

Regulation of embryonic haematopoietic multipotency by EZH1

Linda T. Vo^{1,2,3}, Melissa A. Kinney^{1,2}, Xin Liu⁴, Yuannyu Zhang^{4,5}, Jessica Barragan^{1,2}, Patricia M. Sousa^{1,2}, Deepak K. Jha^{1,2}, Aream Han^{1,2}, Marcella Cesana^{1,2}, Zhen Shao⁵, Trista E. North⁶, Stuart H. Orkin^{2,3,7}, Sergei Doulatov⁸, Jian Xu⁴ & George Q. Daley^{1,2,3}

All haematopoietic cell lineages that circulate in the blood of adult mammals derive from multipotent haematopoietic stem cells (HSCs)¹. By contrast, in the blood of mammalian embryos, lineage-restricted progenitors arise first, independently of HSCs, which only emerge later in gestation^{2,3}. As best defined in the mouse, 'primitive' progenitors first appear in the yolk sac at 7.5 days post-coitum^{2,3}. Subsequently, erythroid–myeloid progenitors that express fetal haemoglobin⁴, as well as fetal lymphoid progenitors⁵, develop in the yolk sac and the embryo proper, but these cells lack HSC potential. Ultimately, 'definitive' HSCs with long-term, multilineage potential and the ability to engraft irradiated adults emerge at 10.5 days post-coitum from arterial endothelium in the aorta-gonad-mesonephros and other haemogenic vasculature³. The molecular mechanisms of this reverse progression of haematopoietic ontogeny remain unexplained. We hypothesized that the definitive haematopoietic program might be actively repressed in early embryogenesis through epigenetic silencing⁶, and that alleviating this repression would elicit multipotency in otherwise lineage-restricted haematopoietic progenitors. Here we show that reduced expression of the Polycomb group protein EZH1 enhances multi-lymphoid output from human pluripotent stem cells. In addition, *Ezh1* deficiency in mouse embryos results in precocious emergence of functional definitive HSCs *in vivo*. Thus, we identify EZH1 as a repressor of haematopoietic multipotency in the early mammalian embryo.

The differentiation of pluripotent stem cells to haematopoietic lineages generates robust erythroid–myeloid lineage-restricted progenitors but not HSCs. This pattern bears marked similarities to early haematopoietic ontogeny. We hypothesized that the same epigenetic factors actively repress multipotency in embryogenesis and differentiation from pluripotent stem cells. To identify these factors, we adopted a loss-of-function screen using lentivirally delivered short hairpin RNAs (shRNAs) that target 20 DNA- and histone-modifying factors (Extended Data Fig. 1a, Supplementary Table 1). Erythroid–myeloid progenitors differentiated from human pluripotent stem cells marked by CD34 and CD45 were expanded with five transcription factors (5F). They retained embryonic features, including lack of lymphoid potential⁷, and this enabled us to screen for reactivation of lymphoid potential as a measure of multipotency. 5F cells were transduced with individual shRNAs and screened for T cell potential on OP9-DL1 stromal cells (Fig. 1a). The knockdown of six factors independently enhanced CD4⁺CD8⁺ T cell potential from 5F cells (Fig. 1b, Extended Data Fig. 1b).

Prospective validation revealed that only *EZH1* knockdown (using shEZH1) elicited robust T (16.3 ± 7.4%; mean ± s.e.m.) and B (22.5 ± 7.3%) cell potential (Fig. 1c–e), compared to shRNAs targeting a control luciferase gene (shLUC) (T cell 0.002 ± 0.002%; B cell

0.022 ± 0.006%) across multiple induced pluripotent stem (iPS) cell lines (Fig. 1f). *EZH1*-deficient cells retained erythroid–myeloid potential as shown by colony-forming assays (Fig. 1g) and flow cytometry (Fig. 1h, i). *EZH1* knockdown also promoted lymphoid potential independently of the five transcription factors, as evidenced by robust T cell differentiation from naive CD34⁺ haemogenic endothelial cells (26.1 ± 16.5% shEZH1 versus 2.3 ± 0.4% shLUC) (Extended Data Fig. 1c). Further characterization was not possible owing to the limited proliferation of pluripotent stem and haemogenic endothelial cells. By contrast, 5F cells expanded exponentially (Extended Data Fig. 1d) and showed increased CD34⁺ progenitors after shEZH1 transduction (78.8 ± 14.2% versus 29.3 ± 10.0%) (Extended Data Fig. 1e). Taken together, these data show that *EZH1* knockdown activates multipotency in lineage-restricted embryonic haematopoietic progenitors.

EZH1 is a component of the Polycomb repressive complex 2 (PRC2), which mediates epigenetic silencing of genes via methylation of lysine residue 27 of histone H3⁸. To dissect the role of PRC2 in repressing haematopoietic multipotency, we assessed T cell differentiation upon depletion of each PRC2 subunit. In addition to *EZH1*, *SUZ12* knockdown also enhanced T cell potential, albeit to a lesser extent. By contrast, knockdown of *EED* or *EZH2* had no effect on T cell potential and dual *EZH1* and *EZH2* knockdown phenocopied that of *EZH2* depletion (Fig. 2a, b). To determine whether the catalytic SET domain was required, we overexpressed full-length mouse *Ezh1* or mutant *Ezh1* lacking the SET domain (mEzh1ΔSET) (Fig. 2c). Overexpression of mouse *Ezh1* completely abrogated T cell potential in shEZH1 cells, whereas the mutant mEzh1ΔSET did not (Fig. 2c, d, Extended Data Fig. 2d–g). Furthermore, overexpression of mouse *Ezh2* failed to suppress T cell potential, despite the remarkable homology of the SET domains (Extended Data Fig. 2e, h, i). These data show that specific inhibition of *EZH1*, rather than antagonism of canonical PRC2, unlocks lymphoid potential and the catalytic SET domain is required for this function.

To understand the molecular changes upon *EZH1* knockdown, we performed RNA sequencing (RNA-seq), assay for transposase-accessible chromatin using sequencing (ATAC-seq) and chromatin immunoprecipitation followed by sequencing (ChIP-seq). Upregulated genes after *EZH1* knockdown were enriched for biological processes such as defence response ($P = 6.8 \times 10^{-9}$), immune response ($P = 1.2 \times 10^{-7}$) and T cell co-stimulation ($P = 0.03$) (Fig. 3a, b). Human haematopoietic gene signatures⁹, such as of HSCs (stem), multi-lymphoid progenitors (MLP) (early lymphoid) and ProB, were highly enriched in shEZH1 cells, consistent with stem and lymphoid potential (Fig. 3c). We also performed RNA-seq and ATAC-seq on emergent haematopoietic stem and progenitor cells (HSPCs) at 10.5 days post-coitum^{10–12} from the yolk sac and aorta-gonad-mesonephros (AGM) of wild-type, *Ezh1*^{+/-} and

¹Stem Cell Program, Boston Children's Hospital, Boston, Massachusetts, USA. ²Division of Hematology/Oncology, Boston Children's Hospital and Dana Farber Cancer Institute, Boston, Massachusetts, USA. ³Harvard Medical School, Boston, Massachusetts, USA. ⁴Children's Medical Center Research Institute, Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, Texas, USA. ⁵Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. ⁶Department of Pathology, Beth Israel-Deaconess Medical Center, Boston, Massachusetts, USA. ⁷Howard Hughes Medical Institute, Boston, Massachusetts, USA. ⁸Division of Hematology, Department of Medicine, University of Washington, Seattle, Washington, USA.

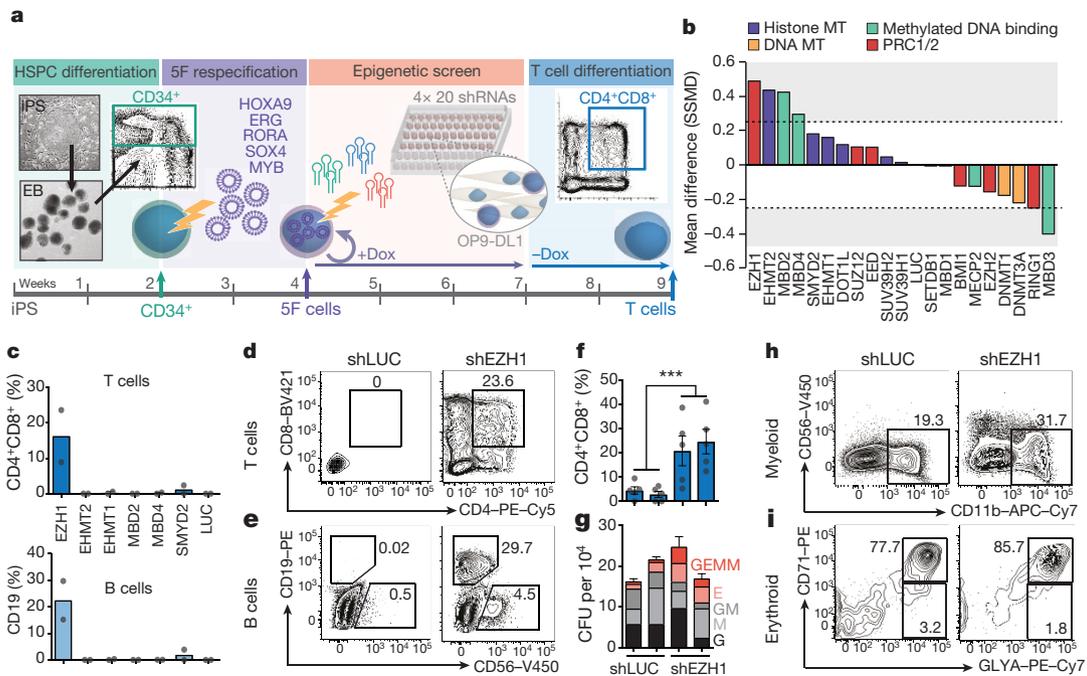


Figure 1 | *In vitro* screen for epigenetic modifiers that restrict lymphoid potential. **a**, Scheme for human pluripotent stem-cell differentiation into haematopoietic progenitors. CD34⁺ cells were transduced with the five transcription factors (5F) HOXA9, ERG, RORA, SOX4 and MYB. 5F cells were then transduced with individual shRNAs (×4 each) that targeted each epigenetic modifier and seeded onto OP9-DL1 stromal cells to induce T cell differentiation. Dox, doxycycline; EB, embryoid body. **b**, Strictly standardized mean difference (SSMD) of CD4⁺CD8⁺ T cell frequencies across all four shRNAs targeting each epigenetic modifier in 5F cells using two iPS cell lines, CD45-iPS and MSC-iPS1, in two independent experiments. MT, methyltransferase. **c**, Prospective analysis of T and B cell frequencies from 5F cells plus shRNA targeting top candidates ($n = 2$ biological replicates). **d**, Flow analysis of CD4⁺CD8⁺

T cell development of 5F cells with shRNAs targeting luciferase (shLUC) or *EZH1* (shEZH1) after 5 weeks of differentiation on OP9-DL1 stromal cells. **e**, Flow analysis of CD19⁺ B cell potential. **f**, Quantification of T cell potential of 5F plus shEZH1 cells compared to 5F plus shLUC cells pooled across two hairpins and five independent experiments ($n = 10$) using several iPS cell lines (CD34-iPS, CD45-iPS and MSC-iPS1). Individual values obtained for each hairpin are shown in the Source Data. *** $P = 0.001$ by unpaired two-tailed t -test. **g**, Quantification of colony-forming potential in three independent experiments. E, erythroid; GM, granulocyte, monocyte; M, monocyte; G, granulocyte; GEMM, granulocyte, erythroid, monocyte, megakaryocyte. **h, i**, Flow analysis of myeloid (CD11b⁺) (h) and erythroid (CD71⁺GLYA⁺) (i) potential. Experiments replicated at least twice. Data are mean \pm s.e.m.

Ezh1^{-/-} mouse embryos (Fig. 4a). Interestingly, in wild-type embryos, the expression of *Ezh1* was lower in the AGM than in the yolk sac, whereas *Ezh2* and *Eed* were higher in the AGM (Fig. 4b). Notably, *Ezh1* deficiency *in vivo* also induced genes enriched for angiogenesis, haematopoietic/lymphoid development and immune system processes (Extended Data Fig. 3a–d).

Regions of increased chromatin accessibility (1,610 ATAC peaks) in shEZH1 cells exhibited concomitantly increased gene expression upon *EZH1* knockdown and were associated with T cell development and lymphocyte activation pathways, as well as HSC, HSC/MLP, B and T cell signatures (Fig. 3d, e, Extended Data Fig. 3e–g). *EZH1* knockdown also increased accessibility to HSC/lymphoid transcription factors, such as *HLF*, *FOXO1* and *ARID5B*^{13–15} (Fig. 3f). Downregulated peaks were enriched for alternative developmental processes and importantly, embryonic haematopoiesis (Fig. 3e, Extended Data Fig. 3e). *In vivo*, upregulated ATAC peaks in *Ezh1*-deficient AGM cells were enriched for immune response, T cell activation, lymphocyte differentiation pathways, as well as HSC and HSC/MLP signatures (Fig. 4a, c, d, Extended Data Fig. 3h, i); furthermore, *Ezh1* deficiency increased accessibility to target genes of master haematopoietic transcription factors, including *Runx1* (Extended Data Fig. 3k, l).

We hypothesized that these molecular changes upon *EZH1* knockdown were mediated by bivalent, or poised, chromatin domains, often implicated in the control of developmentally regulated genes¹⁶. Consistent with previous reports, *EZH1* was broadly associated with repressive (H3K27me3), bivalent (H3K27me3 and H3K4me3) and active (H3K4me3) histone methylation marks^{17,18} (Fig. 3g, Extended Data Fig. 4a). Although active genes were associated with housekeeping functions (Extended Data Fig. 4b), *EZH1*-bound bivalent and

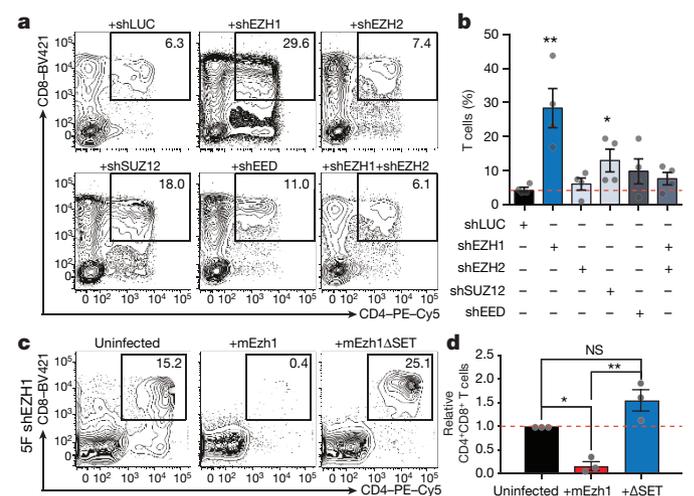


Figure 2 | Repression of canonical PRC2 subunits does not activate lymphoid potential. **a**, Representative flow plots of T cell potential of 5F cells with shRNAs targeting individual components of PRC2. **b**, Quantification of T cell potential of 5F cells plus shRNA targeting the indicated subunit in **a**, shown as using two hairpins across two independent experiments ($n = 4$). * $P = 0.0457$, ** $P = 0.0061$ by unpaired two-tailed t -test. **c**, Representative flow analysis of T cell potential in 5F cells plus shEZH1, with co-expression of full-length mouse *Ezh1* (mEzh1) or mutant mouse *Ezh1* lacking the SET domain (mEzh1 Δ SET, + Δ SET). **d**, Quantification of flow analysis in **c** ($n = 3$ biological replicates). * $P = 0.0146$, ** $P = 0.0011$ by one-way ANOVA. All plots are gated on CD45⁺. Data are pooled across two independent experiments. Data are mean \pm s.e.m. NS, not significant.

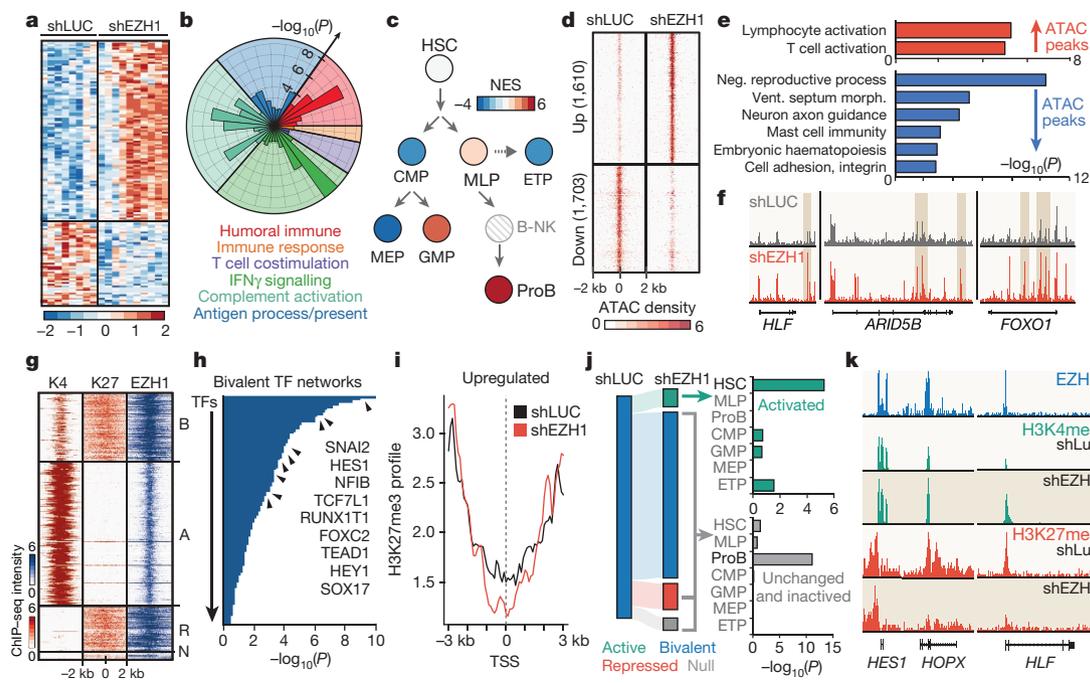


Figure 3 | EZH1 directly binds to and modulates expression and chromatin accessibility of HSC and lymphoid genes. **a**, Heat map of upregulated (104) and downregulated (49) genes (>2-fold; Benjamini–Hochberg corrected *t*-test, $P < 0.1$) from RNA-seq analysis of CD34⁺CD38⁻ HSPCs 5F plus shEZH1 cells ($n = 10$ biological replicates) compared to 5F plus shLUC cells ($n = 8$ biological replicates). **b**, Gene Ontology (GO) analysis of biological processes associated with significantly upregulated genes in **a**, subdivided by GO hierarchical categories with *P* values labelled along the radius. **c**, Enrichment of human HSC and progenitor signatures by gene set enrichment analysis (GSEA) in 5F plus shEZH1 compared with 5F plus shLUC cells, overlaid on the map of human HSPC hierarchy. CMP, common myeloid progenitor; MEP, megakaryocyte–erythroid progenitor; GMP, granulocyte–monocyte progenitor; ETP, early thymic progenitor; NES, normalized enrichment score. **d**, Density map of upregulated and downregulated ATAC peaks by MANorm²⁹ in 5F plus shEZH1 compared to 5F plus shLUC cells ($n = 2$ biological replicates). **e**, GO terms of enriched biological processes of

repressed genes were enriched for developmental and morphogenic processes (Extended Data Fig. 4c, d). *EZH1* knockdown increased the expression of bivalent genes, which were associated with HSC and early lymphoid lineages (Extended Data Fig. 4e, f). These genes included the targets of HSC transcription factors such as *RUNX1T1* and *SOX17*, and NOTCH factors *HES1*, *HEY1* and *FOXC2*¹⁹ (Fig. 3h). *EZH1* directly bound the promoters of HSC and ProB transcription factors including *HLF*, *PRDM16*, *LMO2*, *ETS1*, *MEIS1*, *RUNX1* and *HOX* clusters (Extended Data Fig. 4e). We also observed a global reciprocal relationship between H3K27me3 and gene transcription (Fig. 3i, Extended Data Fig. 4g–k), with poised HSC genes exhibiting loss of H3K27me3 and increased expression upon *EZH1* knockdown (Extended Data Fig. 4h, i). In total, 27 out of 29 of these activated HSC genes are direct targets of *EZH1*, including *HOPX*, *HLF*, *MEIS1* and *HES1* ($P = 7.8 \times 10^{-5}$; Fig. 3j, k).

EZH2 also bound activated HSC genes, consistent with its ability to target the same regions⁸ (Extended Data Fig. 4l); however, recent analysis of SET domain-swapping revealed context-specific sensitivity to an *EZH2*-specific inhibitor, further suggesting that although *EZH1* and *EZH2* can bind a common subset of HSC targets, these enzymes are likely to have distinct functions on chromatin²⁰. Concordant with our observation that *SUZ12* knockdown partially phenocopies *EZH1* loss (Fig. 2a, b), we observed specific enrichment of *EZH1* and *SUZ12* at activated HSC and ProB genes, consistent with non-canonical targets of the *EZH1*–*SUZ12* complex¹⁷ (Extended Data Fig. 4m–q).

ATAC peaks in **d** by GREAT analysis³⁰. **f**, Tracks of representative genes that acquire a significant ATAC peak upon *EZH1* knockdown. **g**, ChIP-seq density map of *EZH1* peaks within bivalent (B), repressed (R), active (A) or null (N) promoter groups ($n = 2$ biological replicates). K4, H3K4me4; K27, H3K27me3. **h**, Waterfall plot of CellNet³¹ predicted regulators of *EZH1*-bound bivalent gene networks. TF, transcription factor. **i**, Sitepro quantitative analysis³² of H3K27me3 levels at all upregulated genes around the transcription start site (TSS) upon *EZH1* knockdown, relative to shLUC ($n = 2$ biological replicates). **j**, Left, Sankey diagram illustrating histone methylation changes of all bivalent genes in shLUC control cells and after *EZH1* knockdown ($n = 2$ biological replicates). Right, genes that lose H3K27me3 (become activated) are specifically enriched in the HSC signature, whereas bivalent genes that are unchanged or inactivated are enriched in the ProB signature by Fisher's exact test. **k**, ChIP-seq tracks of *EZH1*, H3K4me3 and H3K27me3 at representative HSC promoter regions in shLUC and shEZH1 cells. Experiments replicated at least twice.

Similarly, upregulated ATAC peaks in *Ezh1*-deficient AGM were also enriched for *SUZ12* binding, but not *EZH2*, indicating a conserved role for non-canonical PRC2 regulation *in vivo* (Extended Data Fig. 4r). These data suggest that in addition to the canonical function of *EZH1*–PRC2 in mediating H3K27me3 changes at poised HSC loci, *EZH1* also regulates ProB genes through a complementary non-canonical *EZH1*–*SUZ12* complex, highlighting an *EZH1*-specific function that is not phenocopied by *EZH2*.

The emergence of bona fide HSCs, defined by the capacity to repopulate irradiated adult recipients, marks the transition from embryonic to definitive haematopoiesis. We isolated AGM and yolk sac from embryonic day (E)10.5 wild-type, *Ezh1*^{+/-} and *Ezh1*^{-/-} embryos and transplanted adult non-obese diabetic (NOD)/severe combined immunodeficiency (SCID)/*Il2rg*^{-/-} (NSG) recipients (Fig. 4a). We detected peripheral blood reconstitution from wild-type AGM in 3 out of 7 mice ($11.9 \pm 7.9\%$) at 4 weeks, but chimaerism decreased by 16 weeks (2 out of 7, $12.2 \pm 8.1\%$); this corresponds to 1 repopulating unit in approximately 10.4 embryo equivalents (ee), consistent with HSCs being exceedingly rare at E10.5^{10,21}. By contrast, 5 out of 8 mice transplanted with *Ezh1*^{-/-} AGM cells were engrafted at 4 weeks ($39.2 \pm 9.4\%$) and stabilized at 16 weeks ($34.6 \pm 14.6\%$). Notably, *Ezh1*^{+/-} AGM transplant recipients had the highest initial chimaerism ($41.2 \pm 16.3\%$; 4 out of 5), which increased by 16 weeks ($68.9 \pm 17.8\%$), and was predominantly multilineage (3 out of 5) (Fig. 4e, Extended Data Fig. 5a, c). This corresponds to 1 repopulating unit in 3.6 *Ezh1*^{-/-} and

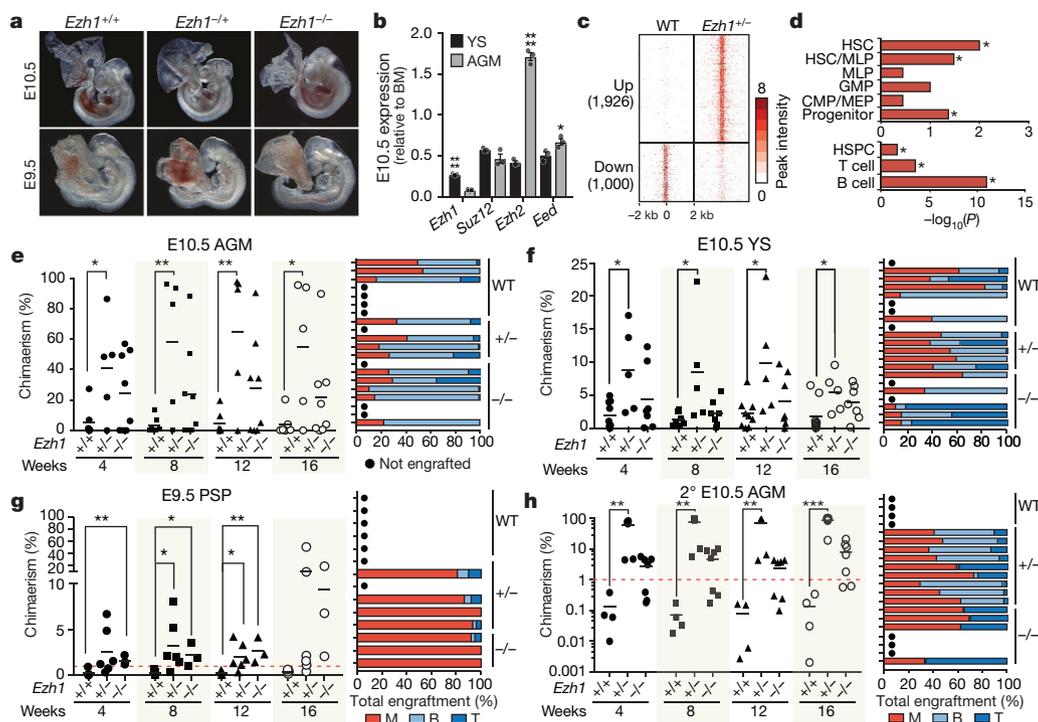


Figure 4 | *Ezh1* deficiency increases lymphoid potential and engraftment of embryonic HSPCs. **a**, Representative images of E9.5 and E10.5 embryos ($n > 50$ embryos). **b**, Quantitative PCR (qPCR) of each PRC2 subunit in E10.5 wild-type yolk sac (YS) and AGM ($n = 3$ biological replicates). $*P = 0.0439$, $****P < 0.0001$ by unpaired two-tailed t -test. Data are mean \pm s.e.m. BM, bone marrow. **c**, ATAC density map of c-Kit⁺VE-cadherin⁺CD45⁺ HSPCs sorted from 30 pooled embryos of E10.5 wild-type (WT) and *Ezh1*^{+/-} AGM. **d**, Significantly upregulated ATAC peaks were compared to HSPC, T and B cell networks and signatures of the human HSPC hierarchy⁴⁵. $*P < 0.05$ by Fisher's exact test. **e**, Left, engraftment of E10.5 AGM (3.5 ee) in sublethally irradiated adult NSG females. Donor chimaerism marked by CD45.2⁺ was measured in peripheral blood every 4 weeks up to 16 weeks post-transplantation.

2.2 *Ezh1*^{+/-} ee, or an approximately fivefold increase in HSC frequency compared to wild type.

At E10.5, the yolk sac is thought to contain few, if any, HSCs²¹. We detected low-level engraftment of wild-type yolk sac cells in 5 out of 9 recipients at 4 weeks ($3.4 \pm 0.7\%$), and in 3 out of 9 mice at 16 weeks ($4.3 \pm 1.6\%$). Most *Ezh1*^{-/-} ($4.5 \pm 0.9\%$, 6 out of 7 engrafted) and all of the *Ezh1*^{+/-} yolk-sac-transplanted mice ($5.4 \pm 1.4\%$, 5 out of 5 engrafted) showed stable long-term engraftment at 16 weeks. The number of repopulating units calculated was similar to that of the AGM (about 1 in 12.3 ee wild-type mice; 1 in 2.6 ee *Ezh1*^{-/-} mice, 1 in <2 ee *Ezh1*^{+/-} mice). All engrafted mice were multilineage (Fig. 4f, Extended Data Fig. 5a, c). Importantly, up to 75% of peritoneal B cells in *Ezh1*^{+/-} AGM-engrafted mice were of the adult-like B-2 phenotype, as opposed to the embryonic B-1 cells (Extended Data Fig. 6a). Moreover, up to 95% of donor-derived CD45.2⁺CD3⁺ T cells expressed adult-type TCR β , as opposed to embryonic TCR $\gamma\delta$, in *Ezh1*^{-/-} and *Ezh1*^{+/-} AGM- and yolk-sac-engrafted mice (Extended Data Fig. 6b). These data provide compelling evidence that *Ezh1* deficiency, and in particular haploinsufficiency, stimulates generation of definitive HSCs and adult-like lymphopoiesis.

The para-aortic splanchnopleura (PSP) at E9.5 lacks HSCs as determined by transplantation studies³. Transplantation of E9.5 wild-type PSP cells (Fig. 4a) failed to engraft adult recipients (0 out of 5)^{21,22}; by contrast, we detected chimaerism in recipients of *Ezh1*^{-/-} (3 out of 3, $1.6 \pm 0.3\%$) and *Ezh1*^{+/-} (4 out of 6 mice, $3.6 \pm 1.3\%$) PSP at 4 weeks post-transplantation (Fig. 4g, Extended Data Fig. 5b). By 16 weeks, chimaerism increased in *Ezh1*^{-/-} (3 out of 3, $9.4 \pm 5.1\%$) and *Ezh1*^{+/-}

Each dot represents a single transplant recipient; lines denote mean values. Right, lineage distribution of engrafted mice showing T cell (T), B cell (B), and myeloid (M) contribution. **f**, Left, engraftment of E10.5 yolk sac (5 ee). Right, lineage distribution of engrafted mice. **g**, Left, engraftment of E9.5 PSP (10 ee). Right, lineage distribution of engrafted mice. **h**, Left, serial transplantation of whole bone marrow from primary recipients of E10.5 AGM cells in **e**. Secondary transplant (2°) was carried out after 24 weeks of primary transplant. Right, lineage distribution of engrafted mice ($n \geq 3$ mice per group). $*P < 0.05$, $**P < 0.01$, $***P < 0.0001$ by unpaired two-tailed t -test. See Supplementary Information for exact P values per time point. Data are pooled across four (**e–g**) or three (**h**) independent experiments; experiment in **c** was performed once.

(5 out of 6, $13.1 \pm 9.5\%$) recipients, and grafts were fully multilineage (Extended Data Fig. 5c). Thus, *Ezh1* deficiency stimulates precocious generation of bona fide HSCs during embryogenesis.

To assess the self-renewal capacity of *Ezh1*-deficient HSCs, we performed secondary transplantation. No mice showed engraftment with E10.5 wild-type AGM (0 out of 4) or yolk sac (0 out of 7). By contrast, 4 out of 7 *Ezh1*^{-/-} ($4.4 \pm 0.5\%$) and 9 out of 9 *Ezh1*^{+/-} ($57.8 \pm 10.2\%$) AGM-derived secondary recipients were engrafted (Fig. 4h, Extended Data Fig. 5d). Of note, although no *Ezh1*^{-/-} yolk sac recipients (0 out of 10) were engrafted, we observed secondary chimaerism from *Ezh1*^{+/-} yolk sac cells (5 out of 7, $1.5 \pm 0.3\%$), which increased by 16 weeks (6 out of 7, $5.1 \pm 1.9\%$) (Extended Data Fig. 5d, e). All engrafted secondary recipients were multilineage, with no evidence of leukaemic transformation (Fig. 4h, Extended Data Fig. 5c, e). Taken together, these data indicate that genetic *Ezh1* deficiency elicits precocious emergence of bona fide HSCs *in vivo*.

It has long been a curiosity that haematopoietic ontogeny progresses in reverse order, with haematopoietic progenitors appearing first in embryonic development independently of HSCs^{2,3}. We propose that *EZH1* represses definitive loci in primitive blood progenitors differentiated from human pluripotent stem cells and in mouse embryos, which precludes precocious HSC emergence during gestation. *EZH1* deficiency promotes multipotency in lineage-restricted blood progenitors and enables precocious emergence of HSCs. Although PRC2 is a well-characterized HSC regulator, our data contribute compelling evidence for the distinct molecular functions of *EZH1* and *EZH2*, and suggest a putative role for non-canonical PRC2, involving *EZH1* and

SUZ12. Homozygous loss of *Suz12* in mice impairs HSC function and lymphopoiesis, but heterozygosity for *Suz12* or *Eed* enhances HSC self-renewal^{23,24}. Consistent with this, our data reinforce the concept that HSCs are exquisitely sensitive to PRC2 dosage, with partial reduction or increase affecting function^{23–26}. Interestingly, *Runx1* haploinsufficiency also promotes premature HSC generation²⁷. Our data unify these observations; EZH1 marks many transcription factor-binding sites, whereas *Ezh1* deficiency enhances accessibility to targets of key HSC transcription factors, including *Runx1*, to promote HSC emergence (Extended Data Fig. 3k, l). We identify *Ezh1* as a molecular regulator of lineage-restricted potential of the first blood progenitors in the mammalian embryo, which accounts in part for why early embryonic progenitors lack multipotency. Beyond developmental implications, our findings suggest that resolution of EZH1-marked domains may be essential for physiological specification of HSCs from pluripotent stem cells, as a complementary approach to the synthetic reactivation of stem-cell programs by HSC transcription factors^{7,28}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 October 2016; accepted 5 December 2017.

Published online 17 January 2018.

- Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631–644 (2008).
- Medvinsky, A., Rybtsov, S. & Taoudi, S. Embryonic origin of the adult hematopoietic system: advances and questions. *Development* **138**, 1017–1031 (2011).
- Dzierzak, E. & Speck, N. A. Of lineage and legacy: the development of mammalian hematopoietic stem cells. *Nat. Immunol.* **9**, 129–136 (2008).
- McGrath, K. E. *et al.* Distinct sources of hematopoietic progenitors emerge before HSCs and provide functional blood cells in the mammalian embryo. *Cell Rep.* **11**, 1892–1904 (2015).
- Böiers, C. *et al.* Lymphomyeloid contribution of an immune-restricted progenitor emerging prior to definitive hematopoietic stem cells. *Cell Stem Cell* **13**, 535–548 (2013).
- Cedar, H. & Bergman, Y. Epigenetics of haematopoietic cell development. *Nat. Rev. Immunol.* **11**, 478–488 (2011).
- Doulatov, S. *et al.* Induction of multipotential hematopoietic progenitors from human pluripotent stem cells via respecification of lineage-restricted precursors. *Cell Stem Cell* **13**, 459–470 (2013).
- Shen, X. *et al.* EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol. Cell* **32**, 491–502 (2008).
- Laurenti, E. *et al.* The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nat. Immunol.* **14**, 756–763 (2013).
- North, T. E. *et al.* Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. *Immunity* **16**, 661–672 (2002).
- Rybtsov, S. *et al.* Hierarchical organization and early hematopoietic specification of the developing HSC lineage in the AGM region. *J. Exp. Med.* **208**, 1305–1315 (2011).
- Boisset, J. C. *et al.* *In vivo* imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature* **464**, 116–120 (2010).
- Riddell, J. *et al.* Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors. *Cell* **157**, 549–564 (2014).
- Papaemmanuil, E. *et al.* Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat. Genet.* **41**, 1006–1010 (2009).
- Tothova, Z. *et al.* FoxOs are critical mediators of hematopoietic stem cell resistance to physiologic oxidative stress. *Cell* **128**, 325–339 (2007).
- Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
- Xu, J. *et al.* Developmental control of Polycomb subunit composition by GATA factors mediates a switch to non-canonical functions. *Mol. Cell* **57**, 304–316 (2015).
- Margueron, R. *et al.* Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol. Cell* **32**, 503–518 (2008).
- Jang, I. H. *et al.* Notch1 acts via Foxc2 to promote definitive hematopoiesis via effects on hemogenic endothelium. *Blood* **125**, 1418–1426 (2015).

- Hohmann, A. F. *et al.* Sensitivity and engineered resistance of myeloid leukemia cells to BRD9 inhibition. *Nat. Chem. Biol.* **12**, 672–679 (2016).
- Müller, A. M., Medvinsky, A., Strouboulis, J., Grosveld, F. & Dzierzak, E. Development of hematopoietic stem cell activity in the mouse embryo. *Immunity* **1**, 291–301 (1994).
- Yoder, M. C. *et al.* Characterization of definitive lymphohematopoietic stem cells in the day 9 murine yolk sac. *Immunity* **7**, 335–344 (1997).
- Lee, S. C. *et al.* Polycomb repressive complex 2 component *Suz12* is required for hematopoietic stem cell function and lymphopoiesis. *Blood* **126**, 167–175 (2015).
- Xie, H. *et al.* Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner. *Cell Stem Cell* **14**, 68–80 (2014).
- Majewski, I. J. *et al.* Polycomb repressive complex 2 (PRC2) restricts hematopoietic stem cell activity. *PLoS Biol.* **6**, e93 (2008).
- Kamminga, L. M. *et al.* The Polycomb group gene *Ezh2* prevents hematopoietic stem cell exhaustion. *Blood* **107**, 2170–2179 (2006).
- Cai, Z. *et al.* Haploinsufficiency of AML1 affects the temporal and spatial generation of hematopoietic stem cells in the mouse embryo. *Immunity* **13**, 423–431 (2000).
- Sugimura, R. *et al.* Haematopoietic stem and progenitor cells from human pluripotent stem cells. *Nature* **545**, 432–438 (2017).
- Shao, Z., Zhang, Y., Yuan, G. C., Orkin, S. H. & Waxman, D. J. MAAnorm: a robust model for quantitative comparison of ChIP-seq data sets. *Genome Biol.* **13**, R16 (2012).
- McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
- Cahan, P. *et al.* CellNet: network biology applied to stem cell engineering. *Cell* **158**, 903–915 (2014).
- Shin, H., Liu, T., Manrai, A. K. & Liu, X. S. CEAS: cis-regulatory element annotation system. *Bioinformatics* **25**, 2605–2606 (2009).
- Doulatov, S. *et al.* Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat. Immunol.* **11**, 585–593 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank T. Jenuwein for sharing the *Ezh1* mutant mice, which were generated at the Research Institute of Molecular Pathology (IMP, Vienna) in 2000 by D. O'Carroll (laboratory of T. Jenuwein) with the help of M. Sibilia (laboratory of E. Wagner). We also thank T. Schlaeger and the HESC Core Facility at Boston Children's Hospital for providing pluripotent stem-cell lines, R. Mathieu from BCH Flow Cytometry Core, and M. J. Chen for technical advice. This work was supported by grants from the NIH NIDDK (R24-DK092760, R24-DK49216) and NHLBI Progenitor Cell Biology Consortium (U01-HL100001); NHLBI R01HL04880 and NIH R24OD017870-01. L.T.V. is supported by the NSF Graduate Research Fellowship. M.A.K. is supported by T32 NIH Training Grant from BWH Hematology. M.C. is supported by a fellowship from the Leukemia and Lymphoma Society. S.D. is supported by K99 NIH NHLBI award (1K99HL123484). S.H.O. is an Investigator of the Howard Hughes Medical Institute. J.X. is supported by NIH grants (K01DK093543 and R01DK111430) and a Cancer Prevention and Research Institute of Texas (CPRIT) New Investigator award (RR140025). G.Q.D. was supported by the Howard Hughes Medical Institute, and is an associate member of the Broad Institute and an investigator of the Manton Center for Orphan Disease Research.

Author Contributions L.T.V., S.D. and G.Q.D. conceived the project. L.T.V. designed all experiments, performed all pluripotent stem-cell and mouse transplantation studies and interpreted data. M.A.K. analysed RNA-seq, ChIP-seq and ATAC-seq data, performed all network analyses and interpreted data. X.L. performed ChIP-seq and ATAC-seq experiments. Y.Z. and Z.S. analysed ChIP-seq and ATAC-seq data. J.B. performed and analysed qPCR and western blot validations, assisted with tissue culture, animal dissections and mouse transplantation studies. P.M.S. assisted with timed matings, animal dissections and mouse transplantation studies. D.K.J. performed western blot validations, cloned the *Ezh2*-mCherry overexpression construct, assisted with ChIP-seq optimization and interpreted data. M.C. assisted with ChIP-seq optimization. A.H. assisted with RNA-seq analysis. T.E.N., S.H.O., S.D., J.X. and G.Q.D. supervised research, interpreted data and participated in project planning. L.T.V., T.E.N., S.D. and G.Q.D. wrote the manuscript with input from all co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to G.Q.D. (George.Daley@childrens.harvard.edu).

Reviewer Information *Nature* thanks B. Gottgens, H. Mikkola and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

A step-by-step protocol can be found at the Protocol Exchange³³.

Human iPS cell culture. All experiments were performed using MSC-iPS³⁴, CD34-iPS and CD45-iPS cells, obtained from the Boston Children's Hospital Human Embryonic Stem Cell Core (hESC) and verified by immunohistochemistry for pluripotency markers, teratoma formation and karyotyping. All cells were routinely tested for mycoplasma contamination. Human iPS cells were maintained on mouse embryonic fibroblast (GlobalStem) feeders in DMEM/F12 plus 20% KnockOut-Serum Replacement (Invitrogen), 1 mM L-glutamine, 1 mM non-essential amino acids (NEAA), 0.1 mM β -mercaptoethanol and 10 ng ml⁻¹ bFGF. Medium was changed daily, and cells were passaged 1:4 onto fresh feeders every 7 days using standard clump passaging with collagenase IV.

Embryoid body differentiation. Differentiation of embryoid bodies was performed as previously described³⁵. In brief, human pluripotent stem cell colonies were scraped into non-adherent rotating 10 cm plates at the ratio of 2:1. Embryoid body medium was KO-DMEM plus 20% FBS (Stem Cell Technologies), 1 mM L-glutamine, 1 mM NEAA, 1% penicillin-streptomycin, 0.1 mM β -mercaptoethanol, 200 μ g ml⁻¹ human transferrin and 50 μ g ml⁻¹ ascorbic acid. After 24 h, medium was changed by allowing embryoid bodies to settle by gravity, and replaced with embryoid body medium supplemented with growth factors: 50 ng ml⁻¹ BMP4 (R&D Systems), 200 ng ml⁻¹ SCF, 200 ng ml⁻¹ FLT3, 50 ng ml⁻¹ G-CSF, 20 ng ml⁻¹ IL-6 and 10 ng ml⁻¹ IL-3 (all Peprotech). Medium was changed on days 5 and 10. Embryoid bodies were dissociated on day 14 by digesting with collagenase B (Roche) for 2 h, followed by treatment with enzyme-free dissociation buffer (Gibco), and filtered through an 80- μ m filter. Dissociated embryoid bodies were frozen in 10% DMSO, 40% FBS freezing solution.

Progenitor sorting. Dissociated embryoid body cells were thawed following the Lonza Poetics protocol and resuspended at 1×10^6 per 100 μ l staining buffer (PBS plus 2% FBS). CD34⁺ cells were sorted from bulk embryoid body culture using human CD34 microbeads (Miltenyi Biotec) and run through a magnetic column separator (MACS) as per the manufacturer's instructions.

Lentiviral and shRNA library plasmids. The 5F lentiviral plasmids HOXA9, ERG, RORA, SOX4 and MYB were cloned into pInducer-21 doxycycline-inducible lentiviral vector. The shRNA library targeting 20 epigenetic modifiers³⁶ was obtained from the Broad Institute RNAi Consortium in pLKO.1 or pLKO.5 lentiviral vectors. Lentiviral particles were produced by transfecting 293T-17 cells (ATCC) with the lentiviral plasmids and third-generation packaging plasmids. Viruses were harvested 24 h after transfection and concentrated by ultracentrifugation at 64,965g for 3 h using the Beckman Coulter SW 32 Ti rotor. All viruses were titred by serial dilution on 293T cells.

5F gene transfer and 5F culture. MACS-separated CD34⁺ embryoid body progenitors were seeded on retromectin-coated (10 μ g cm⁻²) 96-well plates at a density of 2×10^4 – 5×10^4 cells per well. The infection medium was SFEM (StemCell Technologies) with 50 ng ml⁻¹ SCF, 50 ng ml⁻¹ FLT3, 50 ng ml⁻¹ TPO (all R&D Systems), 50 ng ml⁻¹ IL-6 and 10 ng ml⁻¹ IL-3 (both from Peprotech). Lentiviral infections were carried out in a total volume of 150 μ l. The multiplicity of infection (MOI) for each factor was as follows: ERG MOI = 5, HOXA9 MOI = 5, RORA MOI = 3, SOX4 MOI = 3, MYB MOI = 3, and MOI = 2 for shRNA. Virus was concentrated onto cells by centrifuging the plate at 924g for 30 min at room temperature. Infections were carried out for 24 h. After gene transfer, 5F cells were cultured in SFEM with 50 ng ml⁻¹ SCF, 50 ng ml⁻¹ FLT3, 50 ng ml⁻¹ TPO, 50 ng ml⁻¹ (all R&D Systems) IL-6, and 10 ng ml⁻¹ IL-3 (Peprotech). Doxycycline (Dox) was added at 2 μ g ml⁻¹ (Sigma). Puromycin was added at 0.3 μ g ml⁻¹ (ThermoFisher Scientific). Cultures were maintained at a density of $<1 \times 10^6$ cells ml⁻¹, and the medium was changed every 3–4 days.

T cell differentiation. After 14 days of respecification, 1×10^5 5F cells were plated in OP9-DL1 stromal co-culture³⁷. Cells were cultured in α -MEM (Gibco), 1% penicillin-streptomycin, 20% FBS (Gemini), and 1 mM L-glutamine with 30 ng ml⁻¹ SCF, 5 ng ml⁻¹ FLT3, 5 ng ml⁻¹ IL-7 (all R&D Systems) for 20 days with 2 μ g ml⁻¹ Dox followed by Dox removal. Cells were collected by mechanical dissociation and filtered through a 40- μ m filter and passaged onto fresh stroma every 5–7 days. T cell development was assessed after 35 days using CD45, CD7, CD3, CD4 and CD8.

B cell differentiation. After 14 days of respecification, 5×10^4 5F cells were plated into a single well of MS-5 stroma in a 6-well NUNC plate. Cells were cultured in Myelocult H5100 (Stem Cell Technologies) supplemented with 50 ng ml⁻¹ SCF, 10 ng ml⁻¹ FLT3, 25 ng ml⁻¹ IL7, 25 ng ml⁻¹ TPO (all R&D Systems) and 1% penicillin-streptomycin for 10 days with 2 μ g ml⁻¹ Dox followed by Dox removal.

Colony assays. After 14 days of respecification, 5×10^4 cells were plated into 3 ml of complete methylcellulose H3434 (StemCell Technologies) supplemented with 10 ng ml⁻¹ IL-6 (Peprotech), 10 ng ml⁻¹ FLT3 (R&D) and 50 ng ml⁻¹ TPO (R&D)

without 2 μ g ml⁻¹ Dox. The mixture was distributed into two 60-mm dishes and maintained in a humidified chamber for 14 days.

Mouse transplantation. NOD/SCID/IL2rg^{-/-} (NSG) (Jackson Laboratory) mice were bred and housed at the Boston Children's Hospital animal care facility. Animal experiments were performed in accordance with institutional guidelines approved by Boston Children's Hospital Animal Care Committee. At least three animals were used per cohort, based on previous transplantation studies. Mice were assigned randomly to groups and blinding was not used. In brief, 8–12-week-old mice were irradiated (2.75 Gy) 24 h before transplant. To ensure consistency between experiments, only female mice were used. Sublethally irradiated adult NSG females were transplanted intravenously with 3.5 ee of whole E10.5 AGM, 5 ee of whole E10.5 yolk sac or 10 ee of whole E9.5 PSP. Mice were bled retroorbitally every 4 weeks to monitor donor chimaerism up to 16 weeks post-transplantation. Twenty-four weeks after primary transplantation, primary recipients from each group were euthanized and 4×10^6 whole bone marrow cells were transplanted into 1–3 secondary recipients. Cells were transplanted in a 200 μ l volume using a 28.5-gauge insulin needle. Sulfatrim was administered in drinking water to prevent infections after irradiation. Data points were combined from all independent experiments and outliers were not excluded.

Flow cytometry. The following antibodies were used for human cells: CD45 allophycocyanin (APC)-conjugated Cy7 (557833, BD Biosciences), CD4 phycoerythrin (PE)-conjugated Cy5 (IM2636U, Beckman Coulter Immunotech), CD8-BV421 (RPA-T8, BD Horizon), CD5-BV510 (UCHT2, BD Biosciences), TCR $\gamma\delta$ -APC (555718, BD Biosciences), TCR $\alpha\beta$ -BV510 (T10B9.1A-31, BD Biosciences), CD3-PE-Cy7 (UCHT1, BD Pharmigen), CD7-PE (555361, BD Pharmigen), CD1a-APC (559775, BD Pharmigen) for T cell staining. For B cell staining: CD45-PE-Cy5 (IM2652U, Beckman Coulter Immunotech), CD19-PE (4G7, BD Biosciences), CD56-V450 (B159, BD Biosciences), CD11b-APC-Cy7 (557754, BD Biosciences), For HSC/progenitor sorting: CD34-PE-Cy7 (8G12, BD Biosciences), CD45-APC-Cy7 (557833, BD Biosciences), CD38-PE-Cy5 (IM2651U, Beckman Coulter) and DAPI. For myeloid and erythroid staining: CD11b-APC-Cy7 (557754, BD Biosciences), GLYA-PE-Cy7 (A71564, Beckman Coulter), CD71-PE (555537, BD Biosciences), CD45-PE-Cy5 (IM2652U, Beckman Coulter Immunotech). All staining was performed with $<1 \times 10^6$ cells per 100 μ l staining buffer (PBS plus 2% FBS), with a 1:100 dilution of each antibody, for 30 min at room temperature in the dark. Compensation was performed by automated compensation with anti-mouse Ig κ and negative beads (BD Biosciences). All acquisitions were performed on a BD Fortessa or BD Aria cytometer.

The following antibodies were used for mouse cells: CD45.2-PE-Cy7 (104, eBioscience), CD45.1-FITC (A20, eBioscience), B220-PB (RA3-6B2, BD Biosciences), Ter119-PE-Cy5 (Ter 119, eBioscience), GR1 (RB6-8C5, BD Bioscience), CD3-APC (145-2C11, eBioscience), CD19-APC-Cy7 (1D3, BD Bioscience), MAC1-AF700 (M1/70, BD Bioscience) for engraftment analyses. For B cell staining: CD45.2-APC-Cy7 (104, BioLegend), CD23-PE-Cy7 (B3B4, eBioscience), Ter119-PE-Cy5 (Ter 119, eBioscience), MAC1-A700 (M1/70, BD Bioscience), CD5-BV510 (53-7.3, BD Biosciences), IgM-eFluor660 (II/41, eBioscience). For T cell staining: CD45.2-PE-Cy7 (104, eBiosciences), TCR β -PE-Cy5 (H57-597, BD Biosciences), CD8-APC-EF780 (53-6.7, eBioscience), CD4-APC (GK1.5, eBioscience), CD3-AF700 (17A2, BioLegend), TCR $\gamma\delta$ -FITC (GL3, BD Biosciences). For HSPC sorting: CD16/32 (93, BioLegend), Ter119-biotin (Ter119, eBioscience), Gr-1-biotin (RB6-8C5, eBioscience), CD3-biotin (17A2, eBioscience), CD5-biotin (53-7.3, eBioscience), CD8-biotin (53-6.7, eBioscience), CD19-biotin (eBio1D3, eBioscience), streptavidin-eFluor450 (eBioscience), CD45-PerCP-Cy5.5 (30-F11, eBioscience), CD144-eFluor660 (eBioBV13, eBioscience), CD117-APC-eFluor 780 (2B8, eBioscience), CD41-PE-Cy7 (eBioMWR30, eBioscience). All staining was performed with $<1 \times 10^6$ cells per 100 μ l staining buffer (PBS plus 2% FBS), with a 1:100 dilution of each antibody, for 30 min on ice in the dark. Compensation was performed by automated compensation with anti-rat and anti-hamster Ig κ and negative beads (BD Biosciences). All acquisitions were performed on a BD Fortessa or BD Aria cytometer.

RNA-seq. Human cells were stained and sorted using CD34-PE-Cy7 (8G12, BD Biosciences), CD38-PE-Cy5 (IM2651U, Beckman Coulter) and DAPI (Beckman Coulter). RNA-seq libraries were prepared using the NEB Ultra (PolyA) kit as per the manufacturer's protocol with 50 ng input RNA. Mouse cells were stained and sorted using the 'HSPC stain' (see 'Flow cytometry'). RNA-seq libraries were prepared using the Clontech SMARTer Universal Low Input kit as per the manufacturer's protocol with 10 ng input RNA. Libraries were sequenced using the 200 cycle paired-end kit on the Illumina HiSeq2500 system. RNA-seq reads were analysed with Tuxedo Tools following a standard protocol³⁸. Reads were mapped with TopHat version 2.1.0 and Bowtie2 version 2.2.4 with default parameters against build hg19 of the human genome, and build hg19 of the RefSeq human genome

annotation. Samples were quantified with the Cufflinks package version 2.2.1. Differential expression was performed using Cuffdiff with default parameters.

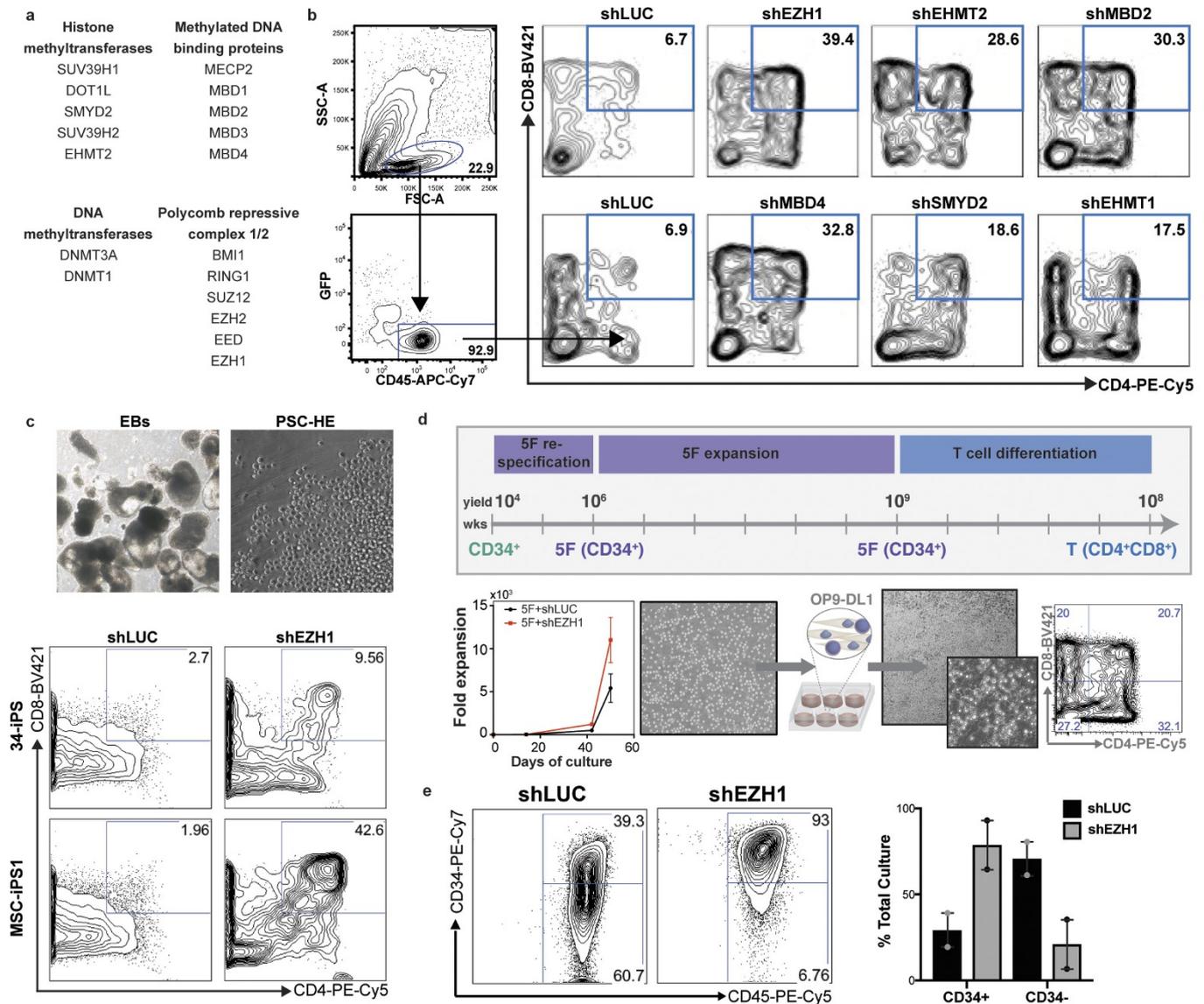
ATAC-seq. ATAC-seq was performed as previously described³⁹. 5×10^3 – 50×10^3 cells were used for each tagmentation using Tn5 transposases. The resulting DNA was isolated, quantified and sequenced on an Illumina NextSeq500 system. The raw reads were aligned to the human genome assembly hg19 using Bowtie⁴⁰ with the default parameters, and only tags that uniquely mapped to the genome were used for further analysis. ATAC peaks were identified using MACS⁴¹.

ChIP-seq. ChIP experiments were performed as previously described⁴² using the antibodies for H3K4me3 (04-745, Millipore) and H3K27me3 (07-449, Millipore) in 5F cells. For bioChIP analysis of EZH1 or EZH2 occupancy, Flag-biotin-tagged EZH1 or EZH2 was stably expressed in 5F cells. The chromatin was isolated and immunoprecipitated with streptavidin Dynabeads (Life Technologies) as previously described⁴³. ChIP-seq libraries were generated using NEBNext ChIP-seq Library Prep Master Mix following the manufacturer's protocol (New England Biolabs), and sequenced on an Illumina NextSeq500 system. ChIP-seq raw reads were aligned to the human genome assembly hg19 using Bowtie⁴⁰ with the default parameters; only tags that uniquely mapped to the genome were used for further analysis. ChIP-seq peaks were identified using MACS⁴¹.

Bioinformatics and statistical analysis. All statistical calculations were performed using GraphPad Prism. Tests between two groups used a two-tailed unpaired Student's *t*-test. Data are presented as mean \pm s.e.m. Where indicated, ANOVA was used, with $P < 0.05$ considered significant. GSEA and GO were run according to default parameters in their native implementations. Statistical enrichment of gene lists was performed using Fisher's exact test. No statistical methods were used to predetermine sample size.

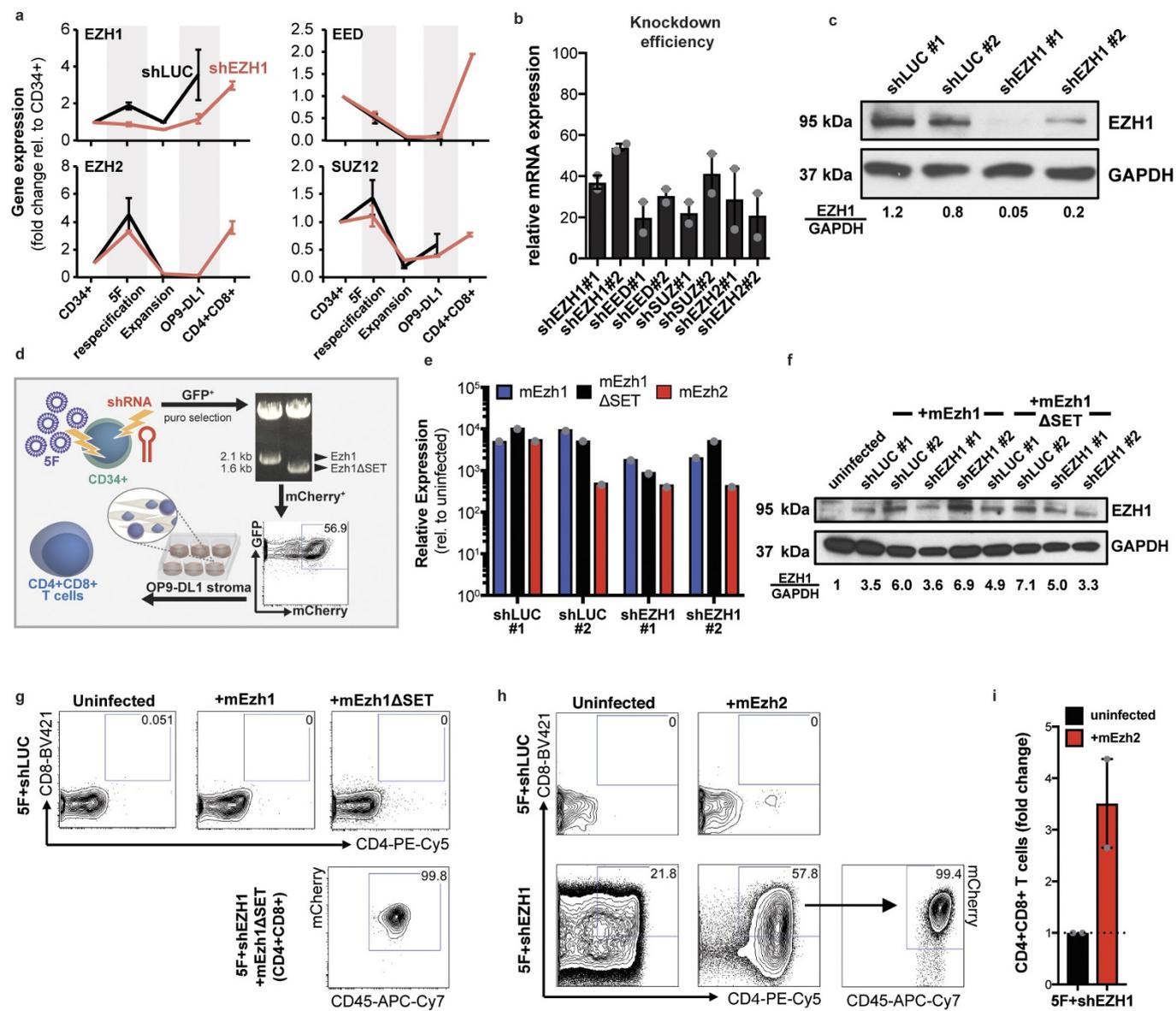
Data availability. All RNA-seq, ATAC-seq and ChIP-seq data have been deposited to the Gene Expression Omnibus (GEO) database under the accession number GSE89418.

33. Vo, L. T., Doulatov, S. & Daley, G. Q. Haematopoietic progenitor and lymphoid differentiation from human pluripotent stem cells. *Protoc. Exch.* <http://doi.org/10.1038/protex.2017.160> (2018).
34. Park, I. H. *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**, 141–146 (2008).
35. Chadwick, K. *et al.* Cytokines and BMP-4 promote hematopoietic differentiation of human embryonic stem cells. *Blood* **102**, 906–915 (2003).
36. Onder, T. T. *et al.* Chromatin-modifying enzymes as modulators of reprogramming. *Nature* **483**, 598–602 (2012).
37. Holmes, R. & Zúñiga-Pflücker, J. C. The OP9-DL1 system: generation of T-lymphocytes from embryonic or hematopoietic stem cells *in vitro*. *Cold Spring Harb. Protoc.* **2009**, <http://doi.org/10.1101/pdb.prot5156> (2009).
38. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46–53 (2013).
39. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
40. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
41. Zhang, Y. *et al.* Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
42. Huang, J. *et al.* Dynamic control of enhancer repertoires drives lineage and stage-specific transcription during hematopoiesis. *Dev. Cell* **36**, 9–23 (2016).
43. Kim, J., Cantor, A. B., Orkin, S. H. & Wang, J. Use of *in vivo* biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. *Nat. Protoc.* **4**, 506–517 (2009).
44. Ditadi, A. *et al.* Human definitive haemogenic endothelium and arterial vascular endothelium represent distinct lineages. *Nat. Cell Biol.* **17**, 580–591 (2015).



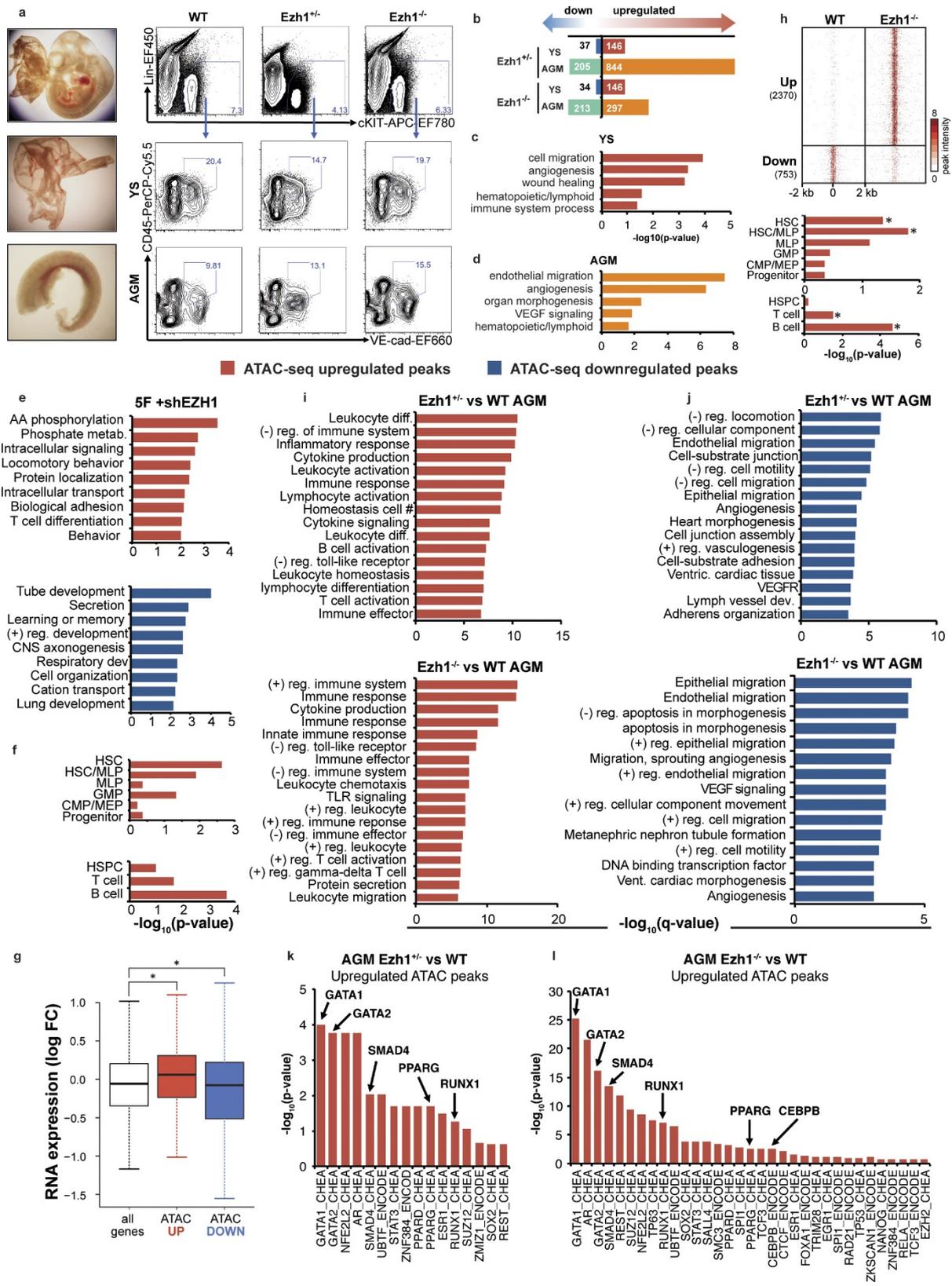
Extended Data Figure 1 | EZH1 knockdown activates lymphoid potential from pluripotent stem cells. **a**, List of all candidate epigenetic modifiers in loss-of-function shRNA screen. **b**, Representative flow plots of CD4⁺CD8⁺ T cell potential across top six candidates from four independent hairpins in two independent experiments ($n = 8$). See Fig. 1. **c**, CD34⁺ cells were isolated after 9 days of embryoid body (EB) differentiation (top left), transduced with shLUC or shEZH1 and cultured under conditions that promote endothelial-to-haematopoietic transition⁴⁴. After 6 days, rounded haematopoietic cells (top right) were collected and co-cultured on OP9-DL1 stroma. Bottom, flow cytometric analysis of T cell potential in shLUC and shEZH1 cells without 5F is shown for two independent iPS lines (34-iPS and MSC-iPS1) in one experiment ($n = 2$ biological replicates). PSC-HE, pluripotent stem-cell-derived haematogenic

endothelium. **d**, Expansion and differentiation potential of 5F plus shEZH1 cells after long-term *in vitro* culture. 5F plus shEZH1 cells were maintained in cultures containing doxycycline for 14 days re-specification (approximately 100-fold expansion), plus an additional 6 weeks (approximately 1,000-fold expansion) and then plated into OP9-DL1 stromal cells for T cell differentiation. Representative flow cytometric analyses of T cell potential of 5F plus shLUC and 5F plus shEZH1 cells after 13 weeks of expansion and differentiation ($n = 2$ biological replicates). **e**, Flow cytometric analysis (left) and quantification (right) of the proportion of CD34⁺ and CD34⁻ haematopoietic progenitors in doxycycline-containing suspension culture at day 25 ($n = 2$ biological replicates).



Extended Data Figure 2 | Ezh1, but not Ezh2, suppresses T cell potential and requires its catalytic domain. **a**, qPCR of PRC2 expression (human genes *EZH1*, *EED*, *EZH2* and *SUZ12*), during the course of differentiation from human pluripotent stem cell-derived CD34⁺ cells, respecification (5F), expansion, OP9-DL1 co-culture and CD4⁺CD8⁺ T cells ($n = 2$ biological replicates in one experiment). **b**, qPCR of mRNA knockdown efficiency of individual shRNAs for PRC2 genes ($n = 2$ biological replicates). See also Fig. 2a, b, c, Western blot for EZH1 and GAPDH protein levels. **d**, Scheme for rescue experiments. GFP⁺ 5F cells were transduced with shRNAs and selected with puromycin. 5F plus shRNA cells were then transduced with full-length mouse *Ezh1* open-reading frame (mEzh1) or mutant mouse *Ezh1* with the catalytic SET domain deleted (mEzh1ΔSET), marked by mCherry fluorescence. Triple-transduced (GFP⁺, puromycin-resistant, mCherry⁺) cells were sorted and seeded onto OP9-DL1. T cells were analysed by flow cytometry

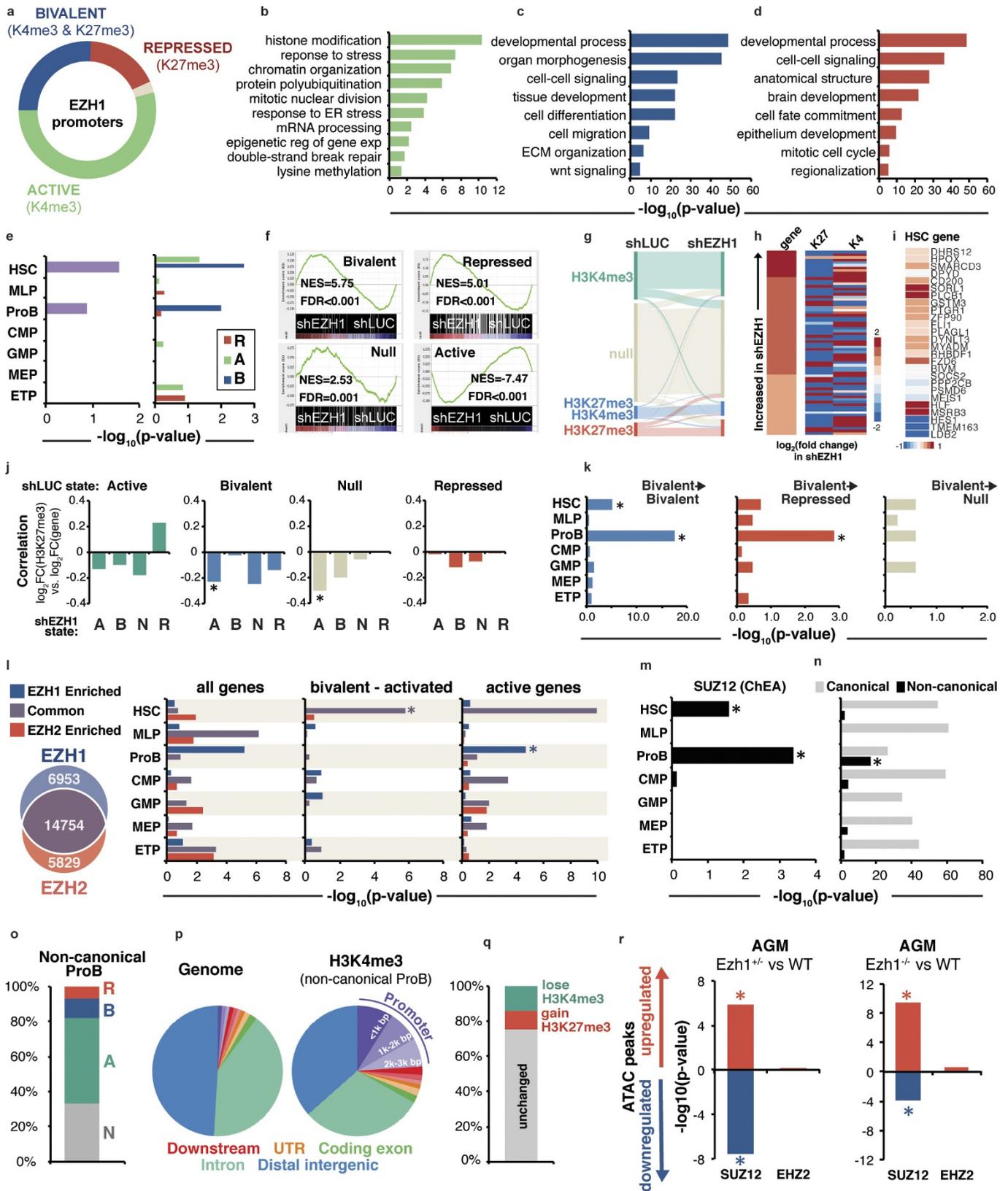
after 5 weeks of differentiation. See also Fig. 2c. **e**, Expression of full-length mouse *Ezh1*, catalytic-deleted mEzh1ΔSET, or full-length mouse *Ezh2* in shLUC and shEZH1 cells by qPCR. **f**, Western blot validation of expression of mouse *Ezh1* or mutant mEzh1ΔSET in shLUC and shEZH1 cells. **g**, Top, representative flow cytometry plots of T cell potential for 5F plus shLUC cells for rescue experiments in **d** ($n = 3$ biological replicates). Bottom, CD4⁺CD8⁺ T cells were verified for mCherry fluorescence. See also Fig. 2c. All plots are gated on CD45⁺. **h**, 5F plus shRNA cells were transduced with full-length mouse *Ezh2* open-reading frame (mEzh2) marked by mCherry fluorescence. Triple-transduced (GFP⁺, puromycin-resistant, mCherry⁺) cells were sorted and seeded onto OP9-DL1 stromal cells. T cells were analysed by flow cytometry after 5 weeks of differentiation. Representative flow plots for two biological replicates in one experiment. **i**, Quantification of data in **h**. Data are mean ± s.e.m.



Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | *Ezh1* regulates haematopoietic and lymphoid programs *in vitro* and *in vivo*. **a**, Representative images of E10.5 embryo (top), yolk sac (middle) and AGM (bottom) from $n > 30$ embryos. $\text{Lin}^- \text{c-Kit}^+ \text{VE-cadherin}^+ \text{CD45}^+ \text{CD41}^+$ cells from E10.5 yolk sac and AGM were FACS-sorted followed by RNA-seq analysis. See also Fig. 4a, c, d. **b**, Genes upregulated and downregulated by more than twofold in *Ezh1*^{+/-} or *Ezh1*^{-/-} yolk sac and AGM compared to those from wild-type mice. **c, d**, GO term annotations of upregulated genes in *Ezh1*^{+/-} and *Ezh1*^{-/-} yolk sac and AGM compared to those from wild-type mice. **e**, GO analysis of enriched pathways of the 1,033 nearest neighbour genes associated with upregulated ATAC peaks (top) and the nearest 1,012 neighbour genes associated with downregulated ATAC peaks (bottom). See also Fig. 3d, e. **f**, Comparison of upregulated ATAC peaks in 5F plus shEZH1 cells with HSPC hierarchy signatures⁴⁵ (top) and HSPC B and T cell networks (bottom). See also Fig. 3d, e. **g**, Box plot of expression of genes associated

with upregulated and downregulated ATAC peaks. * $P < 0.05$ by one-way ANOVA. **h**, ATAC density map of $\text{c-Kit}^+ \text{VE-cadherin}^+ \text{CD45}^+$ HSPCs sorted from approximately 30 embryos of E10.5 wild-type and *Ezh1*^{-/-} AGM (top) from one experiment. Significantly upregulated ATAC peaks were compared to HSPC, T, B cell networks and signatures of the human HSPC hierarchy (bottom). See also Fig. 4c, d. **i**, GO terms of enriched pathways of regions associated with significantly upregulated ATAC peaks annotated by GREAT analysis in *Ezh1*^{+/-} AGM (top) and *Ezh1*^{-/-} AGM (bottom) compared to wild type. See also Fig. 4c, d. **j**, GO terms of enriched pathways of regions associated with significantly downregulated ATAC peaks annotated by GREAT analysis³⁰ in *Ezh1*^{+/-} AGM (top) and *Ezh1*^{-/-} AGM (bottom) compared to wild type. See also Fig. 4c. **k**, Transcription factor binding to genes with upregulated ATAC peaks in *Ezh1*^{+/-} (left) and *Ezh1*^{-/-} (right) AGM from **i** compared to wild-type AGM.

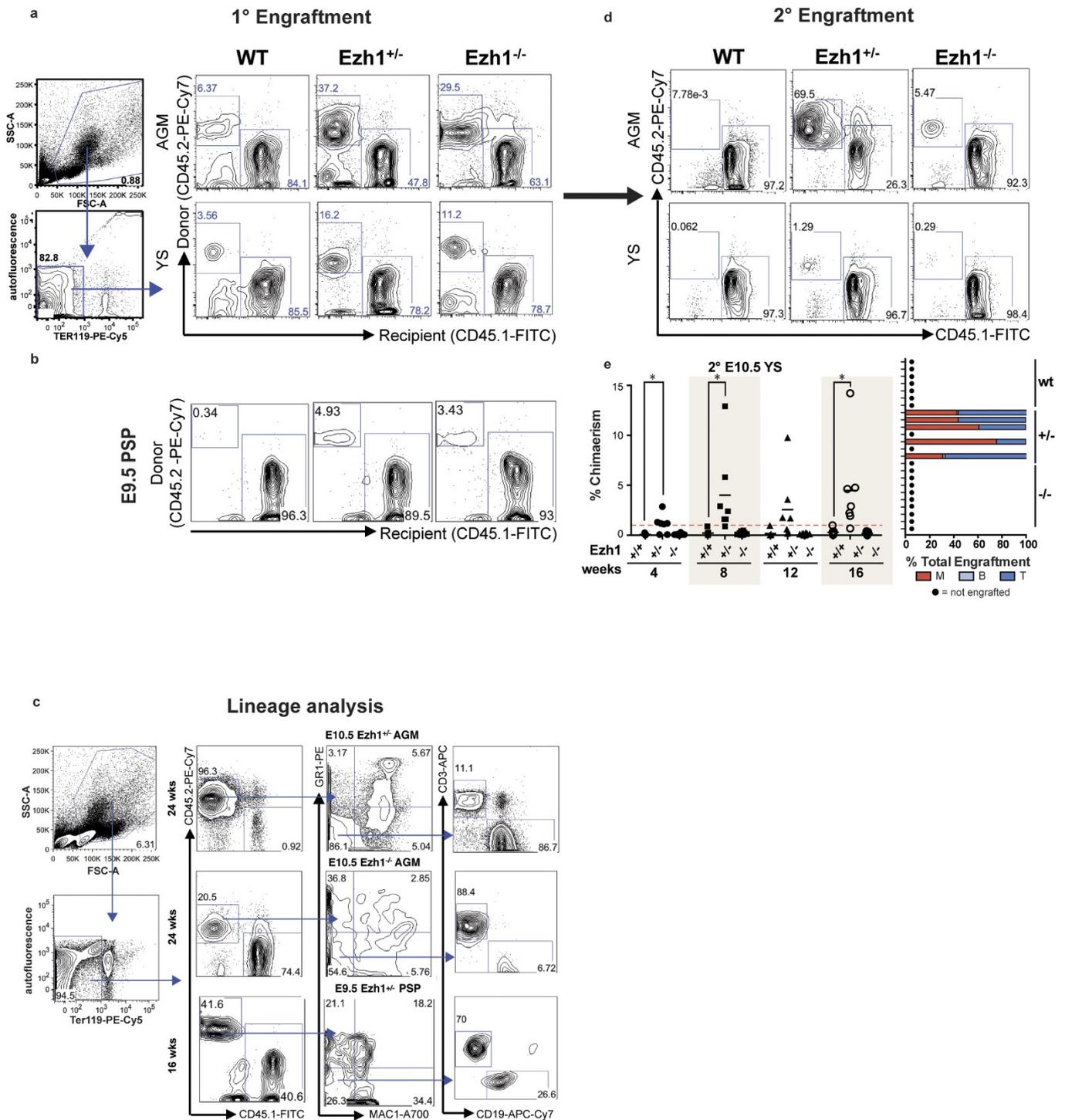


Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | Genome-wide chromatin occupancy reveals EZH1 enrichment at bivalent HSC genes and non-canonical active lymphoid genes.

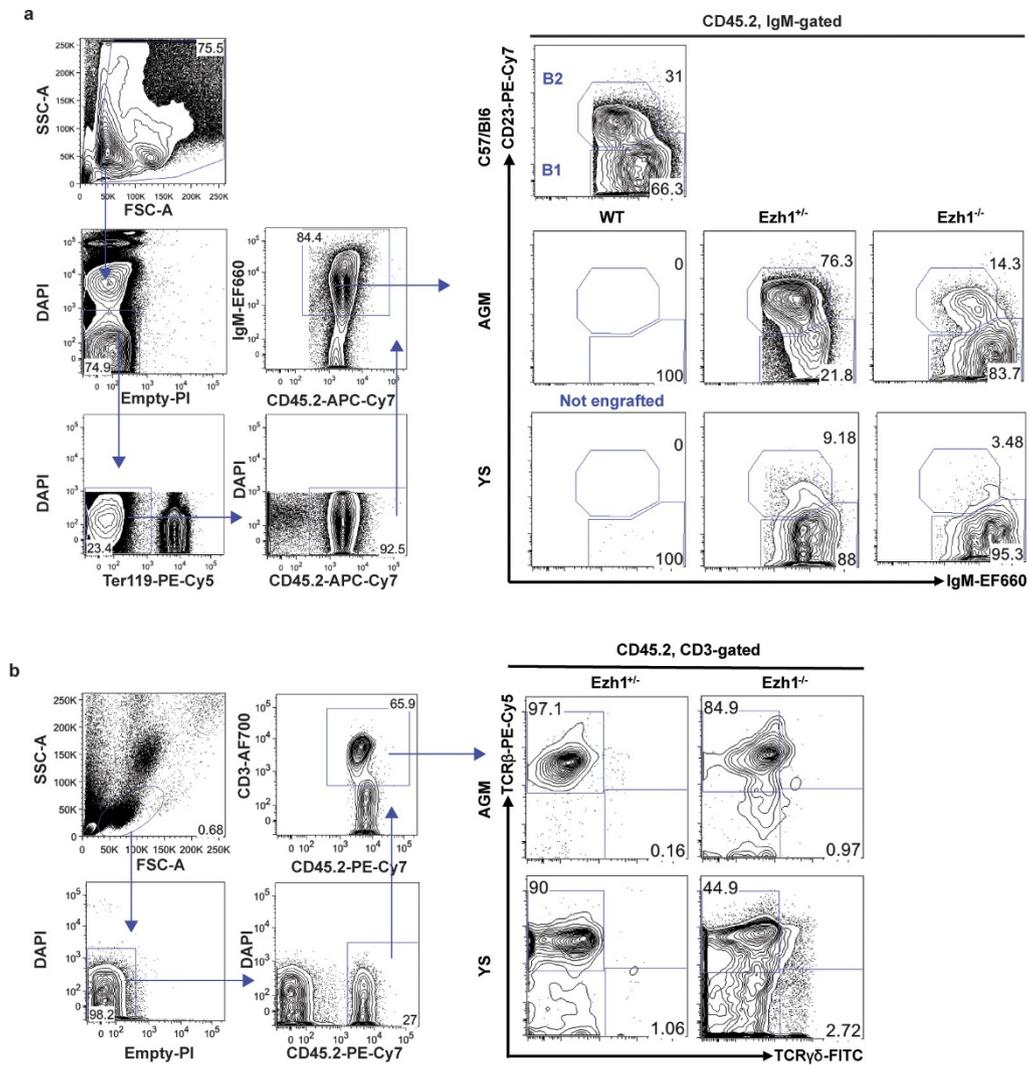
a. Breakdown of EZH1 binding at promoter regions and associated histone marks. **b–d.** GO term analysis of EZH1-bound active (**b**), bivalent (**c**) and repressed (**d**) genes. **e.** Distribution of EZH1-bound genes across the haematopoietic hierarchy (left) and their associated histone marks (right). A, active (H3K4me3-marked); B, bivalent (H3K4me3 and H3K27me3-marked); R, repressed (H3K27me3-marked). **f.** GSEA analysis of EZH1-bound genes correlated with RNA-seq data upon *EZH1* knockdown. **g.** Sankey diagram showing genome-wide changes in histone methylation status upon *EZH1* knockdown. **h.** Upregulated genes exhibit reciprocal decreases in H3K27me3 levels, as quantified by EpiChIP software. K4, H3K4me3; K27, H3K27me3. See also Fig. 3i. **i.** Activated (formerly bivalent) HSC genes exhibit increased gene expression upon *EZH1* knockdown and loss of H3K27me3. See also Fig. 3j. **j.** Correlations between changes in H3K27me3 and gene expression

levels upon *EZH1* knockdown, subdivided by subgroups corresponding to methylation changes. N, null. **k.** Breakdown of bivalent–bivalent (left), bivalent–repressed (centre) and bivalent–null (right) genes upon *EZH1* knockdown across the haematopoietic hierarchy. **l.** Overlap of EZH1- and EZH2-enriched peaks and the distribution of all EZH1-enriched, EZH2-enriched or common genes across the hierarchy (left), or specifically bivalent genes that become activated upon *EZH1* knockdown (middle) and active genes, marked by H3K4me3 in shLUC (right). **m.** *SUZ12* binding (from the ChEA database) across the haematopoietic hierarchy. **n.** Canonical and non-canonical previously identified targets¹⁷ across the haematopoietic hierarchy. **o, p.** Breakdown of histone marks on non-canonical ProB genes (**o**) and the genome-wide distribution from CEAS analysis³² (**p**). **q.** Changes in actively marked, non-canonical ProB genes (green bar in **o**), upon *EZH1* knockdown. **r.** *SUZ12* and *EZH2* binding (ChEA database) at ATAC peaks in *Ezh1*^{+/-} and *Ezh1*^{-/-} AGM. **P* < 0.05 by one-way ANOVA.



Extended Data Figure 5 | *Ezh1* deficiency enhances embryonic HSPC engraftment. **a**, Whole E10.5 AGM and yolk sac were transplanted intravenously into sublethally irradiated NSG adult females. Chimaerism was monitored by retroorbital bleeding every 4 weeks. Representative flow plots are shown for analysis after 4 weeks in $n \geq 3$ mice. See also Fig. 4e, f. **b**, Whole E9.5 PSP was transplanted intravenously into sublethally irradiated NSG adult females. Chimaerism was monitored via retroorbital bleeding every 4 weeks. Representative flow plots are shown for analysis after 8 weeks in $n \geq 3$ mice. See also Fig. 4g. **c**, Representative flow plots of lineage analysis in E10.5 AGM *Ezh1*^{+/-} and *Ezh1*^{-/-} primary transplant recipients after 24 weeks, and in E9.5 PSP *Ezh1*^{+/-} primary

transplant recipient after 16 weeks ($n \geq 3$ mice per group). See also Fig. 4e, g. **d**, Primary recipients in **a** were euthanized 24 weeks post-transplantation and 4×10^6 whole bone marrow was transplanted into sublethally irradiated adult NSG females intravenously. Chimaerism was monitored via retroorbital bleeding. Representative flow plots of E10.5 AGM and yolk sac secondary transplants after 4 weeks in $n \geq 3$ mice. See also Fig. 4h. **e**, Left, secondary transplantation of E10.5 yolk sac primary recipients (Fig. 4f). Right, lineage distribution of E10.5 secondary recipients. Data are pooled across three independent experiments. * $P < 0.05$, ** $P < 0.01$ by unpaired two-sided *t*-test; see Supplementary Information for exact *P* values per time point.



Extended Data Figure 6 | *Ezh1*-deficient embryonic HSPCs contribute to adult-type lymphopoiesis *in vivo*. **a**, Flow analysis of B1 and B2 progenitors in the peritoneal cavity of engrafted primary recipients ($n = 1$

mouse per group). **b**, Flow analysis of TCR β and TCR $\gamma\delta$ frequencies of donor-derived peripheral CD3⁺ T cells from engrafted primary recipients ($n = 1$ mouse per group). See also Fig. 4e, f.

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

► Experimental design

1. Sample size

Describe how sample size was determined.

At least n=3 animals were used per cohort, based on previous transplantation studies. Mice were assigned randomly to groups and not blinded.

2. Data exclusions

Describe any data exclusions.

No data was excluded

3. Replication

Describe whether the experimental findings were reliably reproduced.

Number of replicates per experiment are described in each figure legend.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

At least n=3 animals were used per cohort, based on previous transplantation studies. Mice were assigned randomly to groups and not blinded.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Mice were assigned randomly to groups and not blinded.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a | Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
- A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

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

7. Software

Describe the software used to analyze the data in this study.

All statistical calculations were performed using GraphPad Prism. Tests between two groups used two-tailed unpaired Student's t test. Data are presented as mean \pm s.e.m. Where indicated, ANOVA was used, with p-values less than 0.05 considered significant. Gene set enrichment analysis (GSEA) was run according to default parameters in their native implementations. Statistical enrichment of gene lists in gene ontology (GO) enrichment was performed using Fisher's exact test. ChIP-Seq raw reads were aligned to hg19 using bwa with default parameters and peaks were called using MACS14 with --nomodel. ATAC-Seq raw reads were aligned to hg19 or mm9 using bowtie1 with default parameters and peaks were called using MACS2 with --nomodel. RNA-seq reads were aligned to the human genome/transcriptome (hg19 and corresponding UCSC gene model) using a TopHat2 software with the following parameters (--library-type=fr-unstranded --min-intron-length=10 --no-coverage-search --microexon-search --min-isoform-fraction=0). To estimate expression levels for each gene, we counted aligned reads per gene using a htseq-count tool with the following parameters (--stranded=no --idattr=gene_id -t exon -i gene_name).

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

All unique materials are readily available from the authors.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

The following antibodies were used for human cells: CD45 APC-Cy7 (557833, BD Biosciences), CD4 PE-Cy5 (IM2636U, Beckman Coulter Immunotech), CD8 BV421 (RPA-T8, BD Horizon), CD5 BV510 (UCHT2, BD Biosciences), TCRgd APC (555718, BD Biosciences), TCRab BV510 (T10B9.1A-31, BD Biosciences), CD3 PE-Cy7 (UCHT1, BD Pharmingen), CD7 PE (555361, BD Pharmingen), CD1a APC (559775, BD Pharmingen) for T cell staining. For B cell staining: CD45 PE-Cy5 (IM2652U, Beckman Coulter Immunotech), CD19 PE (4G7, BD), CD56 V450 (B159, BD Biosciences), CD11b APC-Cy7 (557754, BD Biosciences), For HSC/Progenitor sorting: CD34 PE-Cy7 (8G12, BD), CD45 APC-Cy7 (557833, BD Biosciences), CD38 PE-Cy5 (IM2651U, BD), DAPI. For myeloid and erythroid staining: CD11b APC-Cy7 (557754, BD Biosciences), GLYA PE-Cy7 (A71564, Beckman Coulter), CD71 PE (555537, BD Biosciences), CD45 PE-Cy5 (IM2652U, Beckman Coulter Immunotech). All stains were performed with $<1 \times 10^6$ cells per 100 μL staining buffer (PBS + 2% FBS) with 1:100 dilution of each antibody, 30 min at RT in dark. Compensation was performed by automated compensation with anti-mouse Igk and negative beads (BD). All acquisitions were performed on BD Fortessa or BD Aria cytometer. All human antibodies were validated using human cord blood mononuclear cells (MNCs) as a positive control.

The following antibodies were used for mouse cells: CD45.2 PE-Cy7 (104, eBioscience), CD45.1 FITC (A20, eBioscience), B220 PB (RA3-6B2, BD Biosciences), Ter119 PE-Cy5 (Ter 119, eBioscience), GR1 (RB6-8C5, BD Bioscience), CD3 APC (145-2C11, eBioscience), CD19 APC-Cy7 (1D3, BD Bioscience), MAC1 A700 (M1/70, BD Bioscience) for engraftment analyses. For B cell staining: CD45.2 APC-CY7 (104, Bio-Legend), CD23 PE-Cy7 (B3B4, eBioscience), Ter119 PE-Cy5 (Ter 119, eBioscience), MAC1 A700 (M1/70, BD Bioscience), CD5 BV510 (53-7.3 BD Biosciences), IgM EF660 (11/41, eBioscience). For T cell staining: CD45.2 PE-Cy7 (104, eBioscience), TCRb PE-Cy5 (H57-597, BD Biosciences), CD8 APC-EF780 (53-6.7, eBioscience), CD4 APC (GK1.5, eBioscience), CD3 AF700 (17A2, BioLegend), TCRgd FITC (GL3, BD Biosciences). For HSPC sorting: CD16/32 (93, Biolegend), Ter119 Biotin (Ter119, eBioscience), Gr-1 Biotin (RB6-8C5, eBioscience), CD3 Biotin (17A2, eBioscience), CD5 Biotin (53-7.3, eBioscience), CD8 Biotin (53-6.7, eBioscience), CD19 Biotin (eBio1D3, eBioscience), Streptavidin eFluor450 (eBioscience), CD45 PerCP-Cy5.5 (30-F11, eBioscience), CD144 eFluor660 (eBioBV13, eBioscience), CD117 APC-eFluor 780 (2B8, eBioscience), CD41 PE-Cy7 (eBioMWRReg30, eBioscience). All stains were performed with $<1 \times 10^6$ cells per 100 μL staining buffer (PBS + 2% FBS) with 1:100 dilution of each antibody, 30 min on ice in dark. Compensation was performed by automated compensation with anti-rat and anti-hamster Igk and negative beads (BD). All acquisitions were performed on BD Fortessa or BD Aria cytometer. All mouse antibodies were validated using mouse peripheral blood mononuclear cells as a positive control.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

All experiments were performed using MSC-iPS1, CD34-iPS and CD45-iPS, obtained from the Boston Children's Hospital Human Embryonic Stem Cell Core (hESC) and verified by immunohistochemistry for pluripotency markers, teratoma formation and karyotyping. All cells were routinely tested for mycoplasma contamination.

b. Describe the method of cell line authentication used.

All experiments were performed using MSC-iPS182, CD34-iPS and CD45-iPS, obtained from the Boston Children's Hospital Human Embryonic Stem Cell Core (hESC) and verified by immunohistochemistry for pluripotency markers, teratoma formation and karyotyping. All cells were routinely tested for mycoplasma contamination.

c. Report whether the cell lines were tested for mycoplasma contamination.

All cells were routinely tested for mycoplasma contamination.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

No commonly misidentified cell lines were used.

► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

NOD/LtSz-scidIL2Rgnull (NSG) (Jackson Labs) mice were bred and housed at the Boston Children's Hospital animal care facility. Animal experiments were performed in accordance to institutional guidelines approved by BCH animal care committee. At least n=3 animals were used per cohort, based on previous transplantation studies. Mice were assigned randomly to groups and not blinded. Briefly, 8-12 week old mice were irradiated (275 rads) 24 hours before transplant. To ensure consistency between experiments, only female mice were used.

Policy information about [studies involving human research participants](#)

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

This study did not involve human research participants.

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

▶ Data presentation

For all flow cytometry data, confirm that:

- 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- 3. All plots are contour plots with outliers or pseudocolor plots.
- 4. A numerical value for number of cells or percentage (with statistics) is provided.

▶ Methodological details

- | | |
|--|--|
| 5. Describe the sample preparation. | All stains were performed with $<1 \times 10^6$ cells per 100 μ L staining buffer (PBS + 2% FBS) with 1:100 dilution of each antibody, 30 min on ice in dark. Compensation was performed by automated compensation with anti-mouse Igk and negative beads (BD). |
| 6. Identify the instrument used for data collection. | All acquisitions were performed on BD Fortessa or BD Aria cytometer. |
| 7. Describe the software used to collect and analyze the flow cytometry data. | All flow cytometry data was analyzed using FlowJo 8.7. |
| 8. Describe the abundance of the relevant cell populations within post-sort fractions. | Sorted samples were confirmed for purity post-sort via flow cytometry. Sorted populations were confirmed to be >95% purity. |
| 9. Describe the gating strategy used. | <p>All human cells were first gated on FSC/SSC according to cell size and granularity, using stained human cord blood mononuclear cells (MNCs) as a positive control and reference for cell size, granularity and staining intensity. All human hematopoietic populations were subsequently gated on CD45+, as indicated. Unstained samples were used to set up negative gates, and stained human MNCs were used to set up positive gates. Dead cell populations were excluded using DAPI staining.</p> <p>All mouse cells were first gated on FSC/SSC according to cell size and granularity using mouse peripheral blood mononuclear cells as a positive control and reference for cell size, granularity and staining intensity. All populations are gated on mouse CD45. Mouse red blood cells are excluded based on Ter119-. Unstained samples were used to set up negative gates, and stained mouse peripheral blood mononuclear cells from C57/Bl6 (CD45.2+, donor) and NSG (CD45.1, recipient) were used to set up positive gates.</p> |

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

ChIP-seq Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

▶ Data deposition

1. For all ChIP-seq data:

- a. Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- b. Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

2. Provide all necessary reviewer access links.

The entry may remain private before publication.

GSE89417

Reviewer access link:

https://urldefense.proofpoint.com/v2/url?u=https-3A__www.ncbi.nlm.nih.gov_geo_query_acc.cgi-3Ftoken-3Dmlylwqaizrufflyn-26acc-3DGSE89418&d=DQIBAg&c=qS4goWBT7poplM69zy_3xhKwEW14JZMSdioCoppxeFU&r=wLJRYiBrsG-PS2GD4hZHKWQq8iQjOcUKvJHPHyGzlgFmgig0WcXViH4r8aeQ76Ao&m=YXVat83CVCZixJ6s2rlaaUHD1dgh0W0vL1EmdvaqTz0&s=r-85jzzF3NnNgS-QoAjNcmBmL2cqBt8VRie8HHeit8&e=

3. Provide a list of all files available in the database submission.

.wig, Peaks .bed files and raw .fastq files for all samples:

ChIP-seq_5F_BirA-Only_Input
ChIP-seq_5F_BirA-Only_rep1
ChIP-seq_5F_BirA-Only_rep2
ChIP-seq_5F_FB-Ezh1_rep1
ChIP-seq_5F_FB-Ezh1_rep2
ChIP-seq_5F_H3K4me3
ChIP-seq_5F_H3K27me3
ChIP-seq_5F_shLuc_H3K4me3_rep1
ChIP-seq_5F_shLuc_H3K4me3_rep2
ChIP-seq_5F_shEZH1_H3K4me3_rep1
ChIP-seq_5F_shEZH1_H3K4me3_rep2
ChIP-seq_5F_shLuc_H3K27me3_rep1
ChIP-seq_5F_shLuc_H3K27me3_rep2
ChIP-seq_5F_shEZH1_H3K27me3_rep1
ChIP-seq_5F_shEZH1_H3K27me3_rep2

4. If available, provide a link to an anonymized genome browser session (e.g. [UCSC](#)).

N/A

▶ Methodological details

5. Describe the experimental replicates.

Two biological replicate ChIP-seq experiments were performed for 5F-BirA-Only (control), FB-EZH1, H3K4me3 and H3K27me3 ChIP-seq in shLuc (control) and shEZH1 5F cells.

6. Describe the sequencing depth for each experiment.

ChIP-seq_5F_BirA-Only_Input: 39,071,701
ChIP-seq_5F_BirA-Only_rep1: 15,496,062
ChIP-seq_5F_FB-Ezh1_rep1: 15,605,335
ChIP-seq_5F_BirA-Only_rep2: 15,762,602
ChIP-seq_5F_FB-Ezh1_rep2: 12,394,022
ChIP-seq_5F_H3K4me3: 29,420,109
ChIP-seq_5F_H3K27me3: 21,638,483
ChIP-seq_5F_H3K4me3_shLuc_rep1: 15,050,713

- ChIP-seq_5F_H3K4me3_shLuc_rep2: 12,869,259
ChIP-seq_5F_H3K4me3_shEZH1_rep1: 17,808,384
ChIP-seq_5F_H3K4me3_shEZH1_rep2: 14,077,682
ChIP-seq_5F_H3K27me3_shLuc_rep1: 12,205,582
ChIP-seq_5F_H3K27me3_shLuc_rep2: 11,776,746
ChIP-seq_5F_H3K27me3_shEZH1_rep1: 13,239,898
ChIP-seq_5F_H3K27me3_shEZH1_rep2: 10,510,864
7. Describe the antibodies used for the ChIP-seq experiments.
- H3K4me3: Millipore 04-745, lot: 1997493
H3K27me3: Millipore 07-449, lot: 2064519
Dynabeads™ MyOne™ Streptavidin T1: Thermo-Fisher 65601, lot:123612210
8. Describe the peak calling parameters.
- Peak detection was performed with the Model-based Analysis of ChIP-Seq (MACS) algorithm (<http://liulab.dfci.harvard.edu/MACS/>) using default parameter.
9. Describe the methods used to ensure data quality.
- ChIP-seq peaks were identified by MACS using p-value 1×10^{-5} as the cutoff. All peaks have p-values $< 1 \times 10^{-5}$.
10. Describe the software used to collect and analyze the ChIP-seq data.
- Sequencing reads were aligned to human genome assembly hg19 (NCBI version 37) using Bowtie v0.12.7 with the following parameters: `-v 2 -m 3 --strata --best`. Duplicate reads were removed after the alignment with the Picard command-line tools. The wig files were generated by a moving window of size 200bp. The tag count in the window was further normalized in unit RPKM (# read per kb per million total reads) for ChIP-seq data generated by NextSeq500.