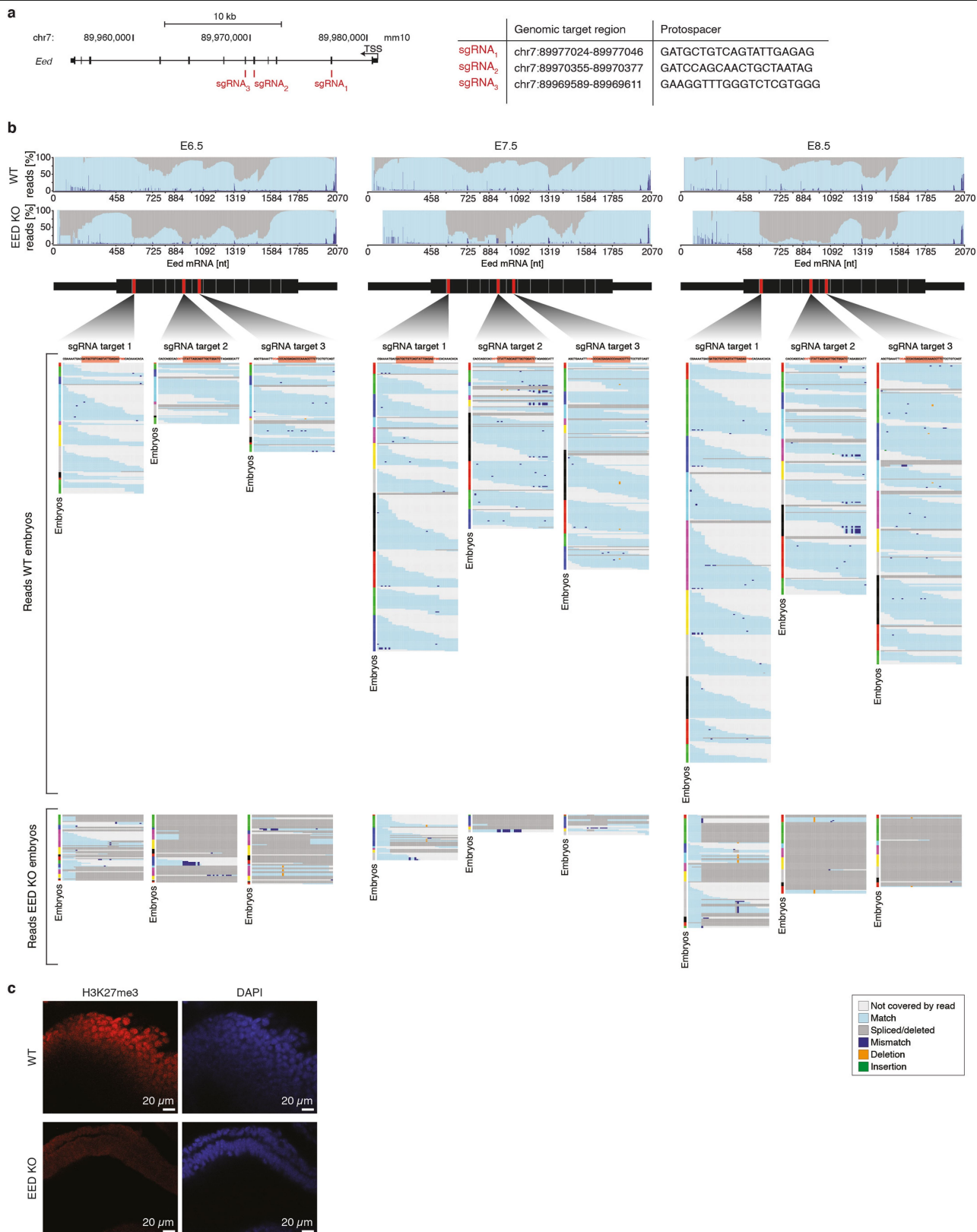
**Extended Data Fig. 7** | See next page for caption.

**Extended Data Fig. 7 | Molecular abnormalities of Polycomb group regulator mutants. a.** Left, large, multi-kilobase DMVs associated with developmental genes gain DNA methylation in PRC-mutant embryos. We clustered 8,972 DMVs that exist within the wild-type E6.5 epiblast according to their methylation in our PRC regulator mutants. A discrete set of 248 is specifically methylated within our EED, RNF2 and KDM2B mutants (cluster 1). Compared to the non-dynamic set (no change), these differentially methylated DMVs are enriched for marker genes as identified by this study, the modification H3K27me3, and for CGI hypermethylation within the Xecto lineage. They are also approximately 4.3 times larger than constitutively hypomethylated DMVs (mean span = 12.2 kb for dynamically methylated, 2.8 kb for 'no change'). Enrichment is calculated as an odds ratio (OR) or fold change (FC) compared to 'no change'. DMV methylation status across these regulator mutants is available as Supplementary Table 8.  $n = 248$  DMVs in cluster 1 versus  $n = 6,888$  for 'no change'. Right, DNA methylation violin plots of the 248 DMVs that gain CpG methylation within the E6.5 epiblast of our PRC mutants. 'DMV' measures methylation of all non-CGI CpGs within DMV boundaries for 'cluster 1', whereas 'CGI' measures those for all CGIs positioned within DMV boundaries ( $n = 529$ ). 'CGI (Xecto hyper)' measures the methylation for the subset of DMV-associated CGI that are specifically de novo methylated in wild-type Xecto ( $n = 191$ ). In epiblast, DMV methylation is highest for RNF2 mutants and lower for the same regions in EED- mutants. By contrast, KDM2B mutation shows substantial heterogeneity, with >55% of DMVs showing lower methylation compared to EED. The DMVs that gain DNA methylation in the epiblast of PRC regulator mutants are generally naturally de novo methylated in the Xecto

(including methylation of CGIs). Here, the CGIs in the EED mutant Xecto pose an exception as they show a specific loss of methylation. **b.** Heat maps showing the wild-type expression status of 303 genes contained within differentially methylated DMVs. In PRC regulator mutants, the loss of epigenetic repression may prime genes for induction. However, there is no clear correlation between the genes located within differentially methylated DMVs and the lineages that are ultimately overproduced. Although the exact relationship remains unclear, our DNA methylation analysis indicates that aspects of the PRC mutant phenotype begin to manifest within the pre-gastrula embryo, leading to similar epigenetic changes within the promoters of master regulators associated with all three germ layers. Left, mean DMV methylation for each mutant and wild-type embryo as calculated in **a** (with CGI CpGs excluded). Middle, row-normalized expression of DMV-associated genes across our 42 wild-type states. Right, fraction of mutant cell states where a given gene is recurrently up- or downregulated. DMVs (rows) are clustered by methylation status and cell states (columns) by DMV-associated gene expression. Top, identity and presence of cell states in E8.5 regulator mutants (black, present; white, absent). States are designated as early, middle or late (most prevalent in wild-type embryos at E6.5 to E7.0, E7.5, or E8.0 to E8.5, respectively). The cumulative number of DMV-associated genes expressed within each state in wild-type embryos is also provided. **c.** The percentage of DMV-associated genes that are expressed in our 42 wild-type states collapsed into early, middle or late based upon when states emerge (E6.5–E7.0, E7.5 or E8.0–E8.5). In general, differentially methylated DMV-associated genes are normally expressed in the middle or late periods of our time series.



**Extended Data Fig. 8 | PRC1 and 2 converge to block non-CGI hypermethylation within DMVs. a**, Scatterplots of the difference between EED- and RNF2-mutant CGI methylation compared to wild-type embryos for E6.5 Epiblast (left) and Xecto (right), respectively. Although overall EED and RNF2 mutants share a similar DNA methylation landscape within the epiblast, we identify some regions in which the RNF2 mutant is differentially methylated and the EED mutant more closely resembles wild-type embryos. EED-mutant embryos show a more substantial loss of CGI methylation specifically within Xecto, whereas RNF2-mutant embryos show increased levels in epiblast that is primarily due to changes in flanking areas (see Fig. 3h). **b**, Genome browser WGBS methylation tracks for representative loci as they are regulated within the E6.5 Epiblast (upper, dark grey) and Xecto (lower, light grey) in wild-type, EED-, RNF2- or KDM2B-mutant embryos. Genes include master regulators from all three germ layers: *Hand2* and *Tbx1*, mesoderm; *Gata4*, endoderm; *Pax6*, *Otx2* and *Sox1*, neural ectoderm. CGI and local CpG density tracks are provided below. Promoter regions of these developmental genes are generally preserved as extended multi-kilobase hypomethylated domains. However, in EED and RNF2 mutants, non-CGI CpGs become hypermethylated whereas the CGIs remain unmethylated. This trend is also observed for KDM2B-mutant embryos but to a substantially lesser degree. Changes to promoter methylation status appear to be independent of the gene's association with particular lineages or expression status at E6.5: mesodermal, endodermal and ectodermal regulators are affected. These regions are also extensively de novo methylated within the Xecto lineage during normal development, including at the CGIs themselves. Within the Xecto, EED mutation specifically causes loss of CGI methylation. Notably, EED mutation-specific methylation changes within the Xecto are also found at loci that do not acquire methylation changes in Epiblast, such as for *Otx2*. **c**, Genome browser WGBS tracks for the *Prdm14* locus in the epiblast (left) and Xecto (right) of wild-type, EED-, RNF2-, KDM2B- and L3MBTL2-mutant embryos. Although this region is retained as a hypomethylated DMV in the wild-type and EED-mutant epiblast, it is specifically methylated in PRC1 subunit mutations. In Xecto, the *Prdm14* promoter is naturally methylated but specifically unmethylated in EED-mutant embryos.



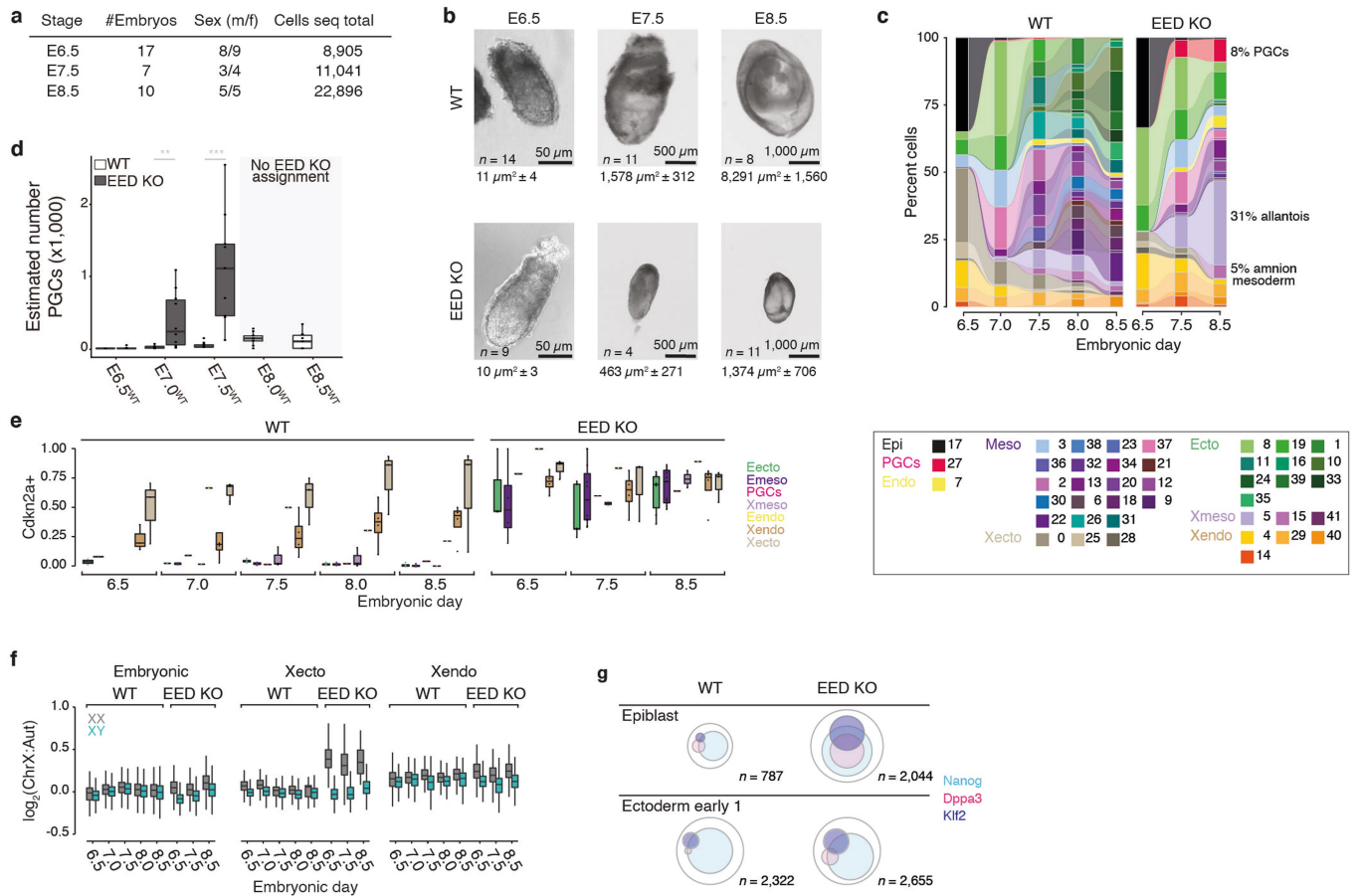
**Extended Data Fig. 9** | See next page for caption.



# Article

**Extended Data Fig. 9 | Efficient Cas9-mediated zygotic disruption of the *Eed* locus across an expanded time series.** **a**, Description of our zygotic perturbation strategy for the *Eed* locus. Three sgRNAs were designed to balance high efficiency cutting, off target potential, and coverage across the first half of the coding sequence (Methods). Then, selected sgRNAs were injected as a pool to provide a high likelihood of functionally disruptive mutations. **b**, Comprehensive analysis of scRNA-seq reads aligned to the *Eed* transcript from E6.5, E7.5 and E8.5 wild-type and EED-mutant embryo data. Top, composite plot showing the fraction of reads that map continuously (light blue) or discontinuously due to spliced or deleted sequences (dark grey) to the *Eed* mRNA annotation. Substantially more reads map discontinuously in mutant compared to wild-type embryos, reflecting alterations in the transcripts as a result of Cas9-mediated genetic disruption. Middle, read-level analysis of our E6.5, E7.5 and E8.5 wild-type embryo data. Positions of the three sgRNA target sequences (red) within the *Eed* mRNA are shown. The sgRNA target regions are magnified with aligned scRNA-seq reads from each embryo shown below (colour bar to the left of each read stack). Each row of the read stack represents the mapped sequence of an scRNA-seq read. Reads are colour coded as exactly matched (light blue) or spanning the deleted/spliced out

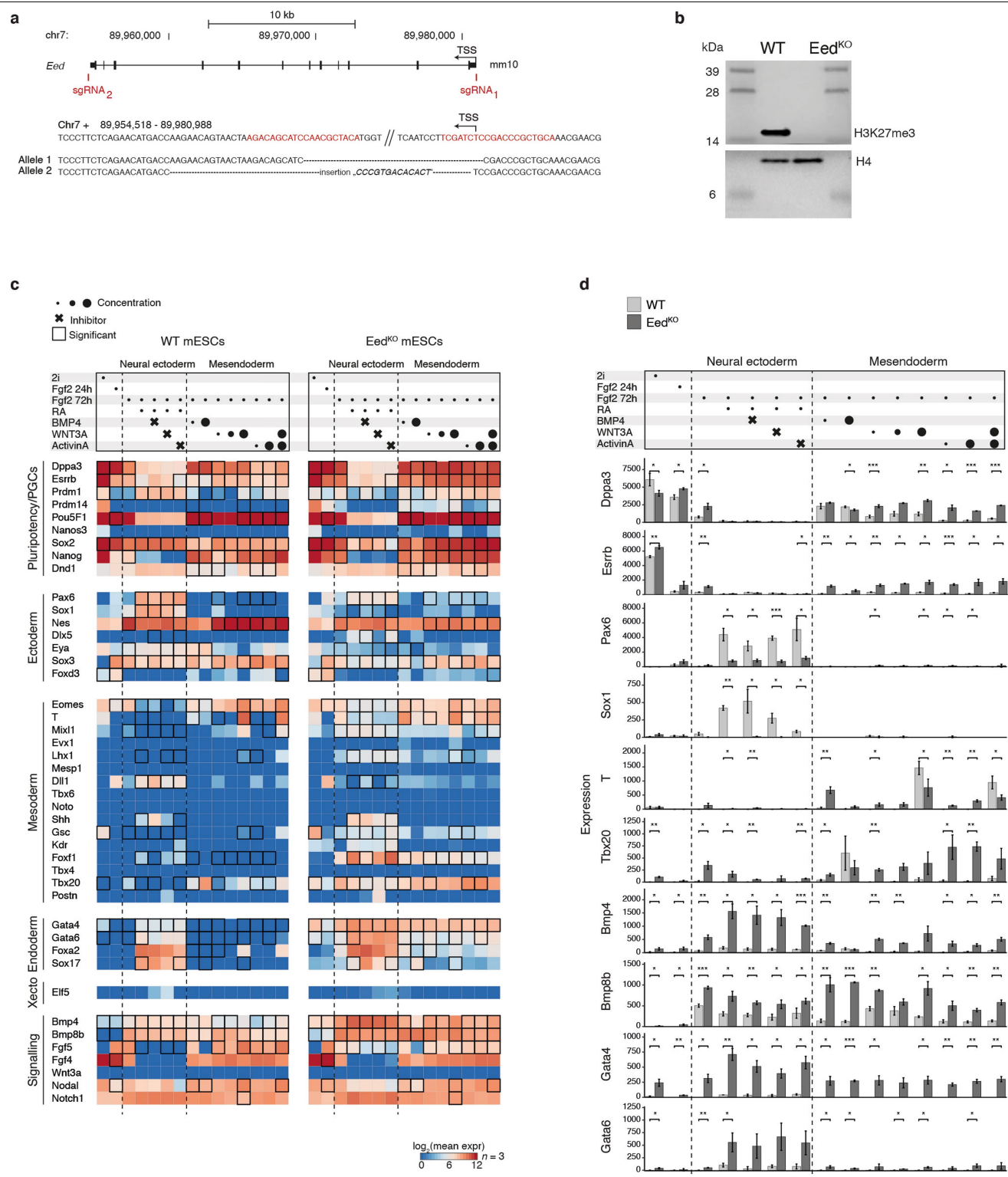
target site (dark grey). Light grey indicates no data for this read at a given position (read ends). Even though the scRNA-seq strategy preferentially profiles the 3' end of transcripts, many reads can be found that span sgRNA target regions in data from wild-type embryos, with a subset covering the entire target site without mismatches, insertions or deletions. Bottom, read-level analysis for our EED-mutant embryos from each time point. Compared to wild-type embryos, a much lower number of reads from the EED-mutant data match the target sites, probably a result of nonsense mediated decay or improper transcript processing. Moreover, aligned reads are imperfect, either spanning a deleted/spliced out target site (dark grey) or mapping with mismatches (dark blue), local deletions (green) or insertions (orange indicates the nucleotide to the right of an insertion). **c**, Representative immunofluorescence staining of H3K27me3 in wild-type and EED-mutant embryos. Single z-stack displaying an anterior region of size-matched wild-type (E7.5) and EED-mutant (E8.5) embryos (H3K27me3, red; nuclei stained by DAPI, blue). The nuclear signal for H3K27me3 is readily detectable in wild-type but absent in the EED mutant. Two independent experiments were conducted with similar results.



#### Extended Data Fig. 10 | Developmental roles of PRC2 during gastrulation.

**a**, Our scRNA-seq profiled EED-mutant series isolated at E6.5, E7.5 and E8.5. See Supplementary Tables 1 and 5 for information on sex-typing and cell-state composition of individual embryos. **b**, Representative wild-type and EED-mutant embryos at gestational days E6.5, E7.5 and E8.5, with size information (image area occluded by an embryo in  $\mu$ m<sup>2</sup>,  $n$  denotes embryos imaged, all experiments had been replicated at least once, with similar results). EED-mutant embryos initially appear similar to wild-type embryos in size and morphology, but become substantially smaller and more variable in morphology, consistent with previous reports using transgenic models<sup>29,69,70</sup>. The initial lack of obvious abnormalities at E6.5 may indicate a later biological requirement or mitigating effects of maternally loaded PRC2, which is detectable until E3.5 (ref.<sup>71</sup>). Complete *Eed* disruption is supported by the consistency of the resulting phenotype, as *Eed*<sup>-/-</sup> mice are viable and appear phenotypically normal during this period<sup>72</sup>. Wild-type embryos shown here are from natural matings isolated at the same gestational age. **c**, Connected bar plots of median cell-state composition across developmental stages for wild-type and EED-mutant embryos, respectively. Wild-type embryos rapidly increase in complexity, whereas EED mutants advance more slowly and become substantially biased towards PGCs and extraembryonic mesoderm. The lack of more advanced neural ectoderm (dark greens) and embryonic mesoderm (purples) may be due to developmental delay or the abnormality of precursor states. Outermost extraembryonic tissues (Xendo, Xecto) can be technically variable during isolation and their proportions should be taken with caution. **d**, Absolute PGC numbers estimated for individual embryos (dots) show that EED mutants overproduce PGCs beyond what is observed for wild-type embryos over gastrulation. EED-mutant embryos are presented after

accounting for their developmental delay (that is, PGC numbers of E8.5-isolated mutant embryos that match developmental stage E7.5 are displayed for E7.5). Wilcoxon test, two-sided,  $P$  values: 0.322 (E6.5), 0.008 (E7.5), 0.0003 (E8.5); \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ;  $n = 10, 15, 9, 10, 11, 9, 10$  and 10 embryos (left to right). **e**, Fraction of cells positive for the *Cdkn2a* transcript in recovered cell states (dots), shown per lineage across our wild-type and EED-mutant time series. *Cdkn2a* is broadly derepressed across lineages in EED-mutant embryos from E6.5 onward.  $n = 10; 19; 1; 3; 1; 4$  and 3 cell states. **f**, Ratio of X chromosomal to autosomal transcripts for all male and female cells isolated across our wild-type and EED mutant time series, separated according to preimplantation lineage (Embryonic, Xendo and Xecto). Derepression of chromosome X-linked genes happens as early as E6.5 in EED-mutant females. Xecto also becomes substantially underproduced in these same embryos over time ( $n = 428, 295$  and 19 female EED-mutant cells from E6.5–E8.5). Xendo and embryonic lineages show increased chromosome X transcription, but not to the degree that is observed in Xecto. Embryonic:  $n = 581; 424; 7,119; 3,714; 4,188; 18,730; 9,519; 11,551; 18,004; 3,468; 3,541; 2,410; 6,041; 2,263; 7,459; 11,264$ ; Xecto:  $n = 325; 355; 830; 360; 120; 2,371; 485; 552; 9; 47; 428; 649; 295; 363; 19; 1,594$ ; Xendo:  $n = 295; 300; 780; 415; 134; 1,879; 533; 704; 767; 220; 930; 947; 1,531; 548; 1,013; 1,547$  cells. **g**, Venn diagrams of epiblast or early ectoderm 1 cells (states 17 and 8, respectively) that are positive for key transcription factors associated with germline formation<sup>73,74</sup>. These transcripts are more abundant in EED-mutant embryos and more frequently present within the same cells, suggesting a PGC-supporting subnetwork within early embryonic cell states, possibly due to insufficient or unstable silencing before gastrulation.  $N$  = total cells per state.



Extended Data Fig. 11 | See next page for caption.

**Extended Data Fig. 11 | Differentiation of EED-knockout mES cells recapitulates many features of the in vivo mutant phenotype.** **a**, Generation of a homozygous EED-knockout mES cell line. The *Eed* gene was deleted in V6.5 mES cells by simultaneous Cas9-targeting of flanking sequences (red) to create a >20-kb deletion. Sanger sequencing confirmed complete deletion by non-homologous end joining for both alleles (sequences aligned to chromosomal sequence above with dashes for missing nucleotides). **b**, Western blot of histone extracts for H3K27me3 confirms depletion of H3K27 trimethylation and homozygous *Eed* deletion. Histone 4 served as loading control. **c**, Transcript counts of 44 genes associated with pluripotency, early germ layers, and the germline<sup>75</sup> over directed differentiation experiments. Wild-type and EED-knockout mES cells were maintained in conditions supporting a naive, inner cell mass-like state (2i), then subjected to low concentrations of bFGF for 24 h followed by culture in neural ectodermal or mesendodermal inducers for an additional 48 h. Top, the combination of signalling molecules and/or inhibitors used. Concentration ranges are indicated by circle diameter and small molecule inhibitors by crosses. Inhibitors were included to promote neural ectodermal gene induction by counteracting competing pathways. 12 ng ml<sup>-1</sup> bFGF; 0.25 μM RA; 5 and 500 ng ml<sup>-1</sup> BMP4; 10, 100, 1,000 ng ml<sup>-1</sup> WNT3A; 10 and 1,000 ng ml<sup>-1</sup> Activin A; 0.5 μM BMP4 pathway inhibitor LDN-193189; 3.3 μM Wnt pathway inhibitor XAV939; 10 μM TGFβ-activin-nodal pathway inhibitor SB431542. Bottom, heat map of log<sub>2</sub>-transformed molecule counts (red being highly expressed) for wild-type and EED-knockout mES cells, separately. Black tile frames indicate significant changes between wild type and knockout. During differentiation, many

pluripotency factors associated with the germline remain expressed in EED-knockout mES cells, especially within mesendodermal supporting conditions. Many mesodermal genes are also induced in EED-knockouts in neural ectodermal supporting conditions. Retinoic acid treatment directs a fraction of wild-type mES cells to an extraembryonic endodermal fate<sup>76</sup>. This appears to be favoured in EED-knockout mES cells, in which many Xendo-associated genes are particularly sensitive to retinoic acid. Finally, many regulators of the endoderm, early mesendoderm and extraembryonic mesoderm, such as *Gata4*, *Tbx20* and *Bmp4*, are broadly expressed in EED-knockout mES cells already in 2i. *n* = 3 experimental replicates profiled with the PlexSet assay on the NanoString nCounter SPRINT instrument. Significance (two-sided *t*-test) was tested for genes in conditions where at least one of the two cell lines produces a minimum average of 50 counts above background. **d**, Bar plots of normalized, absolute Nanostring molecule counts for the genes and conditions presented as fold change in Fig. 4f. Mesodermal genes exhibit some responsiveness to exogenous signals and can be induced to different degrees under supportive conditions (BMP4 and WNT3A). However, many are also more highly expressed without these stimuli, particularly genes associated with posterior, extraembryonic mesodermal fates. Raw counts were normalized based on the expression of four housekeeping genes and a conservative background subtraction to account for the potential of false negative signals. \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001, two-sided *t*-test. Bars show the mean of *n* = 3 experimental replicates and error bars represent s.d. *P* values are provided in the Source Data file Nanostring\_pval.tsv.

## Reporting Summary

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

### Statistics

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

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

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

### Software and code

Policy information about [availability of computer code](#)

Data collection	10X Cell Ranger (v3.0), STAR(v2.5.3a), SNPSplit(v0.3.2), BSmap (v2.9.0), MOABS (v1.3.2), ZEN2 (blue edition) V1.0en.3.1, Metamorph software (7.8.2.0)
Data analysis	R (v3.5.1), R packages: Seurat (v3.0.0), Complex Heatmap (v1.99.5), Gviz (v1.24.0), ggplot2 (v3.1.0); python (v3.7.3) and python modules: scanpy (v1.4.3), velocyto python tool (v0.1.18), ImageJ (1.52p); private code at <a href="https://github.com/HeleneKretzmer/EpigeneticRegulators_MouseGastrulation">https://github.com/HeleneKretzmer/EpigeneticRegulators_MouseGastrulation</a>

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

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

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

All datasets have been deposited in the Gene Expression Omnibus and are accessible under GSE137337. Source data behind Figures 1a, b, 2, 3a, b, d-f, h, i, 4a, c-f and Extended Data Figures 1b, c, e-i, 2b-f, 3, 4b-f, 5a-c, e, 6, 7, 8a, 9b, 10b-g, 11c, d are available at <https://oc-molgen.gnz.mpg.de/owncloud/s/F8g3y5F79JZRyof>. Previously published data used in this study include H3K27me ChIPseq data (GSE98149), WGBS data for sperm and oocyte (GSE112320), preimplantation samples, including 8 cell stage embryos and the ICM and trophectoderm (TE) of the E3.5 blastocyst (GSE84236), and late stage samples including an average of somatic tissues and the E14.5 placenta (GSE42836).



## Field-specific reporting

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

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

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

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

Sample size	No statistical methods were used to predetermine sample size. The number of embryos reported is the maximal number of embryos we could recover and reasonably submit for scRNA-seq. To improve the coverage of E8.0 and E8.5 stage embryos, we ran 2x lanes of the Chromium 10x system to double the representation of cells from these embryos (representing ~2% total coverage). Our methodology is designed to ensure robust testing of given genetic perturbations by providing 7-17 embryo replicates per scRNA-seq run and thousands to tens of thousands of individual single cell transcriptomes. We provide the full variation in cell states of every WT and E8.5 KO sample collected in terms of cell state in Extended Data Figs. 1g and 2f. For imaging of embryos to estimate size or detect Prdm14 reporter activity, sample sizes are described in the associated figure legends.
Data exclusions	<p>We included all embryos collected for each of our scRNA-seq data sets and identified by genotyping. Experimental embryos were isolated from the deciduae of pseudopregnant females and included with minimal or no selection criteria: either every viable embryo from a given experiment was collected or isolation was stopped after recovering a sufficient number of embryos. During collection, embryos may be excluded only if they were randomly damaged (with embryonic material lost) or if the decidua indicated embryo rejection (characterized by a collapsed embryo and extensive maternal blood).</p> <p>Single cell transcriptomes called by Cell Ranger v3 were excluded if they did not include <math>\geq 1,000</math> CAST/EIJ SNPs or if they failed our genotyping criteria. Autosomal SNP ratios from each cell were regenerated after down sampling the data to 20% and iterated 100 times. Those that showed an unstable assignment, either due to lower coverage or because they represent doublets, were excluded from further analysis.</p> <p>Cells assigned to cell states that were discovered with less than 30 cells (except PGC state) were excluded in the KO cohort due to less reliable expression estimation.</p>
Replication	All results obtained and reported in this study were reproducible across the embryo replicates examined. We demonstrate reproducibility by examining between 7 and 17 embryos for each WT or experimental condition, often isolated from multiple pseudopregnant females transferred with different batches of WT or mutant blastocysts at the same time. Each embryo replicate represents an individual experiment downstream of a unique Cas9-generated genetic lesion. The reproducibility of cell state assignments is shown for every WT and E8.5 KO sample in Extended Data Figs. 1g and 2f.
Randomization	Embryos for every experiment were collected without a preconceived selection strategy or prioritization by morphology. When more embryos than can reasonably be covered by a single scRNA-seq run were isolated, those subjected to pooling, trypsinization, and scRNA-seq were selected randomly.
Blinding	Blinding was not relevant for the scRNA-seq strategies used in this study. However, our analytical pipeline followed uniform criteria applied to all samples, allowing us to analyze our data in an unbiased manner.

## Reporting for specific materials, systems and methods

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

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

### Antibodies

Antibodies used	Tri-methyl-histone H3 antibody (Cell, Signaling, C36B11, at 1:500 dilution); Histone H3 Lysine 27 tri-methylation (Abcam ab6002, at 1:200 dilution); histone H4 antibody (Millipore, 07-108, at 1:1,000 dilution)
-----------------	---

## Validation

The tri-methyl-histone H3 antibodies were used on wild-type mouse embryos or mouse embryonic stem cells and validated in knock-out embryos or mouse embryonic stem cells.

## Eukaryotic cell lines

Policy information about [cell lines](#)

## Cell line source(s)

The wild-type V6.5 mouse embryonic stem cell line was provided by the lab of Konrad Hochedlinger.

## Authentication

Wild-type V6.5 mouse embryonic stem cells were authenticated by Nanostring expression profiling of mouse pluripotency markers.

## Mycoplasma contamination

Cell lines tested negatively for mycoplasma.

Commonly misidentified lines  
(See [ICLAC](#) register)

None used.

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

## Laboratory animals

Oocytes were isolated from B6D2F1 strain female mice (age 6 to 8 weeks, Jackson Labs), sperm was isolated from B6/CAST F1 mice (>2months of age) which were generated by breeding C57BL/6J strain female mice with CAST/EiJ strain males. Blastocysts were transferred into CD-1 strain female mice (25-35g, Charles River) which had been mated with Vasectomized SW strain males (Taconic).

## Wild animals

The study did not involve wild animals.

## Field-collected samples

The study did not involve field-collected samples.

## Ethics oversight

All procedures follow strict animal welfare guidelines as approved by Harvard University IACUC protocol (#28-21) and by the Max Planck Institute for Molecular Genetics (G0247/13-SGr1).

Note that full information on the approval of the study protocol must also be provided in the manuscript.