



Published in final edited form as:

*Nature*. 2020 August ; 584(7819): 102–108. doi:10.1038/s41586-020-2552-x.

## Epigenetic regulator function through mouse gastrulation

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

<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, Massachusetts, USA

<sup>3</sup>Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts, USA

<sup>4</sup>Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts, USA

<sup>5</sup>Department of Physics, Washington University in St. Louis, St. Louis, Missouri

### Summary paragraph:

During ontogeny, proliferating cells become restricted in their fate through the combined action of cell-type specific transcription factors and ubiquitous epigenetic machinery, which recognize universally available histone residues or nucleotides but are nonetheless deployed in a highly context-dependent manner<sup>1,2</sup>. The molecular functions of these regulators are generally well understood, but assigning direct developmental roles is hampered by complex mutant phenotypes that often emerge following gastrulation<sup>3,4</sup>. Recently, single-cell RNA sequencing (scRNA-seq) and analytical approaches have explored this highly conserved process across numerous model organisms<sup>5–8</sup>, including mouse<sup>9–18</sup>. To elaborate on these strategies, we investigated a panel of ten essential regulators using a combined zygotic perturbation, scRNA-seq platform where many mutant embryos can be assayed simultaneously to recover robust transcriptional and morphological information. Deeper analysis of central Polycomb Repressive Complex (PRC) 1 and 2 members indicate substantial cooperativity, but distinguishes a PRC2-dominant role in restricting the germline that emerges from gross molecular changes within the initial conceptus.

---

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

Correspondence: [meissner@molgen.mpg.de](mailto:meissner@molgen.mpg.de).

Author contributions

SG, ZDS and AM designed and conceived the study. SG, HK, ZDS and AM prepared the manuscript with the assistance of the other authors. ZDS performed the mouse experiments for scRNA-seq and WGBS. SG and ZDS performed all scRNA-seq experiments. LW and ASK performed IVFs and electroporations for reporter experiments. ASK conducted the staining and imaging of the embryos supervised by SG. SG and ZDS made the scRNA-seq libraries with raw data processing by SK, BT and HK. Downstream data processing was performed by HK. Analysis of repetitive elements from scRNA-seq data was performed by SH. Data analysis was performed by SG, HK and ZDS and assisted by SM. SG conducted *in vitro* experiments and analyzed Nanostring data. AM supervised the work.

\*These authors contributed equally to this work.

Competing interests

The authors declare no competing interests.

We believe our experimental framework will eventually allow for a fully quantitative view of how cellular diversity emerges using an identical genetic template and from a single totipotent cell.

---

## Single-cell view of mouse gastrulation

Gastrulation represents a period 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 scRNA-seq data from pools of sibling embryos with a B6/CAST F1 father, allowing us to computationally distinguish single replicates by their randomly inherited CAST genotype, and sex according to ChrX- and Y-linked gene expression (Extended Data Fig. 1a–c, Supplementary Tables 1 and 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 (WT) compendium comprises 88,779 high-quality single-cell transcriptomes from 50 embryos (median of 16,898 transcripts and 3,854 genes per cell and ~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 and 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 WT reference established, we proceeded to zygotically disrupt one of several epigenetic regulators, 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. Dnmt1, G9a, and Dnmt3b mutations are lethal after gastrulation, while 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, while Kdm2b and L3mbtl2 are noncanonical PRC1.1 and PRC1.6 complex subunits, respectively<sup>20</sup>. Finally, we included the Histone-3-Lysine-4 (H3K4)

methyltransferases Kmt2a and Kmt2b, Trithorax group orthologs that promote developmental gene expression in opposition to Polycomb<sup>21</sup>.

Unlike many transcription factors, these genes are expressed across the majority of lineages and cell states, although the *de novo* Dnmts are particularly abundant prior to 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 inspection of 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 (by developmental progression and gross morphology) and molecularly (by transcriptional state). To examine developmental progression, we assigned mutant (hereafter “KO”) cells by marker expression to their closest WT state and examined embryo composition across replicates (Extended Data Fig. 2e, f). Certain regulator KOs 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 levels of the overall lineage hierarchy, suggesting developmental delays. We therefore innovated a stage-matching metric that considers and weighs the types of states and their relative proportions within an embryo compared to our WT data (Extended Data Fig. 3b–d, Supplementary Table 5).

Our approach confirmed many historical observations, including moderately increased severity for mutations of Dnmt1 compared to Dnmt3a or 3b, as well as of Kmt2b compared to Kmt2a, consistent with its primary role in orchestrating early differentiation<sup>22</sup>. Of our E8.5-isolated KO embryos, L3mbtl2 showed the greatest delay, arresting shortly after the onset of gastrulation at ~E7.0, followed by Eed and Rnf2, which gastrulate but largely fail to progress beyond ~E7.5 (Fig. 2b). Notably, many KOs affect development beyond progression or growth. For example, Eed and Rnf2 clearly gastrulate, but fail 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). In contrast, L3mbtl2 KO 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 2 converge to a common gene set

Assigning each cell from our mutant embryos to pre-defined WT states allowed us to measure within-state expression changes in addition to over- or underproduction of certain lineages. To compare our ten KOs, we initially identified differentially expressed genes for each mutant cell state against WT as “up” or “down” regulated (Supplementary Table 6). We then calculated the fraction of cell states where 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 also be affected by lower overall embryonic complexity. We were surprised to see that the ncPRC1 subunit Kdm2b clusters with Eed and Rnf2, even though Kdm2b KO embryos generally produce more mature cell types (Fig. 2c, Extended Data Fig. 3a, e, f).

All three Polycomb-associated mutants also converge towards functional ontologies associated with developmental processes and cell cycle regulation (Extended Data Fig. 3f). For example, the tumor suppressor *Cdkn2a*, a known PRC-regulated locus<sup>25</sup>, is constitutively targeted by both PRC1 and 2. In contrast, L3mbtl2 KO 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 KO data provide the most compelling results in terms of transcriptional and morphological defects to the gastrulation process itself. However, Dnmt1 and G9a KO 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, we generated whole genome bisulfite sequencing (WGBS) data of E6.5 epiblast and extraembryonic ectoderm (Xecto) for each regulator KO, as these tissues represent the latest homogenous progenitors prior to the actions of gastrulation (Extended Data Fig. 4). We see clear global DNA methylation differences for Dnmt1 KO, as well as more subtly for Dnmt3b and Dnmt3a (Extended Data Fig. 4a–d).

We also 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 KO, which both proceed through gastrulation with some developmental delay, but not for our Kmt2a KO, which advances normally through E8.5 (Fig. 2d, Extended Data Fig. 4d, Ref<sup>26</sup>). Kmt2b appears to protect a larger number of CGIs than Kdm2b and also operates within Xecto, suggesting either a broader utility or earlier preimplantation activity (Fig. 2d). Kdm2b and Kmt2b protected promoter CGIs are also ~2.5-fold enriched for H3K27me3-based regulation in the epiblast, though in general their associated genes are lowly expressed in our data, suggesting that their influence may be too subtle to pinpoint with the current scRNA-seq strategy (Extended Data Fig. 4e). In 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 KOs, which otherwise exhibited limited gene expression differences (Extended Data Fig. 5a). Here, the ERVK subfamily of LTRs shows strongly coupled demethylation and transcription within Dnmt1 KOs, specifically methylation-sensitive Intracisternal A-type Particles (IAPs)<sup>27</sup> (Fig. 2e, Extended Data Fig. 5b). This may explain Dnmt1 KO embryos' impeded progression and death within ~1–2 days following E8.5. In contrast, continued IAP

repression in *Dnmt3b* or *G9a* KOs suggests these embryos maintain a sufficient threshold to preserve epigenetic silencing (Fig. 2e).

Finally, we re-examined our *L3mbtl2* KOs to better understand their severe phenotype, including near total embryonic arrest and continued extraembryonic growth (Fig. 2f). We analyzed cells from either the embryonic or Xecto lineage separately for gross expression changes as they relate to promoter DNA methylation. We identify 13 genes that are both highly over-expressed and show lower than expected promoter methylation (Fig. 2g, Extended Data Fig. 5c, d), including several previously reported ncPRC1.6 targets associated with gametogenesis<sup>28</sup>. Examined over early development, we find that *L3mbtl2*-sensitive genes are specifically unmethylated in both gametes and become methylated shortly following gastrulation (Extended Data Fig. 5e). Thus, *L3mbtl2*'s primary developmental function appears to be the active silencing of select gamete-specific genes, whose aberrant and exogenous expression 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* and *Rnf2* KOs overproduce posterior-proximal structures, such as the allantois, without advancing the embryo proper<sup>29,30</sup> (Fig. 3a). Notably, *Eed* KO embryos also substantially overproduce PGC state cells, a result we confirmed by generating KO embryos carrying a *Prdm14* promoter-driven reporter<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* KO females consistently fail to maintain the Xecto lineage and substantially derepress ChrX-linked genes, while *Rnf2* KOs are more subtly deviated and *Kdm2b* KOs appear 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 ChrX transcript ratios within embryonic lineages, though *Eed* participates in random XCI as well (Extended Data Fig. 6b).

In contrast to core PRC1 and 2 components, *Kdm2b* largely appears to play a secondary role on the same overall gene set. For example, *Cdkn2a* is upregulated in all embryonic lineages for all three KOs, but to a substantially higher degree in *Eed* and *Rnf2* KO (Fig. 3f). Despite a canonical role as a tumor suppressor, *Cdkn2a* expression does not explain the overall gastrulation defect observed in *Eed* KO embryos, as co-injecting sgRNAs to *Cdkn2a* and *Eed* does not produce appreciable differences compared to *Eed* KO alone (Extended Data Fig. 6c–e).

In our three PRC KOs, differentially methylated CpGs collect within multi-kb territories, termed DNA methylation valleys (DMVs)<sup>34</sup>, that are normally maintained in a completely unmethylated state. Upon PRC disruption, a subset of DMVs become hypermethylated, though internal CGIs remain protected (Fig. 3g, h, Extended Data Fig. 7a). These PRC-

sensitive DMVs are enriched for H3K27me3 and cover substantially larger genetic territories. They also preferentially harbor promoters of lineage-specific marker genes, although we see no straightforward correlation between epigenetically deregulated genes and the overall morphology of Eed KO embryos (Extended Data Fig. 7b, c, Supplementary Table 8). While we see the same overall change in methylation pattern in Eed, Rnf2, and Kdm2b KO epiblast, the net levels are lower for Kdm2b (Fig. 3g, h, Extended Data Fig. 8a, b). In all examined cases, Kdm2b shows some deregulation consistent with core PRC subunits, but does not appear to pass a critical threshold to substantially alter early embryo patterning.

Our DMV-level analysis also supports the heightened severity of the Eed KO phenotype. For example, although DMVs are preferentially unmethylated within WT epiblast, they are *de novo* methylated within the WT Xecto lineage, including embedded CGIs (Extended Data Fig. 7a, b). However, these CGIs do not gain methylation specifically within the Eed KO Xecto, including promoters of critical early regulators of the epiblast and germline, such as *Otx2* and *Prdm14*, respectively (Fig. 3h, Extended Data Fig. 8b, c). Eed disruption therefore affects developmental gene promoters in both embryonic and extraembryonic compartments, producing an overall similar epigenetic pattern. Unfortunately, *Prdm14* is generally lowly expressed, even within PGCs, limiting our ability to precisely monitor its differential expression as a possible explanation for Eed KO-specific overproduction. However, *Dppa3*, another key PGC marker gene, is more broadly expressed within Eed KO 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 KO scRNA-seq data from E6.5 and E7.5 to see if these phenotypes otherwise obey general principles of WT development, including stepwise induction of committed progenitors by exogenous signals (Extended Data Figs. 9 and 10). Morphologically, we find minimal changes at E6.5, but confirm diminished complexity and delays by E7.5 (Fig. 4a, Extended Data Fig. 10b, c). Notably, *Prdm14* reporter activity indicates that PGCs are positionally specified in Eed KOs, 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 WT, but subsequently become either abnormal or fail to consistently develop (Fig. 4c). By E7.5, products of the primitive streak skew posteriorly towards extraembryonic mesoderm and PGCs, while the axial mesoderm (node, notochord) is highly abnormal and more advanced embryonic mesoderm or neural ectoderm do 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 KO embryos. Many changes detected at E8.5 are already apparent prior to gastrulation: *Cdkn2a* is already active within the embryo proper and ChrX 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*, *Hoxd9* double positive caudal mesodermal tissues<sup>36</sup>. In contrast, *Eed* KO 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*<sup>KO</sup> mESCs) to induce alternate fates *in vitro* (Extended Data Fig. 11a, b, Supplementary Figure 1). Specifically, we directed *Eed*<sup>KO</sup> mESCs into a formative epiblast-like state using FGF, followed by exposure to different concentrations of signaling components for an additional 48h. Across many conditions, *Eed*<sup>KO</sup> mESCs 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 impeding competing mesendodermal and surface ectodermal pathways with small molecule inhibitors (Fig. 4f, Extended Data Fig. 11c, d). Thus, the *Eed* mutant phenotype appears 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, which may be fundamental to a given developmental process. We investigated a number of key epigenetic regulators that produce lethal post-gastrulation phenotypes 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 somewhat greater severity that includes the overproduction of PGCs, broad destabilization of a shared PGC/pluripotency subnetwork, and failure to establish several key epigenetic features within the Xecto. Additional work will be necessary 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 strategies will contribute additional layers regarding how progenitor fields become altered without epigenetic supervision<sup>16</sup>. Finally, the ability to measure multiple 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.

## Materials and Methods

### Embryo generation

Protocols are adapted from those described previously<sup>40</sup>. Briefly, B6D2F1 strain female mice (age 6–8 weeks, Jackson labs) were superovulated by serial injection of Pregnant Mare Serum Gonadotropin (5IU per mouse, Prospec Protein Specialists) followed by Human Chorionic Gonadotropin (5IU, Millipore) 46 hours later. 12–14 hours after priming, MII stage oocytes were isolated in M2 media supplemented with hyaluronidase (Millipore) and stored in 25 µl drops of pre-gassed KSOM with ½ 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 F1 strain sperm in batches of 30–50 oocytes and standard micromanipulation equipment, including a Hamilton Thorne XY Infrared laser, Eppendorf Transferman NK2 and Patchman NP2 micromanipulators, and a Nikon Ti-U inverted microscope. Alternatively, for material subjected to whole genome bisulfite sequencing (WGBS), which does not require 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 ~6 hours of incubation and injected with a cocktail consisting of 200 ng/µl Cas9 mRNA and a 100 ng/µl equimolar ratio of 3–4 single guide RNAs (sgRNAs) targeting different exons of an epigenetic regulator gene locus (designed using ChopChop<sup>42</sup> and IDT's CRISPR-Cas9 guide RNA checker as described previously<sup>24</sup>, see Supplementary Table 10). Preferentially, targeted exons were chosen to be located towards the 5' end and to be shared across isoforms. At ~84 hours postfertilization, 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 hour offset in gestational time to accommodate implantation.

B6/CAST F1 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 in –40°C in 25 µl aliquots for use within 6 months to a year of collection. Cas9 mRNA and sgRNAs were *in vitro* transcribed using the mMESSAGING mMACHINE® T7 Ultra or MEGAscript® 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 (WT) and knockout (KO) embryos were isolated from surrogate mice between E6.5–8.5 at 12-hour intervals for WT and at gestational day E8.5 for epigenetic regulator KO experiments. The emergence of the Eed mutant phenotype was profiled in more detail by additionally sampling E6.5 and E7.5 KO embryos. Outermost extraembryonic tissues (yolk sac, trophoctoderm derived tissues) were preserved if possible. Microscope images recorded embryo number and morphology. Embryos were serially washed through several droplets of 1xPBS/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 minutes at 37°C, with pipetting in 5-minute intervals. The cell suspension was filtered using Scienceware Flowmi Cell Strainers, 40  $\mu$ m. Cells were washed twice with 1 ml 1xPBS/0.4%BSA and centrifugation for 5 minutes at 1200 rpm. The cell concentration was determined using a hemocytometer and cells were subjected to single-cell RNA sequencing (10x Genomic, 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 or 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 Table 1 and 2.

## Imaging embryos for morphology and size measurement

E7.5 and E8.5 wild type and Eed KO 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 10X and 7X objectives 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 50X and E8.5 embryos at 40X objectives. 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 WT embryos were acquired at 4X and KO embryos at 10X. Wild-type E7.5 and 8.5 embryos shown in Extended Data Fig. 10 to provide a size and morphological comparison to Eed KO 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 described previously<sup>43</sup>. The reporter strain was generated by the lab of Mitinori Saitou<sup>31</sup> and the mVenus Prdm14 promoter sperm (B6.Cg-Tg(Prdm14-Venus)1Sait/SaitRbrc; 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>; Reproduction. 2008 136(4): 503–14). PN3 zygotes were washed in M2 medium and prepared for electroporation. Electroporation reactions were setup according to

the Alt-R CRISPR-Cas9 ribonucleoprotein (RNP) complex protocol from Integrated DNA Technologies (IDT). RNP complexes were assembled just prior to electroporation. Briefly, 2  $\mu\text{L}$  of 200  $\mu\text{M}$  tracrRNA and 0.67  $\mu\text{L}$  of each 200  $\mu\text{M}$  crRNA were mixed, heated to 95°C for 5 minutes and allowed to anneal at room temperature for 10 minutes. 3  $\mu\text{L}$  of crRNA-tracrRNA mix, 1  $\mu\text{L}$  of 61  $\mu\text{M}$  Alt-R Hi-Fi Cas9 Nuclease 3NLS was diluted in 46  $\mu\text{L}$  of Opti-MEM medium and incubated at room temperature for 20 minutes.

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 milliseconds was used with an interval of 50 milliseconds, voltage decay of 10% and (+) polarity. Transfer pulse was applied at 5 V for 50 milliseconds with an interval of 50 milliseconds, 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 to E6.5, E7.5 and E8.5. Isolated embryos were washed in cold 1x PBS with 0.4% BSA and fixed overnight in 4% Paraformaldehyde (PFA) at 4°C followed by three washes in cold 1x PBS. Nuclei were stained with 0.24  $\mu\text{g}/\text{mL}$  DAPI for 40 minutes at 4°C. Images were acquired using Zeiss LSM880 at 10X magnification and z-stacks of 5  $\mu\text{m}$  interval. 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 KO 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 1x HBBS and then fixed overnight in 4% PFA at 4°C. Embryos were rinsed three times in 1x PBS and permeabilized with PBT0.5 (0.5% Triton X-100 in 1x PBS) for two hours followed by blocking for an hour with blocking solution (10% FBS in PBT0.5). Embryos were incubated for 72 hours at 4°C with the primary antibody anti-Histone H3 Lysine 27 tri-methylation (Abcam ab6002), diluted in blocking solution to 1:200. The following day, embryos were washed with PBT0.5 four times (30 minutes 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 minutes per wash), and nuclei were counterstained with DAPI (0.24  $\mu\text{g}/\text{mL}$ ) for 40 minutes at 4°C. Embryos were washed with PBT0.5 four times (30 minutes per wash), and post-fixed with 4% PFA for 20 minutes. Final washes (three times, 15 minutes) were performed with 0.02 M phosphate buffer (0.025 M  $\text{NaH}_2\text{PO}_4$ ; 0.075 M  $\text{Na}_2\text{HPO}_4$ , pH 7.4) followed by optical clearing at 4°C for 24 – 48 hours with 1.62 M RIMS clearing agent (Histodenz in 0.02 M phosphate buffer). Images were acquired using Zeiss LSM710 at 63X magnification (oil immersion) and z-stacks of 2.13  $\mu\text{m}$  intervals were generated. Two independent experiments were performed with similar results. A representative z-stack is shown in Extended Data Figure 9c.

## Eed<sup>KO</sup> mESC line generation and fate induction experiments

WT V6.5 mouse ESCs (provided by the lab of Konrad 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 (See Supplementary Table 10). Subclones were expanded and homozygous *Eed* disruption was confirmed by target site amplification and Sanger sequencing.

Our selected Eed<sup>KO</sup> 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-methyl-histone H3 antibody (Cell, Signaling, C36B11, at 1:500 dilution). Histone 4 was detected by anti-histone H4 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.

WT and Eed<sup>KO</sup> mESC 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 prior to induction with exogenous factors<sup>44</sup>. For signaling experiments, 10,000 cells were plated per well into N2B27 media containing 12 ng/ml bFGF. For these experiments, we used human plasma fibronectin (purified protein, Millipore) coated 8-well chamber slides ( $\mu$ -Slide 8 Well, ibidi). After 24 hours, media was exchanged with N2B27 containing 12 ng/ml bFGF and select concentrations of signaling compounds and/or small molecule inhibitors for an additional 48 hours. Final concentrations of growth factors or small molecule inhibitors were as follows: 12 ng/ml Recombinant Human bFGF (R&D Systems); 0.25  $\mu$ M Retinoic acid (Sigma); 5 and 500 ng/ml Recombinant Human BMP-4 Protein (R&D Systems); 10, 100, 1000 ng/ml Recombinant Human WNT-3A Protein (R&D Systems); 10 and 1000 ng/ml Recombinant Human/Murine/Rat ACTIVIN A (Peprotech); 0.5  $\mu$ M ALK2/3 inhibitor LDN-193189 (Stemgent, 10 mM solution) to inhibit the Bmp4 pathway; 3.3  $\mu$ M Tankyrase1/2 inhibitor XAV939 (Tocris) to inhibit the Wnt pathway; and 10  $\mu$ M TGF- $\beta$  RI Kinase Inhibitor VI SB431542 (Millipore) to inhibit Activin/Nodal signaling. Total RNA was isolated by washing wells twice with PBS followed by adding 350  $\mu$ l 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 bFGF after 24 and 72 hours, respectively. mESC 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*, *Gusb*), 400 ng total RNA were used in a NanoString nCounter PlexSet assay to profile 88 RNA samples of the mESC experiments described above (triplicates for all but one condition, duplicate for 100 ng/ml WNT-3A). Probe hybridization was set up according to manufacturer's instructions and performed for 24 hours (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 WT and Eed<sup>KO</sup> 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

If not stated otherwise: All statistics and plots are generated using R version 3.5.1 “Feather Spray”. Boxes indicating the median and quartiles with whiskers reaching up to 1.5 times the interquartile range. The violin plot outlines illustrate kernel probability density, i.e. 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. Heatmaps were plotted using the Complex Heatmap package<sup>45</sup> and browser track figures using the Gviz package<sup>46</sup>.

## Preprocessing

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>47</sup> with default settings and “--outSAMattributes NH HI NM MD.” The hybrid genome was prepared using SNPsplit<sup>48</sup> to mask SNPs between the mouse version mm10 (GRCm38) and the CAST/EiJ strain genomes with the ambiguity base (N). Subsequently, SNPsplit was used to sort reads that cover SNPs by origin (reference genome). Unambiguous and unique alignments of WT 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, i.e. 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 (`kmeans` 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 unstable cell to embryo assignments were often either the result of low UMI counts or of very high counts, most likely 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 *Erdr1* (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 *Erdr1*, *Ddx3y* or *Eif2s3y* were determined to be male (Supplementary Table 1).

### Genotyping - Cluster determination

The cluster determination was split into four main parts and was largely done using the R package Seurat with default settings<sup>49</sup>. The establishment of the WT 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 agnostically clustering WT 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 WT 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 WT and KO 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 centers (WT): 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 centers." This allowed us to detect even rare cell states while preserving embryo-specific variability.

(2) Cell cluster (WT): The embryo specific centers of all WT stages were combined into one analysis to determine variable genes. A PCA was run based on the variable genes and the first 20 PC's were used to cluster the embryo specific centers (parameters adjusted for low 'cell' number: k.param = 8, k.scale = 50, prune.SNN = 1/10). This resulted in 42 clusters of embryo-specific centers and the median expression profile of each cluster was calculated to

form preliminary cell states. Then, as a temporary step, all WT 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 WT reference cell states: WT 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 in the course of the generation of this manuscript. Thus, raw data was reprocessed and the cell state kernels were adjusted by again assigning the WT cells to the kernels.

(4) Cell states of single cells: The WT and KO 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, and 6a). Cell states with an insufficient number of cells from KO embryos ( $< 30$  cells) were discarded from further analysis.

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>50</sup>. Of the 1,530 genes that show biased expression in at least 10% of these tissues within an F1 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 were comprised of several embryos and never resulted from a single embryo.

### Cell states prevalence

Prevalence of cell states with respect to embryo stage (Extended Data Fig. 1h) was evaluated normalizing each state across the developmental stages (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  $\log_2$ -fold change between the mean proportions of developmentally stage-matched KO and WT embryos.

### Correlation of gene expression

Gene expression profiles were compared between either two different cell states or between the WT and Eed KO by correlating the average gene expression profiles of the marker genes (R function cor, Pearson correlation).

### Differential expression

We called differentially expressed genes between WT and KO 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>51</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 KO vs WT cells. A gene was called differentially expressed within a cell state between WT and KO 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) either a minimum difference of 0.1 in percent positive cells or a minimum absolute  $\log_2$  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 KOs 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, extraembryonic endoderm; Xecto, extraembryonic ectoderm) or the embryonically derived lineages (Epiblast; Xmeso, extraembryonic mesoderm; Eendo, embryonic endoderm, Eecto, embryonic ectoderm; Emeso, 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 cutoff (Supplementary Table 7).

### Stage matching metric to assign “developmental stage”

The gestational age of all KO embryos was adjusted for developmental delay by comparing cell state data to the median of the WT embryos from each time point. Because some states may be more informative about developmental stage than others, we performed two distinct principal component analyses (PCA) using the WT 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 (Emeso, Eendo, Eecto, Epiblast, PGCs and Xmeso), since Xendo and Xecto cell state proportions are more sensitive to embryo dissection and single cell dissociation. The R function `prcomp` (parameters: `retx = TRUE`, `center = TRUE`, `scale = TRUE`) was used to calculate PCs for WT embryos and WT medians and the `predict` function transformed KO data according to the WT loadings (Extended Data Fig. 3b–d, Supplementary Tables 1 and 5). The first PCs of both PCAs were used to assign each KO embryo to its closest median WT by Euclidean distance.

### H3K27me3 ChIP-seq data

Publicly available H3K27me3 ChIP-seq data of E6.5 epiblast and extraembryonic ectoderm<sup>52</sup> were used to calculate the average H3K27me3 occupancy of each gene's promoter region (calculated as the region 1500 bp upstream to 500 bp downstream of the TSS). Only the first two replicates were used for each, since these two replicates showed a similar trend when compared to WT gene expression of our epiblast or Xecto cell states, while the third replicate did not show any linear relation to gene expression. For both data sets, 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 therefore set at this threshold.

### PGC number estimation

The total number of PGCs per WT or Eed KO 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 prior to loading (as measured using a hemocytometer, 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

Uniform Manifold Approximation and Projection (UMAP) was used as a dimension reduced visualization of single-cell marker gene expression profiles<sup>53</sup>. Transformation of the WT data was performed using the R function `umap` and subsequently applied to all KO data to project it onto the same manifold as produced for WT.

### RNA velocity

RNA velocity was calculated using the `velocyto` tool<sup>54</sup> and visualized using `scanpy`<sup>55</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 DNA sequence of the respective gene using STAR<sup>47</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 was 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 7 embryos, pooled and processed into WGBS libraries using the Accel-NGS Methyl-seq® kit as previously described using 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. In order 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>56</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 WT tissue.

CpG islands 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 CpG islands were previously defined<sup>24</sup>.

The CpG density of a genomic region was calculated as the fraction of CpG dinucleotides within a 200 bp window (sliding window with 20 bp offset).

DNA methylation valleys (DMVs) were detected using a 2 kb sliding window (500 bp offset). Regions with an average methylation rate below 0.15 in WT (excluding CpG island methylation) were merged given a maximum distance of 1 kb.

### 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 trophoctoderm (TE) of the E3.5 blastocyst (GSE84236), and late stage samples including an average of somatic tissues and the E14.5 placenta (GSE42836).

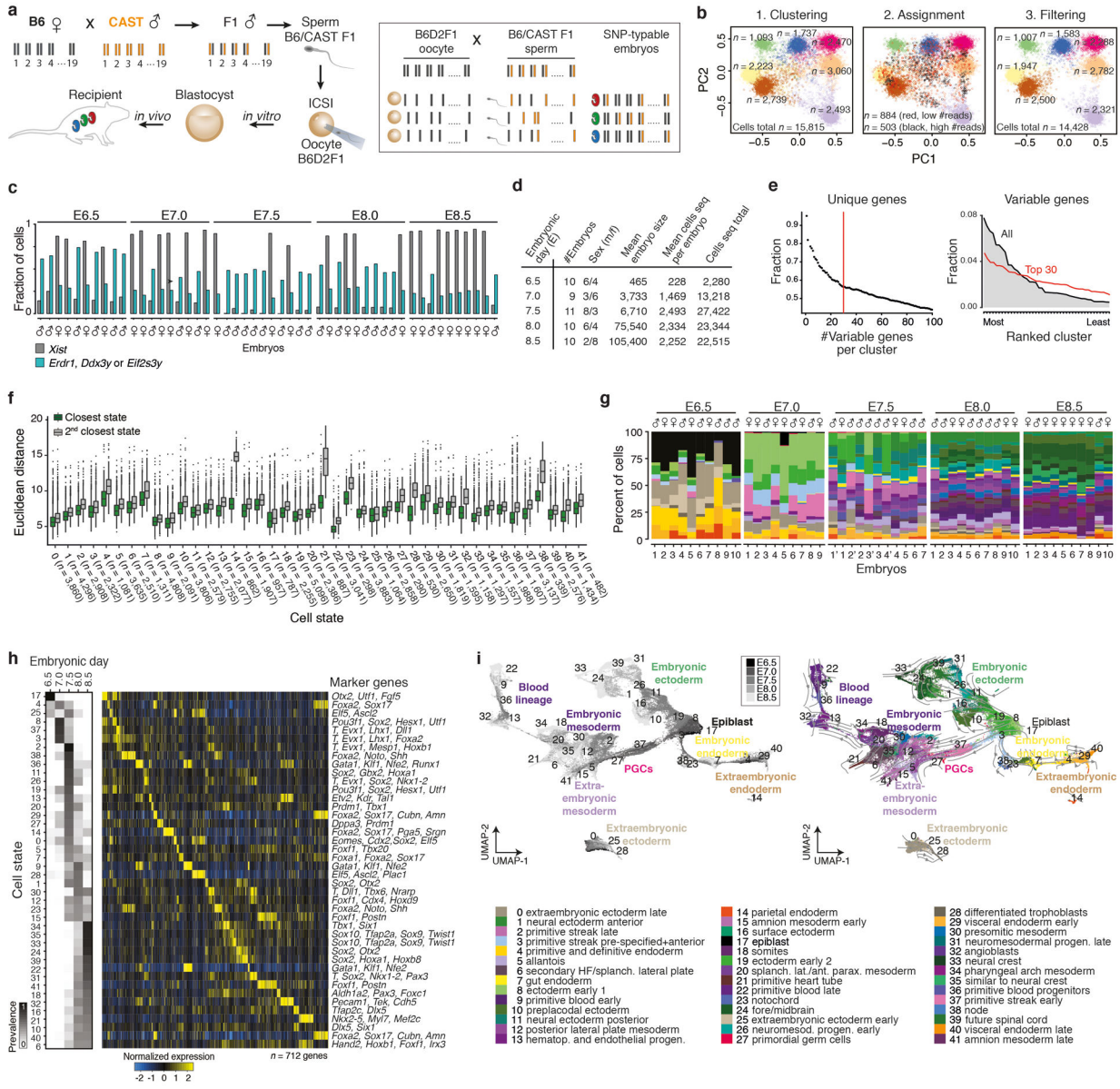
### Code availability

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

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Extended Data



**Extended Data Figure 1. SNP-based genotyping and assignment of single cells into 42 discrete cell states**

**a.** Single nucleotide polymorphism (SNP) based cell-to-embryo assignment strategy. Embryos were generated by intracytoplasmic sperm injection (ICSI) using sperm from hybrid males (C57BL6/J × CAST/EiJ) to confer a randomly inherited CAST/EiJ haplotype. Siblings (individually colored embryos) are pooled prior to single-cell RNA sequencing (scRNA-seq) and computationally deconvoluted based on their embryo-specific SNP profiles. Briefly, the ratios of CAST-specific SNPs (orange) are scored per chromosome to cluster cells into distinct embryos. We use B6D2F1 (C57BL6/J × DBA) oocytes, whose genotypes differ by only ~4.5M SNPs compared to ~17.7M for CAST/EiJ<sup>57</sup>.

**b.** SNP-based deconvolution of seven pooled E7.5 wild-type (WT) embryos. **Left:** Principal Component Analysis (PCA) projection of autosomal CAST SNP ratios for all sequenced cells with 1,000 covered SNPs. Cells are colored by cluster assignment, indicating

individual genotypes (embryos). **Center:** 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, while those with higher counts collect between clusters and likely 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-linked gene transcripts (*Erdr1*, *Ddx3y* or *Eif2s3y*, blue) used for sex-typing. For cell numbers, see Supplementary Tables 1 and 2.

**d.** Summary statistics of profiled WT embryos from E6.5–8.5 ( $n = 50$  total).

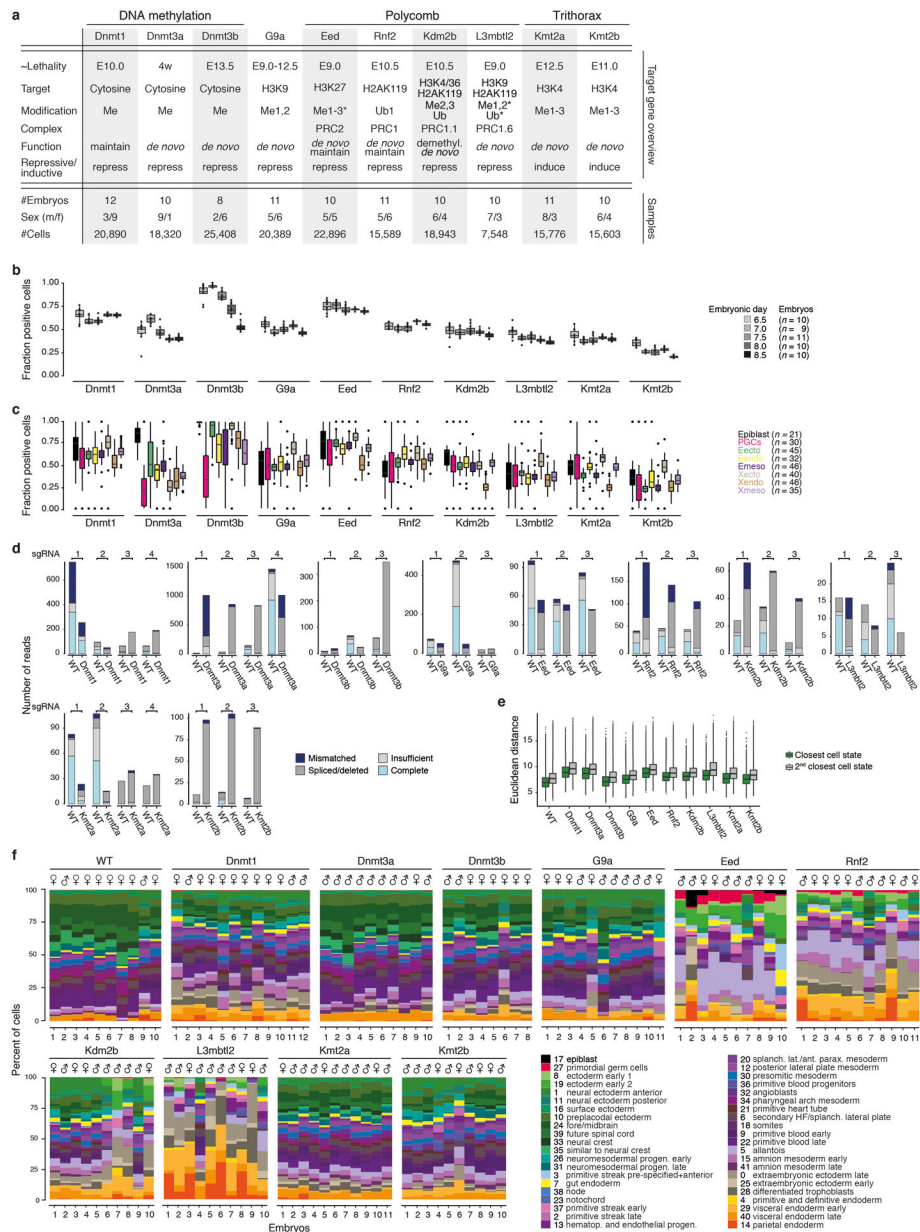
**e. Left:** Fraction of variable genes that are uniquely assigned to a single state when taking the top N-most differentially expressed genes per cluster. We selected the top 30 most unique genes per cluster ( $n = 712$  genes) because it maximizes the information per cluster under the constraint that 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 barplots show percent 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:** Heatmap of cell state prevalence across profiled embryonic stages. The median state proportions are calculated across embryos for each time point, then row normalized across time points to show their dynamics. **Right:** Expression heatmap of our 712 marker genes, with key markers for each state highlighted (see Supplementary Text). Mean state expression for each marker gene is normalized over the column and arranged by maximal expression value across states.

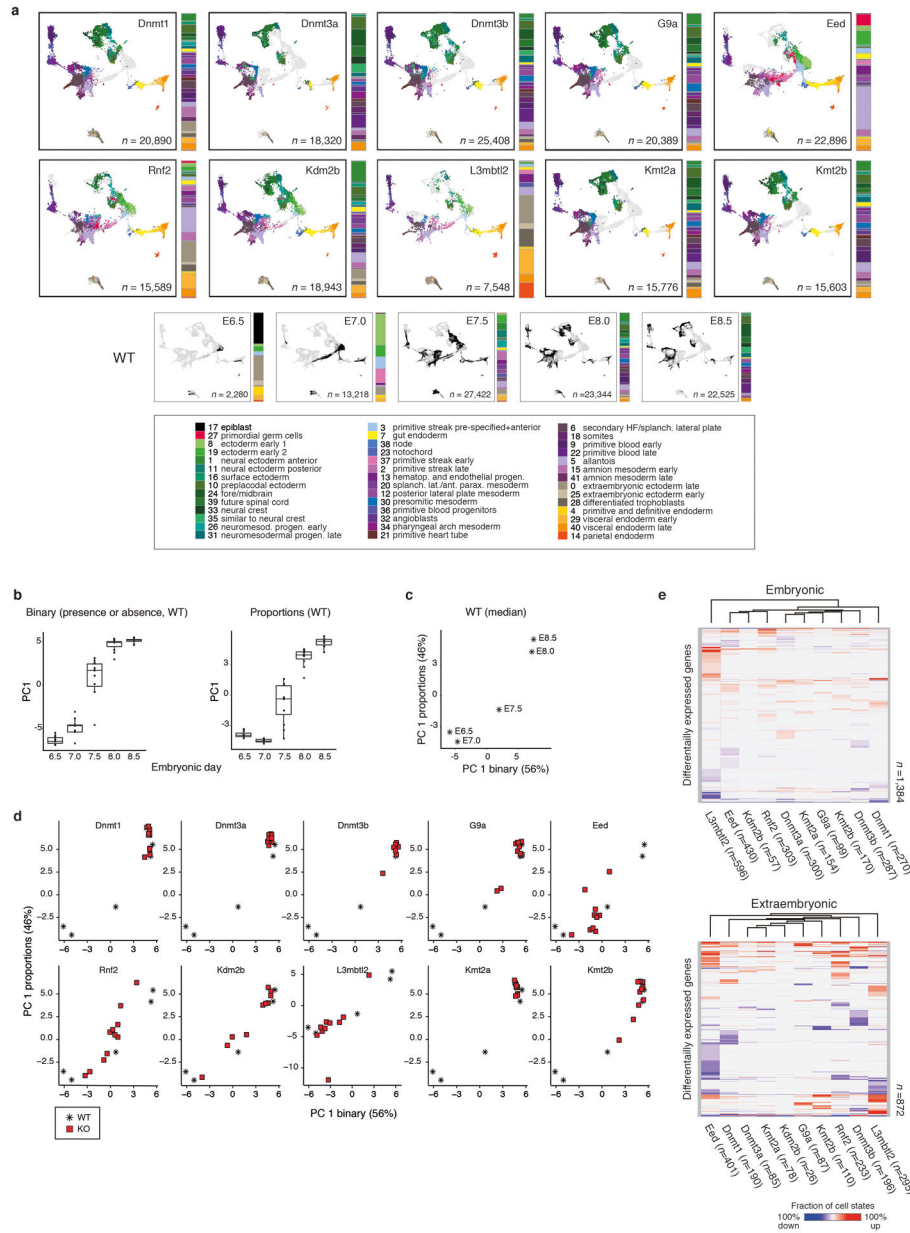
**i. Left:** Uniform Manifold Approximation and Projection (UMAP) of WT cells ( $n = 88,779$ ) colored by time point from dark to light gray. **Right:** WT UMAP overlaid with RNA velocity<sup>54</sup> information as an indicator of transcriptome dynamics between different cell states.



### Extended Data Figure 2. Efficient genetic perturbation of epigenetic regulators and cell state characteristics across KO embryo replicates

**a. Top:** Epigenetic regulators investigated here with information about their target residues and function and grouped into three key pathways: regulation by DNA methylation, Polycomb, or Trithorax. The majority of lethality phenotypes occur soon after our last experimental collection time point (E8.5)<sup>22,26,28–30,58–61</sup>. L3mbtl2 is a methyl-histone binding protein that participates in PRC1 regulation as part of ncPRC1.6. 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 WT 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. *n*'s reflect the number of embryos collected at each time point.
- c.** Fraction of cells positive for selected epigenetic regulator genes in WT for eight major developmental lineages. *n*'s reflect the number of embryos for 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 1 base is a mismatch/deletion/insertion; Spliced/deleted, split read spans 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 E8.5 WT embryos is shown in comparison for each target site. A more comprehensive analysis of zygotic disruption is presented for Eed KO in Extended Data Fig. 9.
- e.** KO embryo cells can be described using WT-defined states. Boxplots show the single cell Euclidean distances to their closest (green) and second closest (grey) states per KO 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 KO cells as we do for the WT cells from which our state kernels were derived.  $n = 88,779; 20,890; 18,320; 25,408; 20,389; 22,896; 15,589; 18,943; 7,548; 15,776; 15,603$  left to right
- f.** Barplots showing the percentage of cells per embryo (x-axis) that were assigned to each of our 42 cell states (colors, y-axis) with E8.5 WT provided for comparison. Notably, KO 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 KOs underproduce early germ layer 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 Figure 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 WT cell states and projected onto our WT-defined gastrulation UMAP. That KO cells fall within WT states cannot confirm equivalent functionality or potential, but does suggest that cell states are largely constrained even without key epigenetic regulators. Instead, many KO embryos differ from WT by cell state composition. Adjacent barplots reflect the median embryo composition. A reference key for our WT time series is provided. *n* = number of cells.
- b.** Distribution across Principal Component 1 (PC1) for WT embryos (dots *n* = 10, 9, 11, 10, 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

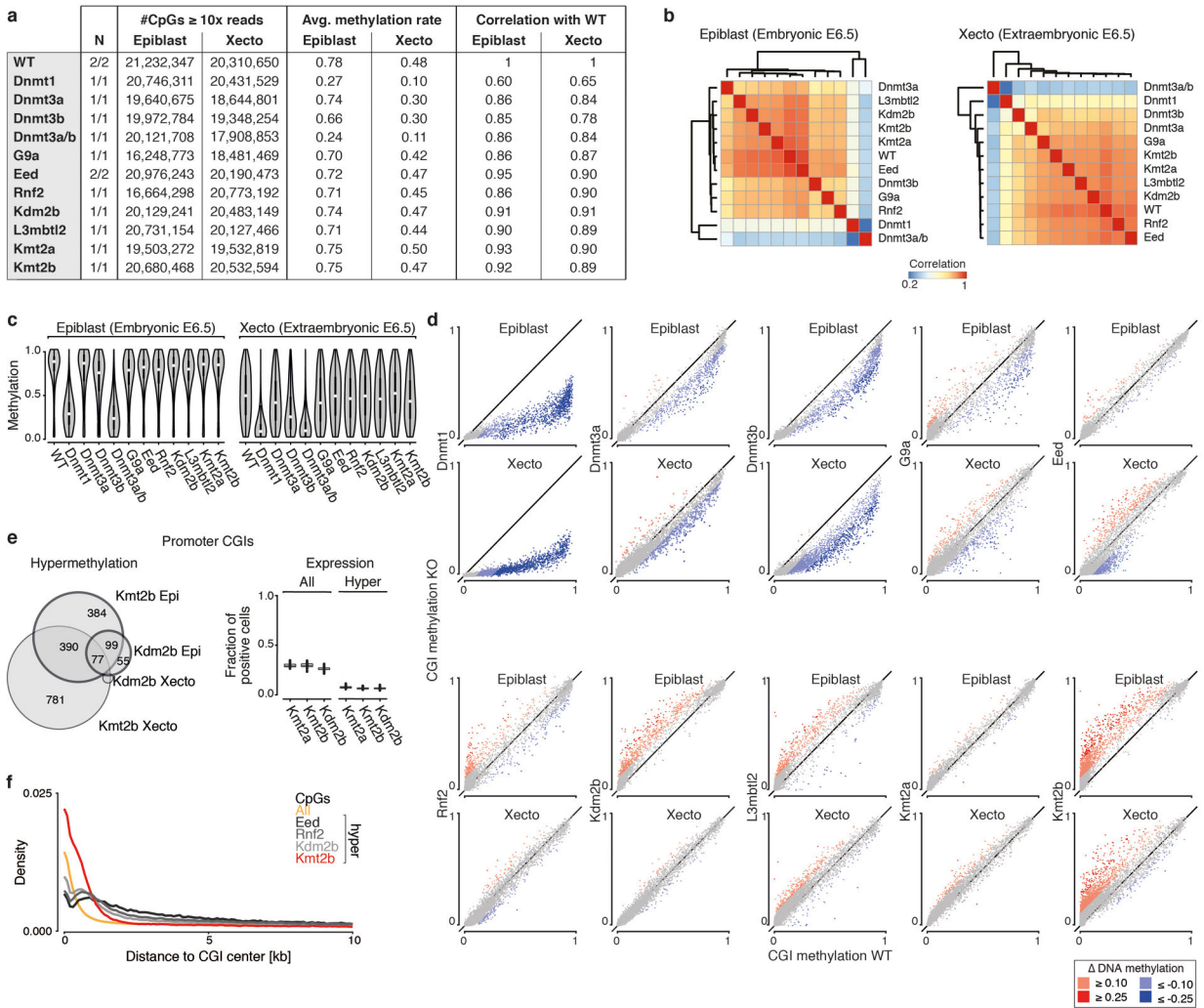
developmental stages (i.e. E6.5-E7.5) are better resolved according to the presence or absence of key states associated with the primitive streak, while later time points (i.e. E8.0 and E8.5) share many of the same states, but at different proportions. Tissues prone to technical recovery biases during embryo isolation (Xecto and Xendo) were excluded from this analysis (Supplementary Table 5).

**c.** PC1 values for median WT embryos ( $n = 5$  time points, stars). 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 KO embryo replicates (squares) by projecting them onto the WT-defined PCA space described in **c.** KOs of the DNA methyltransferases and the histone methyltransferases Kmt2a and G9a show no or mild developmental delays. The Polycomp components Eed, Rnf2, Kdm2b and L3mbtl2 exhibit stronger setbacks in developmental progression, with greater variability. For staging information see Supplementary Table 1.  $n = 12, 10, 8, 11, 10, 11, 10, 10, 11, 10$  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 KO to WT cell state, split by embryonic (**Top**) or extraembryonic (**Bottom**) origin (Supplementary Table 6). Differentially up- or downregulated genes found in 2 states are shown in red and blue, with color 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 KO 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 if lineage-specific regulation has been impeded or if these differences are merely a consequence of subtly offset development. Additional GO term analysis: upregulated genes in KOs of the three Dnmt enzymes significantly overlap with imprinted genes ( $Q = 1.3 \times 10^{-12}$  for Dnmt1 KO embryonic, upregulated) and our G9a KO is statistically enriched for genes upregulated in a transgenic G9a knockout model ( $Q = 0.002$  for G9a KO embryonic, upregulated).





**Extended Data Figure 4. Aberrant DNA methylation in epigenetic regulator mutants at the onset of gastrulation**

**a.** Overview of Whole Genome Bisulfite Sequencing (WGBS) data for epiblast and extraembryonic ectoderm (Xecto) of E6.5 WT and KO embryos. Correlation with WT methylation profiles are lowest for KO of DNA methyltransferases, as expected. Additional data generated using both Dnmt3a and 3b sgRNAs confirms the redundancy of the enzymes and results in a gross reduction in DNA methylation to levels seen for the Dnmt1 KO.

**b.** Correlation heatmaps of global DNA methylation at single CpG resolution between all epigenetic regulator mutants as well as WT, clustered independently for the epiblast and Xecto by Pearson.

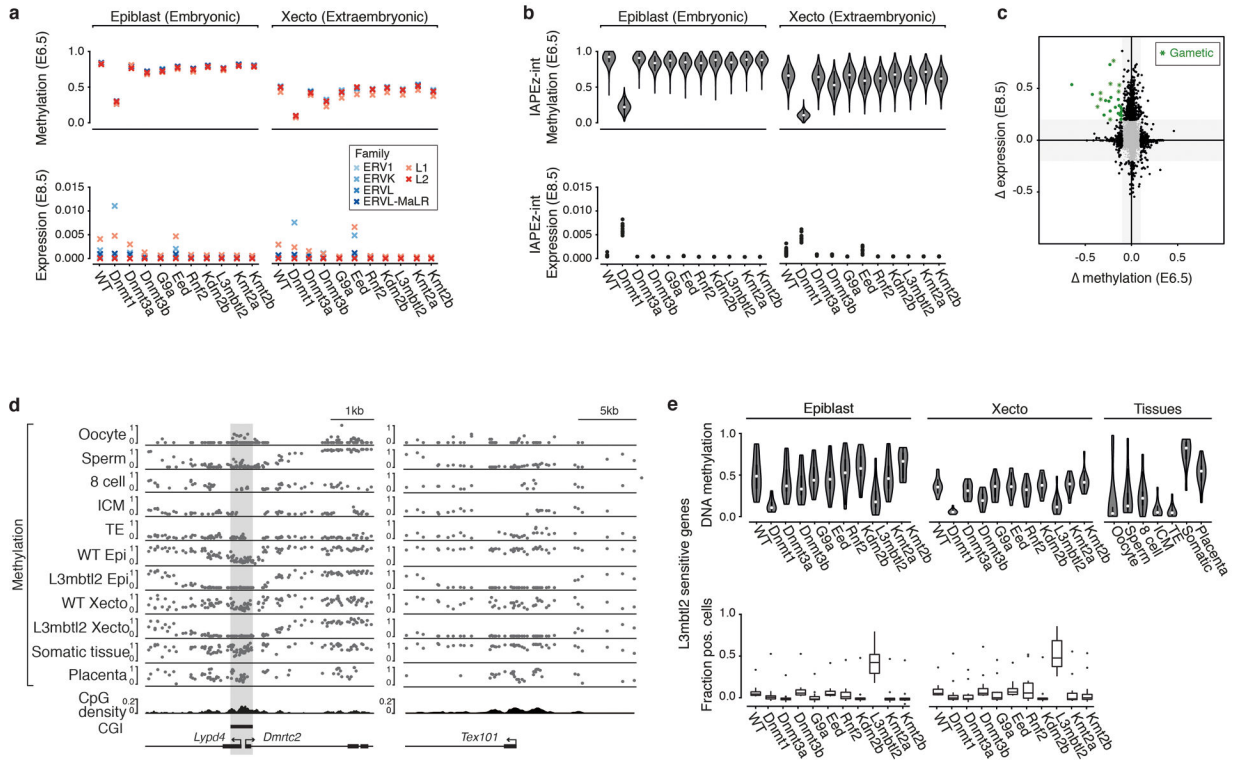
**c.** Violin plots of single CpG methylation status in WT and epigenetic regulator mutants. While most KO do not show obvious differences to WT, large drops in methylation were observed for the KO of the maintenance methyltransferase Dnmt1 and for Dnmt3a and b combined. The effect for Dnmt3b KO alone is substantially weaker albeit more pronounced in Xecto, where it represents the primary *de novo* methyltransferase. Number of CpGs per sample is reported in **a**. Epiblast  $n = 21,232,347; 20,746,311; 19,640,675; 19,972,783; 20,121,708; 16,248,772; 20,976,243; 16,664,297; 20,129,240; 20,731,153; 19,503,271;$

20,680,467; Xecto:  $n = 20,310,650$ ; 20,431,529; 18,644,801; 19,348,253; 17,908,853; 18,481,468; 20,190,473; 20,773,191; 20,483,148; 20,127,465; 19,532,818; 20,532,593 CpGs.

**d.** Scatterplots of CpG island (CGI) methylation in epigenetic regulator mutants (y-axis) versus WT (x-axis) for the E6.5 epiblast or Xecto. Red and blue indicate methylation increases or decreases compared to WT ( $> 0.1$ , light;  $> 0.25$ , dark). Dnmt1 KO shows the greatest loss in both epiblast and Xecto, followed by Dnmt3b KO. Kdm2b KO shows substantial gain specifically within the epiblast, which is also apparent in Rnf2 and L3mbtl2 KOs to lesser degrees. In contrast, Eed KO loses CGI methylation within the Xecto. Kmt2b KO has the greatest increase in CGI methylation within both the epiblast and Xecto.  $n = 12,410$  CGIs displayed across all plots.

**e.** CGI methylation in Kdm2b or Kmt2b KO is largely associated with genes that are lowly or not expressed. **Left:** Venn diagram of hypermethylated CpG island promoters between Kmt2b and Kdm2b KO shows a large overlap. Furthermore, hypermethylated CpG island promoters have a  $\sim 2.5$ -fold enrichment for H3K27me<sub>3</sub>-based regulation compared to background<sup>52</sup>. **Right:** Boxplots showing the expression of genes with CGI-containing promoters, calculated as the fraction of positive cells for each embryonic cell state. Data is shown for all CGI promoter-containing genes or for those that are hypermethylated in either the Kmt2b or Kdm2b KO epiblast (bold circles in left,  $n = 1,026$  hypermethylated CGI-containing promoters total between Kmt2b and Kdm2b KO epiblast). Overall, genes that gain promoter methylation are lowly expressed across lineages independent of methylation state. The Kmt2a KO is shown for comparison because it does not gain promoter methylation at E6.5.

**f.** Distance to the nearest CGI center for all CpGs in the genome as well as for hypermethylated ( $> 0.1$ ) CpGs in Eed, Rnf2, Kdm2b and Kmt2b KO epiblast. Kmt2b hypermethylated CpGs are strongly shifted towards the center, while PRC KOs tend to methylate CpGs in close proximity to, but not within, CGIs.



**Extended Data Figure 5. DNA methylation-dependent changes in gene and retrotransposon expression.**

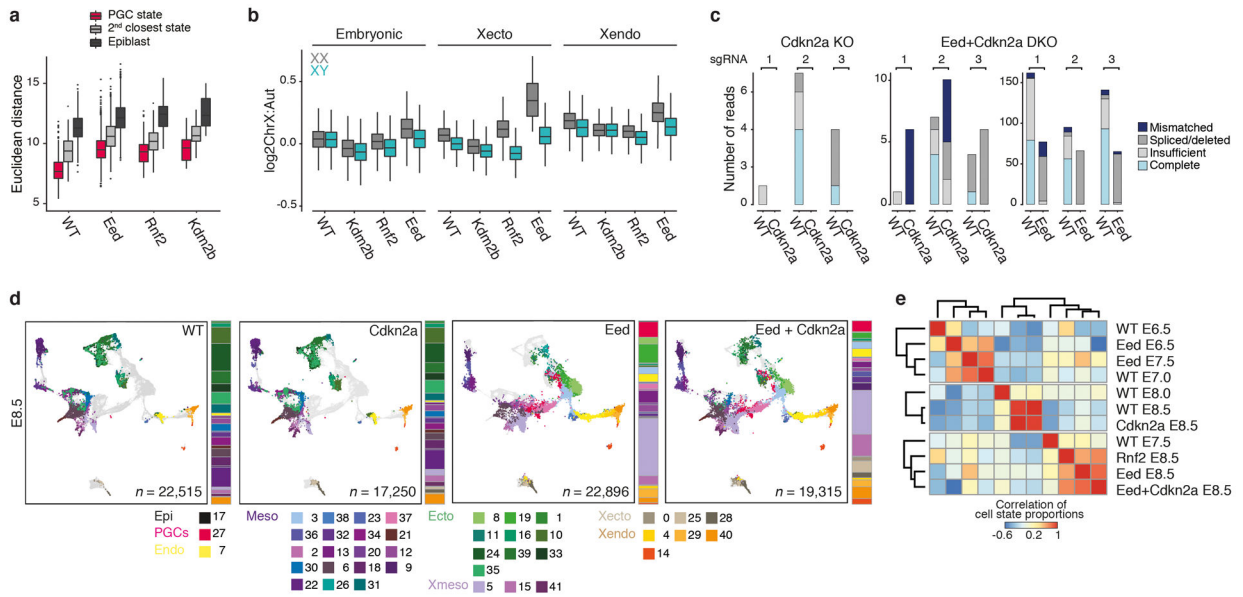
**a.** Average E6.5 DNA methylation (**Top**) and E8.5 expression (**Bottom**) for retrotransposon families. Expression was calculated as the normalized fraction of reads recovered from scRNA-seq data for each subfamily. The Dnmt1 KO shows the strongest reduction in methylation across retrotransposons in the epiblast and the Xecto. The ERVK family of LTRs shows the strongest corresponding increase in expression, which is higher in the embryonic lineage than in Xecto.

**b.** Intracisternal A particle (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 KO. **Bottom:** Mean expression within the embryonic and Xecto lineages of E8.5 KO embryos, shown as the fraction of total reads per cell. Epiblast IAPeZ-int:  $n = 5,585; 5,579; 5,510; 5,440; 5,210$ , Xecto IAPeZ-int:  $n = 5,576; 5,577; 5,498; 5,421; 5,367; 5,575; 5,529; 5,518; 5,500; 5,411; 5,543$ .

**c.** Scatterplot of E6.5 promoter DNA methylation and E8.5 expression differences in the Xecto lineage of L3mbl2 KO compared to WT, as shown for the embryonic lineage in Fig. 2g. Differentially hypomethylated (delta  $-0.1$ ) and derepressed genes (delta  $0.2$  fraction positive cells) in L3mbl2 KO (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 piRNA biogenesis pathway, including the dead-box helicase *Ddx4* (VASA homolog) and *Maelstrom*, 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 the KO while embryonic lineages arrest shortly after gastrulation onset.

**d.** Genome browser tracks of WGBS methylation data for three aberrantly regulated loci in L3mbtl2 KO embryos. The bidirectional genes *Lypd4* and *Dmrtc2* initiate from the same CpG island (CGI), while *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 continues to increase over development. *De novo* methylation does not occur in the L3mbtl2 KO and corresponds with sharp increases in gene expression. WT data from gametes, preimplantation embryos, and late stage samples like somatic tissues and the E14.5 placenta are taken from Ref's<sup>62–65</sup>.

**e.** Promoter DNA methylation (**Top**) and E8.5 expression (**Bottom**, shown as fraction of positive cells per embryo replicate) boxplots of L3mbtl2 sensitive genes ( $n = 13$  genes taken from Fig. 2g, green). Many gametogenesis genes are regulated by “weak” CGI-containing promoters that become methylated during development<sup>66</sup>. In line with this, the promoters of L3mbtl2 KO sensitive genes are hypomethylated in gametes and preimplantation and become *de novo* methylated over postimplantation development. Derepression is specific to L3mbtl2 KO, and does not occur in *Dnmt1* or *Dnmt3b* KOs, where methylation levels drop globally. Expression changes are also not substantial for *Rnf2* or *G9a* KO although these regulators are also expected to participate in ncPRC1.6 complex-directed repression. A single outlier gene, *Ttr*, is expressed in all KOs and WT, but is still upregulated in L3mbtl2 KO. Additional data taken from previous studies<sup>62–64,67</sup>.



**Extended Data Figure 6. Impact of derepressed Polycomb group regulator targets**

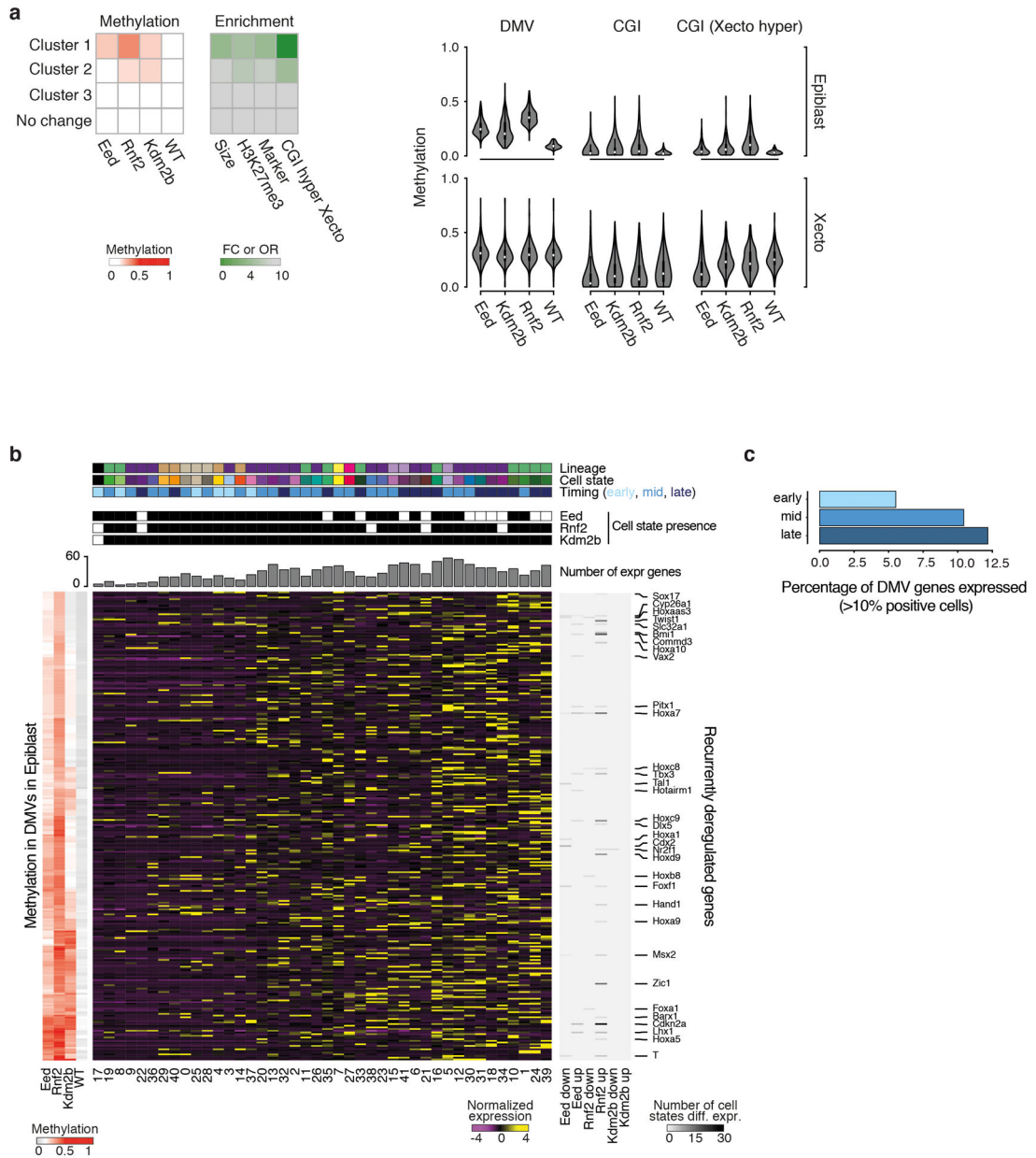
**a.** Euclidean distances of PGC-assigned cells from our WT, Eed, Rnf2, and Kdm2b KOs to the mean marker gene expression of our PGC (state 27, magenta), their second closest (light grey) or the epiblast (state 17, dark grey) cell states. PGC-assigned KO cells are transcriptionally distinct from the next closest or epiblast state, supporting our observation that this state is specifically overproduced in the Eed KO. We include the epiblast state as it shares some master regulators with PGCs and because some cells of this state are still present in the Eed KO. 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 boxplot: center line, median; edges, IQR; whiskers, 1.5xIQR; outliers, individually plotted. Number of recovered PGC state assigned cells is  $n = 290, 1,564, 250,$  and  $44$  for WT, Eed, Rnf2, and Kdm2b.  $P$ -value =  $2.644257e-49, 3.733801e-257, 9.3103e-43,$  and  $1.136868e-13$  for PGC vs 2nd closest state.

**b.** Per cell ChrX to autosome transcript ratios for PRC regulator KOs and WT cells, separated by sex. In our breeding system, X chromosomes are exclusively the B6 genotype, which makes it impossible for us to evaluate mono- vs biallelic transcription. However, these internally normalized measurements reveal increased transcription of ChrX-linked genes within certain KO lineages of female embryos. The Eed KO Xecto is most extreme and shows corresponding proliferation defects at E8.5 (see Fig. 3d). Within the Eed KO, female-specific ChrX deregulation is more subtly observed for the Xendo and embryonic lineages, implying either higher redundancy between PRC1 and 2 after the allocation of the trophectoderm or a lineage-specific failure to renormalize ChrX's transcriptional output within Xecto. In Rnf2 KO, the effects generally follow a similar trend but are more muted. Embryonic cells:  $n = 39,411; 37,887; 5,391; 9,233; 4,448; 5,248; 7,459; 11,264,$  Xecto cells:  $n = 1,769; 3,685; 755; 1,372; 1,465; 1,220; 19; 1,594,$  Xendo cells:  $n = 2,509; 3,518; 773; 1,419; 1,745; 1,463; 1,013; 1,547.$

**c.** Reads spanning the sgRNA protospacer sequences confirms high efficiency disruption of *Eed* and *Cdkn2a* loci in single (*Cdkn2a* KO) and double (*Eed*+*Cdkn2a* DKO) sgRNA injected embryos. Figure as in Extended Data Fig. 2d.

**d.** Single cells from *Cdkn2a* KO and *Eed*+*Cdkn2a* DKO embryos were assigned to one of our 42 WT cell states, projected onto our WT gastrulation UMAP and compared to E8.5 *Eed* KO and WT. Barplot shows the median embryo composition. In general, our DKO resembles the *Eed* KO, demonstrating that the derepression of the *Cdkn2a* locus in *Eed* KO is not responsible for the overall phenotype. The *Cdkn2a* KO is indistinguishable from WT.  $n =$  cells.

**e.** Correlation heatmap of average cell state composition for our *Cdkn2a* KO and *Eed*+*Cdkn2a* DKO embryos compared to WT stages and other core PRC component KOs, including a 24 h resolution *Eed* KO time series described below (Fig. 4, Extended Data Fig. 10). *Cdkn2a* KO clusters with WT E8.0 and E8.5, while the *Eed*+*Cdkn2a* DKO clusters with WT E7.5 as well as our *Eed* and *Rnf2* KOs.  $n = 42$  cell states, Pearson correlation.



**Extended Data Figure 7. Molecular abnormalities of Polycomb group regulator mutants**

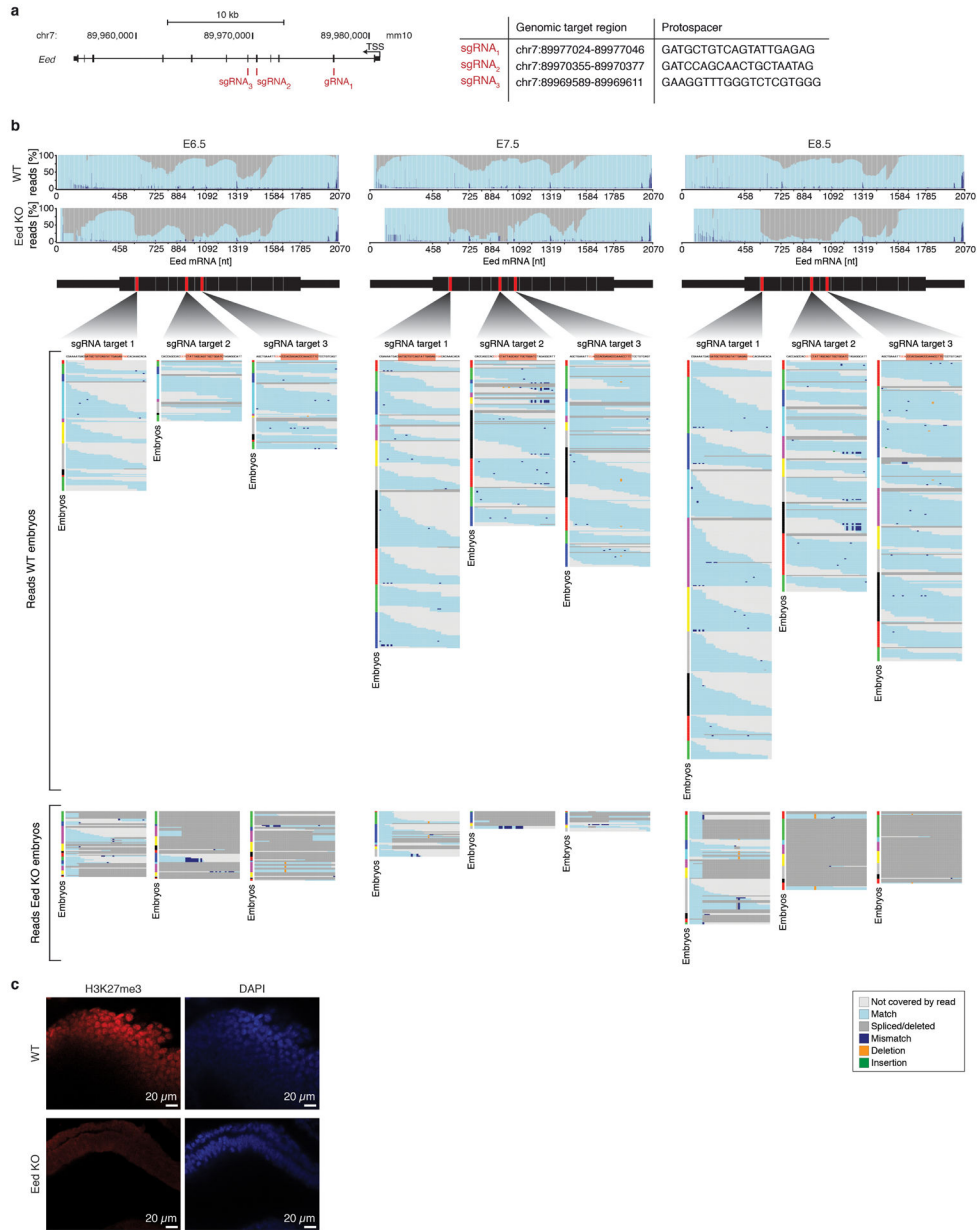
**a. Left:** Large, multi-kb DNA methylation valleys (DMVs) associated with developmental genes gain DNA methylation in PRC KO. We clustered 8,972 DMVs that exist within the WT E6.5 epiblast according to their methylation in our PRC KO. A discrete set of 248 is specifically methylated within our Eed, Rnf2, and Kdm2b KO (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 ~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 KO is available as

Supplementary Table 8.  $n = 248$  DMVs in cluster 1 vs  $n = 6,888$  for no change. **Right:** DNA methylation violin plots of the 248 DMVs that gain CpG methylation within the E6.5 epiblast of our PRC KO. “DMV” measures methylation of all non-CpG island CpGs within DMV boundaries, while “CGI” measures those for all CGI 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 WT Xecto ( $n = 191$ ). In the epiblast, DMV methylation is highest for Rnf2 KO and lower for the same regions in Eed KO. In contrast, Kdm2b KO shows substantial heterogeneity, with >55% of DMVs showing lower methylation compared to the Eed KO. The DMVs that gain DNA methylation in the epiblast of PRC KO are generally naturally *de novo* methylated in the Xecto (including methylation of CGIs). Here, the CGIs in the Eed KO pose an exception as they show a specific loss of methylation.

**b.** Heatmaps showing the WT expression status of 303 genes contained within differentially methylated DMVs. In PRC component KOs, the loss of bivalence 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. While 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 KO and WT as calculated in **a** (with CGI CpGs excluded). **Middle:** Row-normalized expression of DMV-associated genes across our 42 WT states. **Right:** Fraction of KO 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 KOs. States are designated as early, middle or late (most prevalent in WT 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 WT is also provided.

**c.** The percentage of DMV-associated genes that are expressed in our 42 WT states collapsed into early, middle or late based upon when states emerge (E6.5–7.0, 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 gastrulation time series.





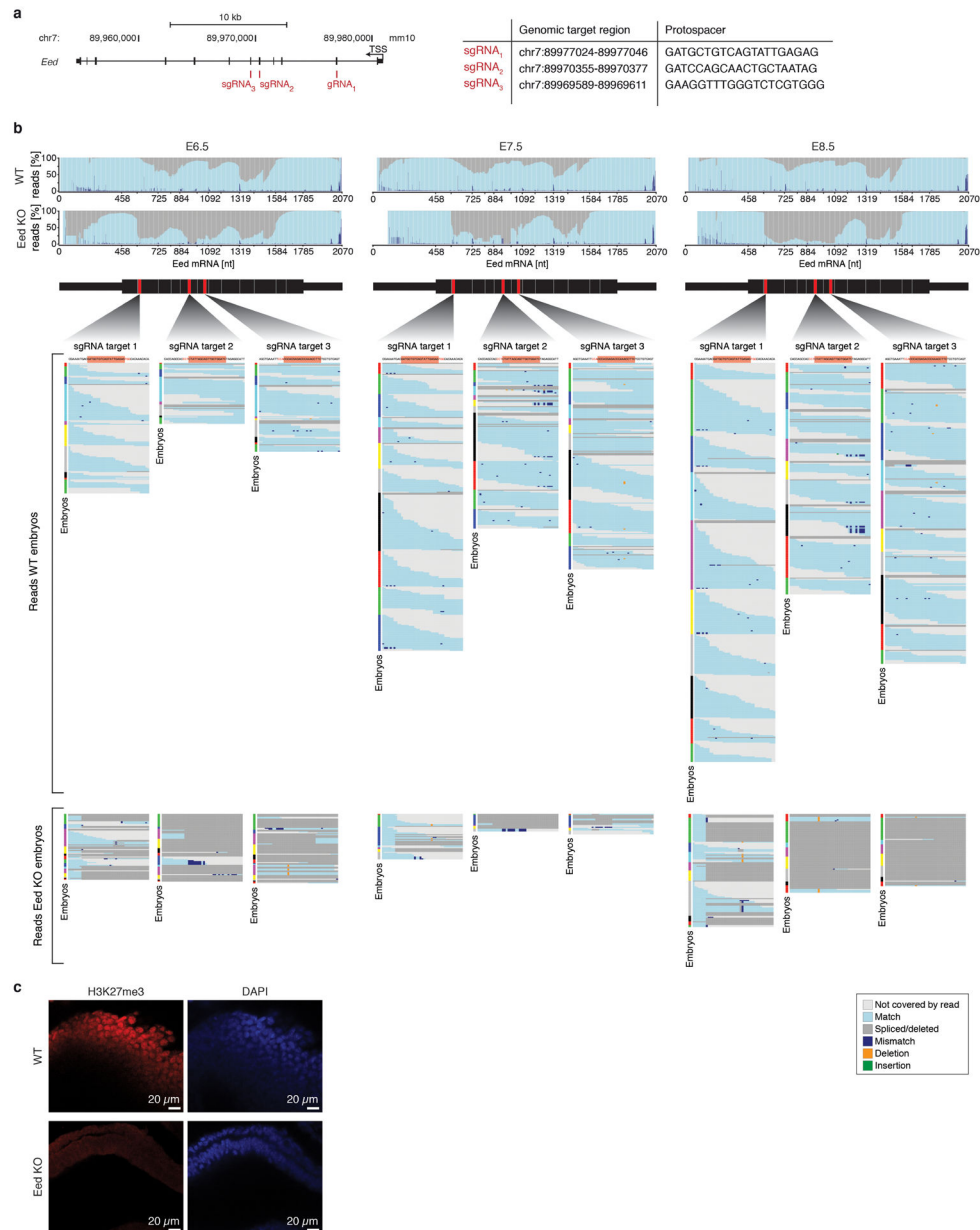
**Extended Data Figure 8. PRC1 and 2 converge to block non-CpG island hypermethylation within DNA methylation valleys (DMVs)**

**a.** Scatterplots of the difference between Eed KO and Rnf2 KO CpG island (CGI) methylation compared to WT for E6.5 epiblast (**Left**) and Xecto (**Right**), respectively. While overall Eed and Rnf2 KO share a similar DNA methylation landscape within the epiblast, we identify some regions where the Rnf2 KO is differentially methylated and the Eed KO more closely resembles WT. Eed KO shows a more substantial loss of CGI methylation specifically within Xecto, while Rnf2 KO shows 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 WT, Eed, Rnf2, or Kdm2b KO. 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-kb hypomethylated domains. However, in Eed and Rnf2 KO, non-CGI CpGs become hypermethylated while the CGI remain unmethylated. This trend is also observed for the Kdm2b KO but to a substantially lower 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 KO specifically causes loss of CGI methylation. Notably, Eed KO-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 WT, Eed, Rnf2, Kdm2b and L3mbtl2 KOs. Although this region is retained as a hypomethylated DMV in the WT and Eed KO epiblast, it is specifically methylated in PRC1 subunit KOs. In Xecto, the *Prdm14* promoter is naturally methylated but specifically unmethylated in Eed KO.



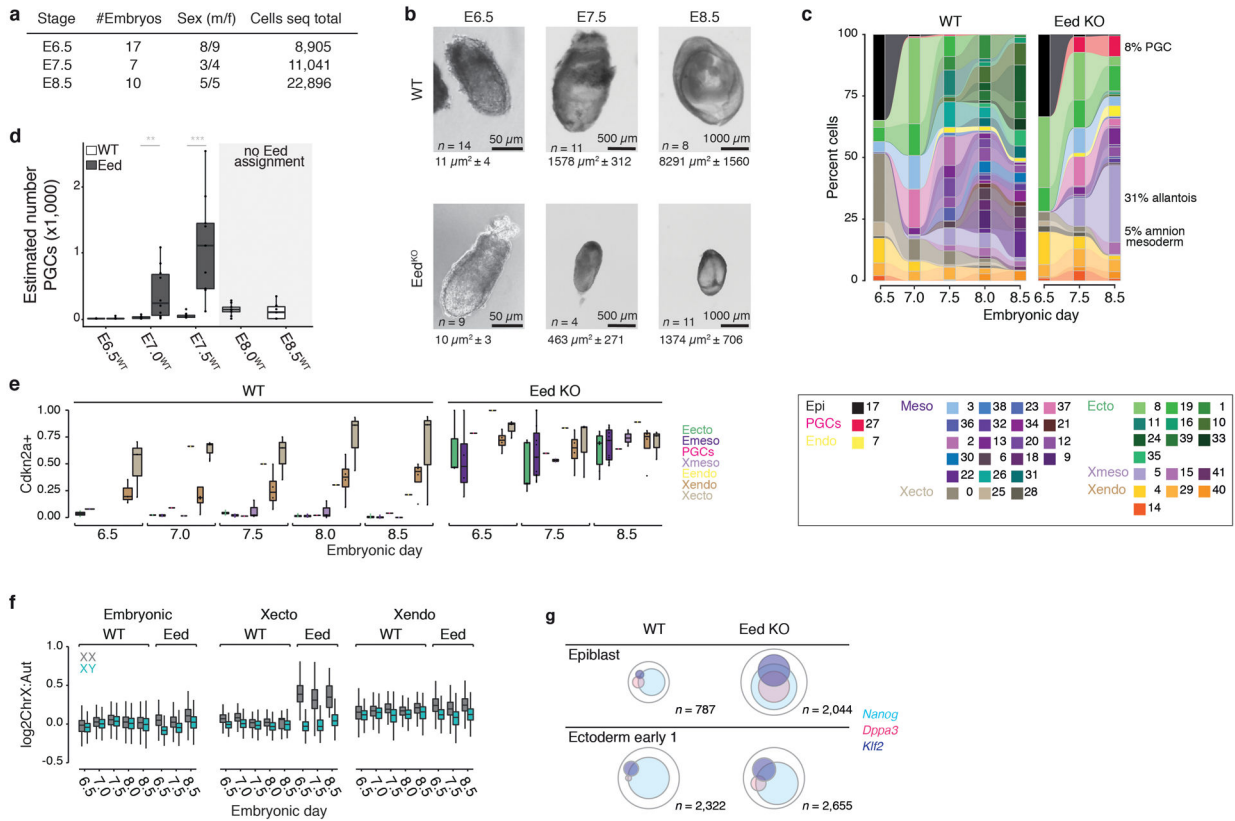
**Extended Data Figure 9. Efficient Cas9-mediated zygotic disruption of the *Eed* locus across an expanded time series**

**a.** Expanded description of our zygotic perturbation strategy for *Eed*. Three sgRNAs were designed to balance high efficiency cutting, off target potential, and coverage across the first half of the coding sequence (see 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, 7.5, 8.5 WT and *Eed* KO 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 the KO compared to the WT, reflecting alterations in the transcripts as a result of Cas9-mediated genetic disruption. **Middle:** Read-level analysis of our E6.5, 7.5, 8.5 WT embryo data.

Position 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 (color 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 color 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 WT embryos, with a subset covering the entire target site without mismatches, insertions or deletions. **Bottom:** Read-level analysis for our Eed KO samples from each time point. Compared to WT, a much lower number of reads from the Eed KO data match the target sites, likely 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 WT and Eed KO embryos. Single z-stack displaying an anterior region of size-matched WT (E7.5) and Eed KO (E8.5) embryos (H3K27me3, red; nuclei stained by DAPI, blue). The nuclear signal for H3K27me3 is readily detectable in WT but absent in Eed KO. Two independent experiments were conducted with similar results.



**Extended Data Figure 10. Developmental roles of PRC2 during gastrulation**

**a.** Our scRNA-seq profiled Eed KO series isolated at E6.5, 7.5, and 8.5. See Supplementary Tables 1 and 5 for information on sex-typing and cell state composition of individual embryos.

**b.** Representative WT and Eed KO embryos at gestational days E6.5, 7.5, and 8.5, with size information (image area occluded by an embryo in μm<sup>2</sup>, n = embryos imaged, all experiments had been replicated at least once, with similar results). Eed KO embryos initially appear similar to WT in size and morphology, but become substantially smaller and more variable in morphology, consistent with previous reports using transgenic models<sup>29,68,69</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>70</sup>). Complete Eed disruption is supported by the consistency of the resulting phenotype, as Eed<sup>+/-</sup> animals are viable and appear phenotypically normal during this period<sup>71</sup>. WT embryos shown here are from natural matings isolated at the same gestational age.

**c.** Connected barplots of median cell state composition across developmental stages for WT and Eed KO embryos, respectively. WT embryos rapidly increase in complexity, while Eed KOs 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.

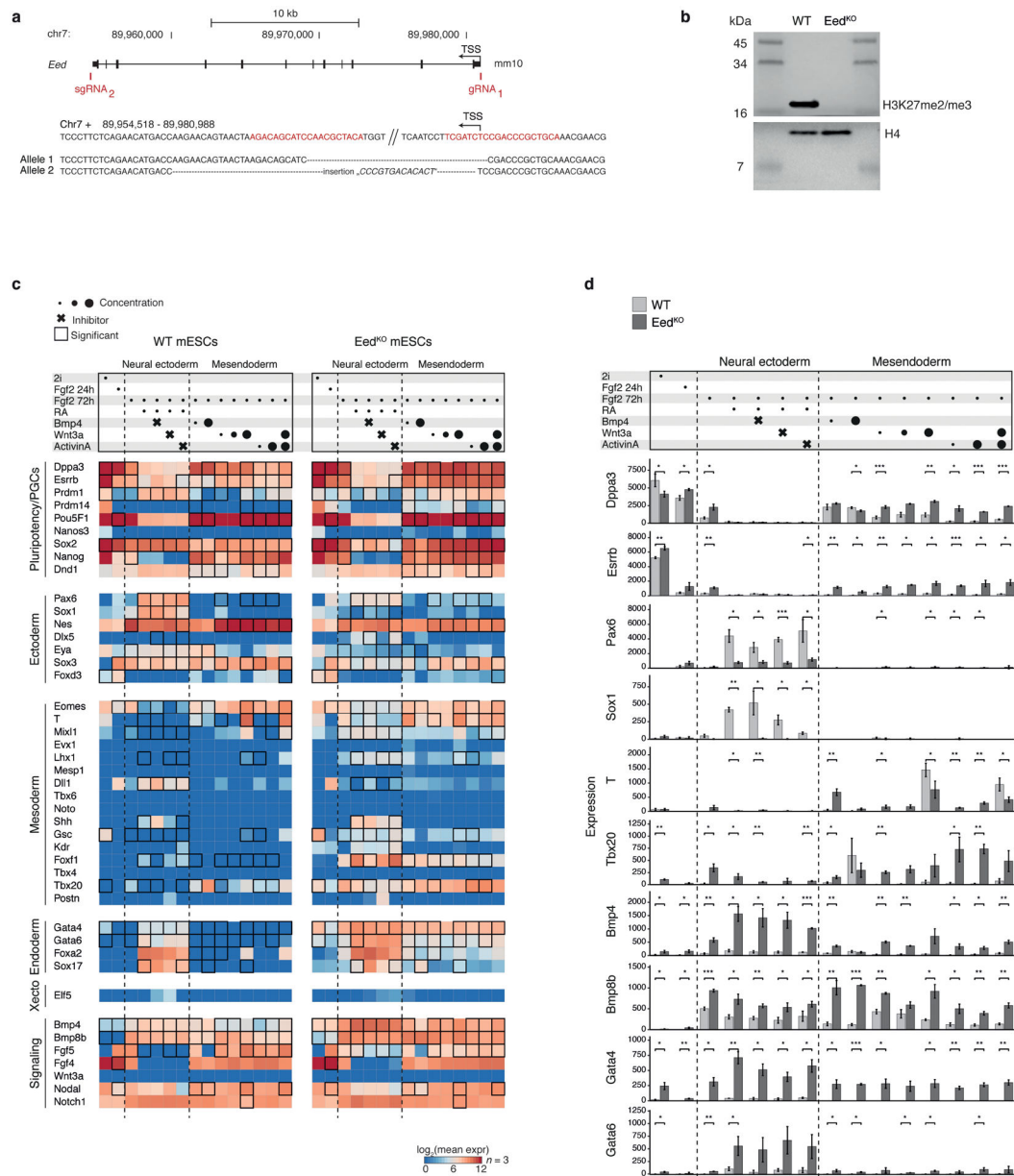
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

- d.** Absolute PGC numbers estimated for individual embryos (dots) show that the Eed KO overproduces PGCs beyond what is observed for WT over gastrulation. Eed KO embryos are presented after accounting for their developmental delay (i.e. PGC numbers of E8.5-isolated Eed KO 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 Cdkn2a transcript positive cells in recovered cell states (dots), shown per lineage across our WT and Eed KO time series. Cdkn2a is broadly derepressed across lineages in Eed KO from E6.5 onward. *n* = 10; 19; 1; 3; 1; 4; 3 cell states
- f.** Ratio of X chromosomal to autosomal transcripts for all male and female cells isolated across our WT and Eed KO time series, separated according to preimplantation lineage (embryonic, Xendo, and Xecto). Derepression of ChrX-linked genes happens as early as E6.5 in Eed KO females. Xecto becomes substantially underproduced in Eed KO females over time (*n* = 428, 295, and 19 female Eed KO cells from E6.5-E8.5). Xendo and embryonic lineages show increased ChrX 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>72,73</sup>. These transcripts are more abundant in Eed KO and more frequently present within the same cells, suggesting a PGC-supporting subnetwork within alternative lineages, possibly due to insufficient silencing prior to gastrulation.



**Extended Data Figure 11. *Eed*<sup>KO</sup> mouse ESC differentiation recapitulates many features of the *in vivo* mutant phenotype**

- a.** Generation of a homozygous *Eed*<sup>KO</sup> mESC line. The *Eed* gene was deleted in V6.5 mESCs 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 homozygous *Eed* deletion and depletion of H3K27 trimethylation. Histone 4 served as loading control.
- c.** Transcript counts of 44 genes associated with pluripotency, early germ layers, and the germline<sup>74</sup> over directed differentiation experiments from *Eed*<sup>KO</sup> mESCs. WT and *Eed*<sup>KO</sup> mESCs were maintained in conditions supporting a naïve, inner cell mass-like state (2i),

then subjected to low concentrations of bFGF for 24 h followed by culture in neural ectodermal and mesendodermal inducers for an additional 48 h. **Top:** The combination of signaling 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 bFGF; 0.25  $\mu$ M RA; 5 and 500 ng/ml BMP4; 10, 100, 1000 ng/ml WNT-3A; 10 and 1000 ng/ml ACTIVIN A; 0.5  $\mu$ M BMP4 pathway inhibitor LDN-193189; 3.3  $\mu$ M Wnt pathway inhibitor XAV939; 10  $\mu$ M TGF- $\beta$ /Activin/NODAL pathway inhibitor SB431542. **Bottom:** Heatmap of log<sub>2</sub>-transformed molecule counts (red being highly expressed) for WT and Eed<sup>KO</sup>, separately. Black tile frames indicate significant changes between WT and KO. During differentiation, many pluripotency factors associated with the germline remain expressed in Eed<sup>KO</sup> mESCs, especially within mesendodermal supporting conditions. Many mesodermal genes are also induced in Eed<sup>KO</sup> in neural ectodermal supporting conditions. Retinoic Acid (RA) treatment directs a fraction of WT mESCs to an extraembryonic endodermal fate<sup>75</sup>. This appears to be favored in Eed<sup>KO</sup> mESCs, where many Xendo-associated genes are particularly sensitive to RA. Finally, many regulators of the endoderm, early mesendoderm and extraembryonic mesoderm, such as Gata4, Tbx20 and Bmp4, are broadly expressed in Eed<sup>KO</sup> mESCs already in 2i/LIF.  $n = 3$  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.** Barplots 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. Two-sided  $t$ -test, \* <0.05, \*\* <0.01, \*\*\*<0.001, bars show the mean of  $n = 3$  experimental replicates and error bars represent standard deviation.  $P$ -values are provided in the Source Data file Nanostring\_pval.tsv.

## Acknowledgements

We thank Adriano Bolondi, Raha Weigert and other members of the Meissner laboratory, Michelle Chan and Denes Hnisz for discussions and advice, Sabine Otto for experimental support characterizing the Eed KO mESC line, Maria Walter for support with embryo isolations, Tobias Ahsendorf for help with initial efforts to optimize our genotyping pipeline, and Daniel Andergassen for discussions on SNP-typing. We are also grateful to Frederick Koch and the transgenic facility, including Miriam Peetz for their feedback and support. We thank Prof. Mitunori 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). Funding: This work was funded by the NIH (1P50HG006193, P01GM099117, 1R01HD078679 and 1DP3K111898) and the Max Planck Society.

## References

1. Hemberger M, Dean W & Reik W Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. *Nature Publishing Group* 10, 526–537 (2009).
2. Meissner A Epigenetic modifications in pluripotent and differentiated cells. *Nat. Biotechnol* 28, 1079–1088 (2010). [PubMed: 20944600]



3. Surani MA, Hayashi K & Hajkova P Genetic and epigenetic regulators of pluripotency. *Cell* 128, 747–762 (2007). [PubMed: 17320511]
4. Rivera-Pérez JA & Hadjantonakis A-K The Dynamics of Morphogenesis in the Early Mouse Embryo. *Cold Spring Harb Perspect Biol* 7, a015867 (2014). [PubMed: 24968703]
5. Plass M et al. Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science* 360, eaq1723 (2018). [PubMed: 29674432]
6. Briggs JA et al. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* 360, eaar5780 (2018). [PubMed: 29700227]
7. Wagner DE et al. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* 360, 981–987 (2018). [PubMed: 29700229]
8. Farrell JA et al. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* 108, eaar3131 (2018).
9. Scialdone A et al. Resolving early mesoderm diversification through single-cell expression profiling. *Nature* 535, 289–293 (2016). [PubMed: 27383781]
10. Peng G et al. Spatial Transcriptome for the Molecular Annotation of Lineage Fates and Cell Identity in Mid-gastrula Mouse Embryo. *Developmental Cell* 36, 681–697 (2016). [PubMed: 27003939]
11. Mohammed H et al. Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. *Cell Rep* 20, 1215–1228 (2017). [PubMed: 28768204]
12. 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* 58, 598 (2018).
13. Peng G et al. Molecular architecture of lineage allocation and tissue organization in early mouse embryo. *Nature* 572, 528–532 (2019). [PubMed: 31391582]
14. Pijuan-Sala B et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* 566, 490 (2019). [PubMed: 30787436]
15. Nowotschin S et al. The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* 569, 361–367 (2019). [PubMed: 30959515]
16. Chan MM et al. Molecular recording of mammalian embryogenesis. *Nature* 570, 77–82 (2019). [PubMed: 31086336]
17. Cao J et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 566, 496 (2019). [PubMed: 30787437]
18. Argelaguet R et al. Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* 576, 487–491 (2019). [PubMed: 31827285]
19. 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). [PubMed: 17477386]
20. Laugesen A & Helin K Chromatin repressive complexes in stem cells, development, and cancer. *Cell Stem Cell* 14, 735–751 (2014). [PubMed: 24905164]
21. Piunti A & Shilatifard A Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science* 352, aad9780 (2016). [PubMed: 27257261]
22. Glaser S et al. Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development* 133, 1423–1432 (2006). [PubMed: 16540515]
23. Rossant J, Chazaud C & Yamanaka Y Lineage allocation and asymmetries in the early mouse embryo. *Philos. Trans. R. Soc. Lond., B, Biol. Sci* 358, 1341–8– discussion 1349 (2003). [PubMed: 14511480]
24. Smith ZD et al. Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. 1–29 (2017). doi:10.1038/nature23891
25. 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). [PubMed: 16921403]
26. Boulard M, Edwards JR & Bestor TH FBXL10 protects Polycomb-bound genes from hypermethylation. *Nat Genet* 47, 479–485 (2015). [PubMed: 25848754]
27. Walsh CP, Chaillet JR & Bestor TH Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* 20, 116–117 (1998). [PubMed: 9771701]

28. 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). [PubMed: 22770845]
29. Faust C, Schumacher A, Holdener B & Magnuson T The eed mutation disrupts anterior mesoderm production in mice. *Development* 121, 273–285 (1995). [PubMed: 7768172]
30. Voncken JW et al. Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc Natl Acad Sci USA* 100, 2468–2473 (2003). [PubMed: 12589020]
31. Yamaji M et al. Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nat Genet* 40, 1016–1022 (2008). [PubMed: 18622394]
32. ylicz JJ et al. The Implication of Early Chromatin Changes in X Chromosome Inactivation. *Cell* 176, 182–197.e23 (2019). [PubMed: 30595450]
33. Wang J et al. Imprinted X inactivation maintained by a mouse Polycomb group gene. *Nat Genet* 28, 371–375 (2001). [PubMed: 11479595]
34. Li Y et al. Genome-wide analyses reveal a role of Polycomb in promoting hypomethylation of DNA methylation valleys. *Genome Biology* 19, 18–16 (2018). [PubMed: 29422066]
35. Leitch HG & Smith A The mammalian germline as a pluripotency cycle. *Development* 140, 2495–2501 (2013). [PubMed: 23715543]
36. Forlani S, Lawson KA & Deschamps J Acquisition of Hox codes during gastrulation and axial elongation in the mouse embryo. *Development* 130, 3807–3819 (2003). [PubMed: 12835396]
37. Saitou M Specification of the germ cell lineage in mice. *Front Biosci (Landmark Ed)* 14, 1068–1087 (2009). [PubMed: 19273116]
38. Nicetto D et al. H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. *Science* 363, 294–297 (2019). [PubMed: 30606806]
39. Tzouanacou E, Wegener A, Wymeersch FJ, Wilson V & Nicolas J-F Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis. *Developmental Cell* 17, 365–376 (2009). [PubMed: 19758561]
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). [PubMed: 23643243]
41. Platt RJ et al. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* 159, 440–455 (2014). [PubMed: 25263330]
42. Montague TG, Cruz JM, Gagnon JA, Church GM & Valen E CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res* 42, W401–7 (2014). [PubMed: 24861617]
43. Nakagata N Cryopreservation of mouse spermatozoa and in vitro fertilization. *Methods Mol Biol* 693, 57–73 (2011). [PubMed: 21080274]
44. Ying Q-L & Smith AG Defined conditions for neural commitment and differentiation. *Meth Enzymol* 365, 327–341 (2003). [PubMed: 14696356]
45. Gu Z, Eils R & Schlesner M Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849 (2016). [PubMed: 27207943]
46. Hahne F & Ivanek R Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol Biol* 1418, 335–351 (2016). [PubMed: 27008022]
47. Dobin A et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21 (2013). [PubMed: 23104886]
48. Krueger F & Andrews SR SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes. *F1000Res* 5, 1479 (2016). [PubMed: 27429743]
49. 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). [PubMed: 29608179]
50. Andergassen D et al. Mapping the mouse Allelome reveals tissue-specific regulation of allelic expression. *eLife* 6, e146 (2017).
51. Ritchie ME et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43, e47–e47 (2015). [PubMed: 25605792]

52. Wang C et al. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat Cell Biol* 20, 620–631 (2018). [PubMed: 29686265]
53. McInnes L, Healy J & Melville J UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. (2018).
54. La Manno G et al. RNA velocity of single cells. *Nature* 1–25 (2018). doi:10.1038/s41586-018-0414-6
55. Wolf FA, Angerer P & Theis FJ SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* 19, 15–5 (2018). [PubMed: 29409532]
56. Sun D et al. MOABS: model based analysis of bisulfite sequencing data. *Genome Biology* 15, R38–12 (2014). [PubMed: 24565500]
57. Keane TM et al. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477, 289–294 (2011). [PubMed: 21921910]
58. Lei H et al. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* 122, 3195–3205 (1996). [PubMed: 8898232]
59. Okano M, Bell DW, Haber DA & Li E DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99, 247–257 (1999). [PubMed: 10555141]
60. 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 & Development* 16, 1779–1791 (2002). [PubMed: 12130538]
61. YU BD, HESS JL, HORNING SE, Brown GA & KORSMEYER SJ Altered Hox expression and segmental identity in Mll-mutant mice. *Nature* 378, 505–508 (1995). [PubMed: 7477409]
62. Hammoud SS et al. Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis. *Cell Stem Cell* 15, 239–253 (2014). [PubMed: 24835570]
63. Smallwood SA et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods* 11, 817–820 (2014). [PubMed: 25042786]
64. 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). [PubMed: 26549683]
65. Hon GC et al. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat Genet* 45, 1198–1206 (2013). [PubMed: 23995138]
66. 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 Biology* 15, 545 (2014). [PubMed: 25476147]
67. Wymeersch FJ et al. Position-dependent plasticity of distinct progenitor types in the primitive streak. *eLife* 5, (2016).
68. Niswander L, Yee D, Rinchik EM, Russell LB & Magnuson T The albino deletion complex and early postimplantation survival in the mouse. *Development* 102, 45–53 (1988). [PubMed: 3416773]
69. Faust C, Lawson KA, Schork NJ, 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). [PubMed: 9778508]
70. Kalantry S & Magnuson T The Polycomb group protein EED is dispensable for the initiation of random X-chromosome inactivation. *PLoS Genet* 2, e66 (2006). [PubMed: 16680199]
71. Niswander L, Yee D, Rinchik EM, Russell LB & Magnuson T The albino-deletion complex in the mouse defines genes necessary for development of embryonic and extraembryonic ectoderm. *Development* 105, 175–182 (1989). [PubMed: 2806116]
72. Han J et al. Tbx3 improves the germ-line competency of induced pluripotent stem cells. *Nature* 463, 1096–1100 (2010). [PubMed: 20139965]
73. Magnúsdóttir E & Surani MA How to make a primordial germ cell. *Development* 141, 245–252 (2014). [PubMed: 24381195]
74. Arnold SJ & Robertson EJ Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat. Rev. Mol. Cell Biol* 10, 91–103 (2009). [PubMed: 19129791]

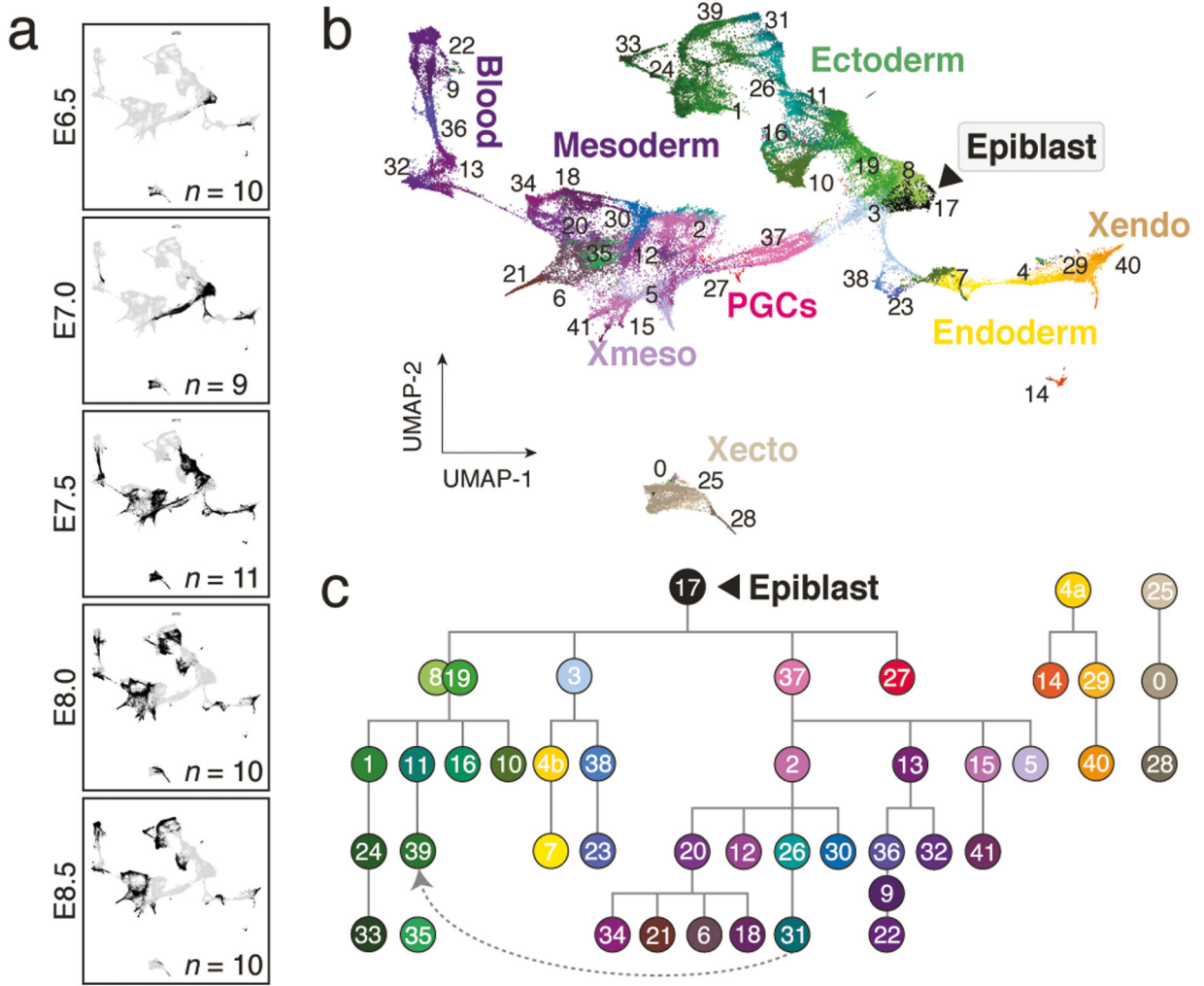
75. Semrau S et al. Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. *Nat Commun* 8, 1096 (2017). [PubMed: 29061959]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



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

**a.** Uniform Manifold Approximation and Projection (UMAP) plots of 88,779 WT single-cell transcriptomes, separated by time point (black dots). *n* = replicate embryos

**b.** UMAP of our WT scRNA-seq time series from panel **a**. Numbers denote the 42 cell states, colors 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 and Supplementary Text). Dashed arrow represents neuromesodermal progenitor (NMP)-containing states that are reported to contribute to neuroectoderm<sup>39</sup>. Extraembryonic ectoderm and endoderm are disconnected to reflect their preimplantation origins<sup>23</sup>.

0:extraembryonic ectoderm late, 1:neural ectoderm anterior, 2:primitive streak late, 3:streak pre-specified|anterior, 4:endoderm primitive(a)|definitive(b), 5:allantois, 6:2° heart field|splanchnic lateral plate, 7:gut, 8:ectoderm early 1, 9:primitive blood early, 10:preplacodal ectoderm, 11:neural ectoderm progenitor, 12:posterior lateral plate mesoderm, 13:hematopoietic|endothelial progenitor, 14:parietal endoderm, 15:amnion mesoderm early, 16:surface ectoderm, 17:epiblast, 18:somites, 19:ectoderm early 2, 20:splanchnic-lateral|

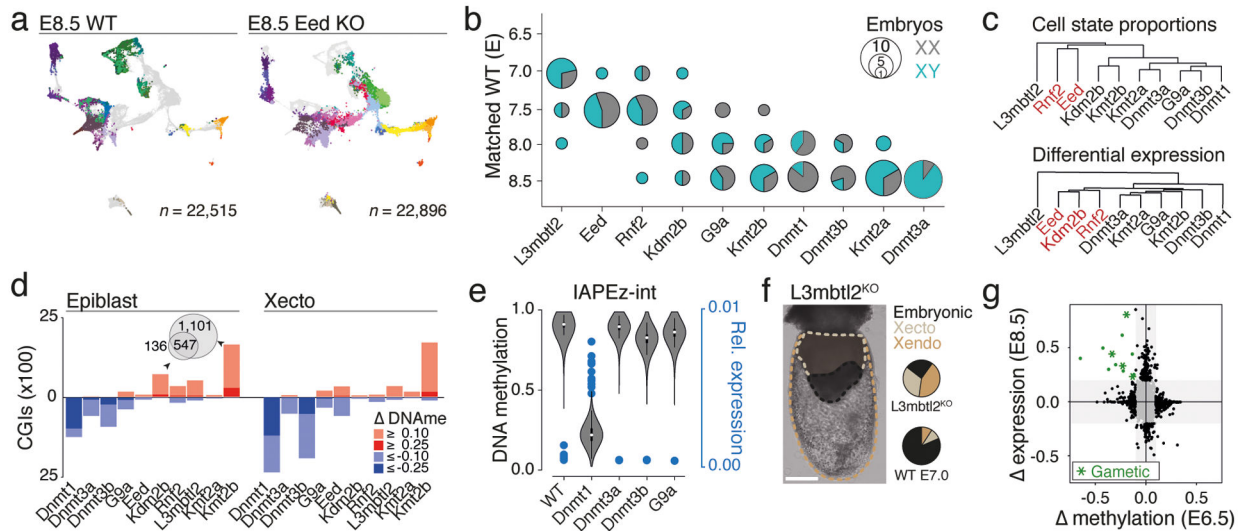
anterior-paraxial mesoderm, 21:primitive heart tube, 22:primitive blood late, 23:notochord, 24:fore|midbrain, 25:extraembryonic ectoderm early, 26:NMPs early, 27:PGCs, 28:differentiated trophoblasts, 29:visceral endoderm early, 30:presomitic mesoderm, 31:NMPs late, 32:angioblasts, 33:neural crest, 34:pharyngeal 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.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 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. KO cells were assigned to WT cell states (colors) and projected onto our gastrulation UMAP (see Extended Data Fig. 3a for all KOs).

**b.** Developmental staging according to cell state composition. Circle size denotes KO embryo number assigned to a given WT stage (y-axis). Colors indicate sex.

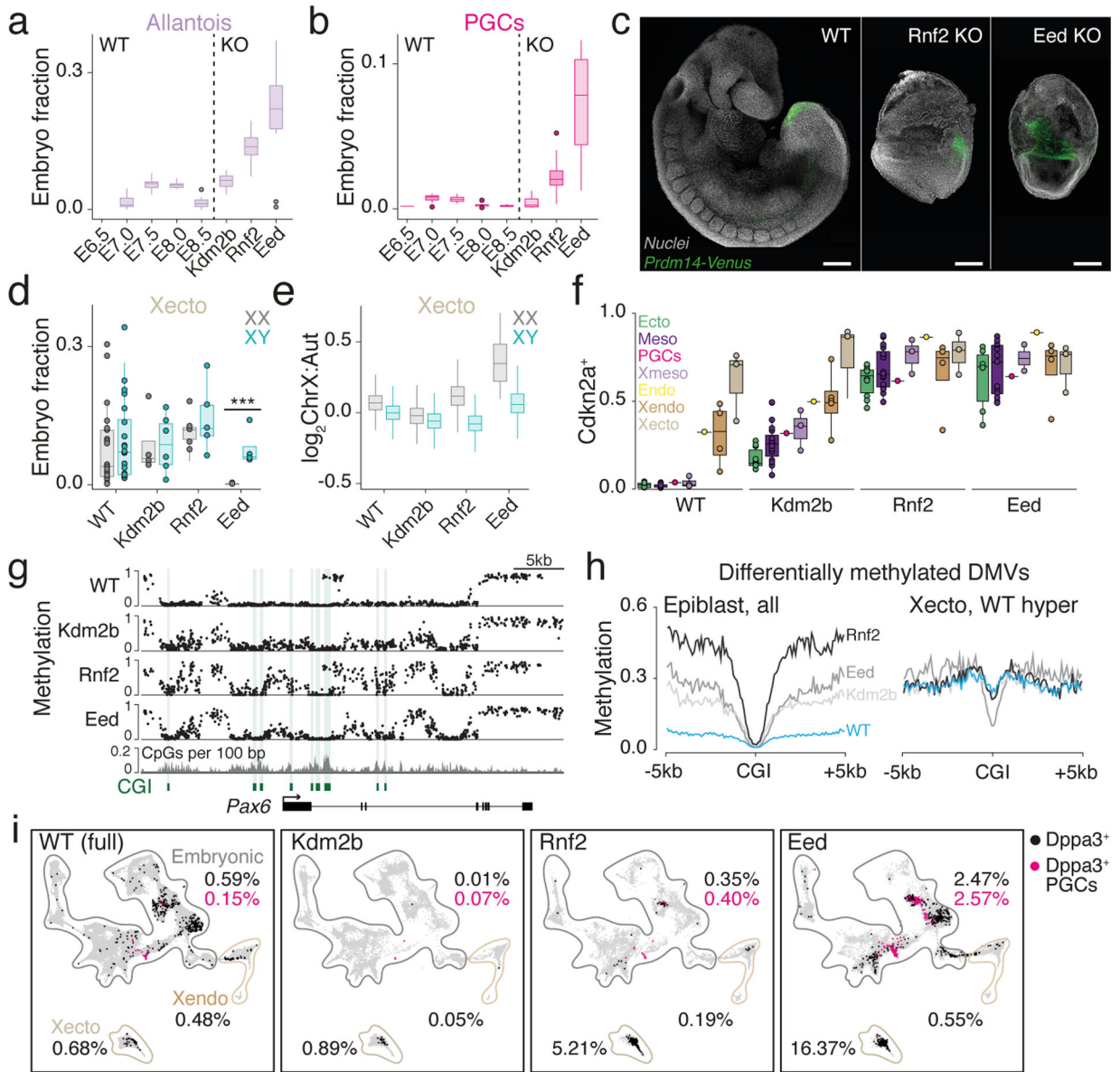
**c.** Hierarchical clustering of KOs based on composition (**Top**) and transcriptional deregulation (**Bottom**) of cell states compared to matching WT stages.

**d.** CGI methylation across our KOs at E6.5. Hyper- or hypomethylation in comparison to WT for absolute changes  $\geq 0.1$  and  $\geq 0.25$ . Dnmt1 KO shows the greatest loss in the epiblast, both Dnmt1 and 3b KOs show large effects in Xecto. Kmt2b and Kdm2b KO show substantial and overlapping CGI methylation in epiblast (Venn diagram), while Xecto is only affected by Kmt2b.

**e.** DNA methylation (violin plots) of IAPs in E6.5 epiblast and per embryo IAP expression (blue dots) as fraction of reads per cell in E8.5 embryonic lineages. White dots, median; edges, the IQR; and whiskers, 1.5xIQR.

**f.** Representative E8.5-isolated L3mbtl2 KO 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 as calculated by scRNA-seq compared to stage-matched WT. Xecto and Xendo are overabundant, while embryonic lineages are substantially impeded. Scale bar, 200 $\mu$ m.

**g.** Scatterplot of changes in E6.5 epiblast promoter methylation (x-axis) and E8.5 embryonic expression (y-axis) for L3mbtl2 KO compared to WT. Green indicates genes that lose promoter methylation ( $\geq 0.1$ ) and increase expression ( $\geq 0.2$  fraction of positive cells). Asterisks, genes that function in gametogenesis.  $n = 13$  gene promoters.



**Figure 3. Phenotypic and molecular abnormalities of PRC regulator mutants**

**a.-b.** Fraction of cells assigned to the allantois (**a**) and PGC state (**b**) per embryo. Dots, outliers;  $n=10;9;11;10;10;11;10$  embryos, left to right

**c.** *Prdm14* reporter<sup>31</sup> activity in representative E8.5-isolated WT, Rnf2 and Eed KO embryos (from total of 4,7,5 embryos, respectively). Scale bars, 200  $\mu$ m.

**d.** Per embryo fractions assigned to Xecto cell states, separated by sex. \*\*\*,  $P$  0.001, two-sided Wilcoxon test,  $n=25;25;4;6;6;5;5$  embryos and  $P$ -value=0.1276;0.9118;0.7618;0.0001 left to right

**e.** ChrX to autosome transcript ratios per Xecto cell, separated by sex. Outliers omitted;  $n=1,769;3,685;755;1,372;1,465;1,220;19;1,594$  cells

**f.** Boxplots of the PRC target *Cdkn2a*, shown as the fraction of positive cells for each state (dots) grouped by lineage (colors). In WT, expression is limited to Xecto and Xendo, with



mixed signal in endoderm (Endo) potentially reflecting its heterogeneous origins<sup>16</sup>.  
 $n=10;19;1;3;1;3;4$  cell states, a,b,d-f boxes, median and quantiles; whiskers, 1.5xIQR

**g.** Developmental gene associated DNA methylation valleys (DMVs) gain methylation in PRC KO in E6.5 epiblast. CpG resolution genome browser tracks of WGBS for the neuroectodermal regulator *Pax6*. CGIs and CpG density are provided.

**h.** Median CpG methylation centered on CGIs within differentially methylated DMVs for E6.5 epiblast (**Left**) and Xecto (**Right**). CGIs that are normally methylated in the Xecto (WT) do not acquire *de novo* methylation in Eed KO (Supplementary Table 8).

**i.** Dppa3-positive cells over our WT time series (E6.5–8.5) and in PRC KOs, 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  $n=18,723; 1,613; 2,560$



**e.** Select mesodermal lineages and PGCs as they stem from the pluripotent epiblast. Grey-scale indicates fraction of  $Hoxb1^+$  and  $Hoxd9^+$  cells. Eed KO embryos induce  $Hoxd9$  prematurely, leading to a profile that resembles extraembryonic lineages. Differential up- or downregulation is highlighted in red or blue, respectively. States that are not produced are dashed.

**f.**  $\log_2$ -fold change between KO and WT for select genes, taken from a total of 44 profiled using Nanostring ( $n = 3$  experimental replicates, Extended Data Fig. 11, Supplementary Table 11). Top rows: circles, morphogen concentrations; crosses, inhibitors. Grey boxes, expression below threshold in both samples.