# MLL3/MLL4 methyltransferase activities regulate embryonic stem cell differentiation independent of enhancer H3K4me1

Guojia Xie[1], Ji-Eun Lee[1], Kaitlin McKernan[1], Young-Kwon Park[1], Younghoon Jang[1], Chengyu Liu[2], Weiqun Peng[3] and Kai Ge[1]*

[1]Laboratory of Endocrinology and Receptor Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

[2]Transgenic Core, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA

[3]Departments of Physics and Anatomy and Cell Biology, The George Washington University, Washington, DC 20052, USA

*To whom correspondence should be addressed. (Email: kai.ge@nih.gov)

**Highlights**

- Simultaneous elimination of MLL3 and MLL4 enzymatic activities leads to early embryonic lethality in mice
- MLL3/4 enzymatic activities are dispensable for ESC differentiation towards the three germ layers
- ESCs lacking MLL3/4 enzymatic activities show cavitation defects during EB differentiation, likely due to impaired VE induction
- MLL3/4-catalyzed H3K4me1 is dispensable for enhancer activation in ESC differentiation

**Abstract**

Enhancers drive cell-type-specific gene transcription and are marked by H3K4me1. MLL4 (KMT2D), a major H3K4me1 methyltransferase with partial functional redundancy with MLL3 (KMT2C), is critical for enhancer activation and cell-type-specific gene induction during cell differentiation and development. However, the roles of MLL3/4-mediated enhancer H3K4me1 and MLL3/4 enzymatic activities in general in these processes remain unclear. Here, we report that MLL3/4 enzymatic activities are partially redundant during mouse development. Simultaneous elimination of both leads to embryonic lethality around E8.5. Using embryoid body (EB) differentiation as an *in vitro* model for early embryonic development, we show that *Mll3* knockout MLL4 enzyme-dead embryonic stem cells (ESCs) are capable of differentiating towards the three germ layers but display severe cavitation defects, likely due to impaired induction of visceral endoderm. Importantly, MLL3/4-catalyzed H3K4me1 is dispensable for enhancer activation during early EB differentiation and lineage-specific neural differentiation. Together, these results suggest a critical, but enhancer H3K4me1-independent, role of MLL3/4 enzymatic activities in early embryonic development and ESC differentiation.

## Introduction

Enhancers are *cis*-regulatory DNA elements recognized by transcription factors (TFs). They communicate with promoters to regulate cell-type-specific gene expression and cell identity. In-depth epigenomic research has uncovered the chromatin signatures of enhancers. Primed enhancers, which are marked by H3K4me1, become activated with the addition of H3K27ac[1,2]. The activation of enhancers during cell fate transition is enabled by several epigenomic regulators sequentially recruited by lineage determining TFs, including histone mono-methyltransferases MLL3 (KMT2C) and MLL4 (KMT2D) followed by histone acetyltransferases CBP and p300, which catalyze the placement of H3K4me1 and H3K27ac, respectively[3-6]. However, the exact role of these epigenomic regulators themselves versus the histone modifications they catalyze, particularly MLL3/4 versus MLL3/4-catalyzed H3K4me1, in enhancer activation has remained elusive.

MLL3 and MLL4 (MLL3/4) are members of the Set1-like family of mammalian H3K4 methyltransferases that are responsible for catalyzing H3K4me1[4]. They are the largest known nuclear proteins (4,903 and 5,588 amino acids in mice, respectively), and associate with the WRAD (WDR5, RbBP5, ASH2L, DPY30) subcomplex as well as NCOA6, UTX, PA1, and PTIP in a large multi-subunit complex[7,8]. Enzymatic activities of MLL3/4 are conferred by the C-terminal SET domain, which is also required for maintaining their protein stability[9,10]. Consistent with their critical roles in enhancer activation, MLL3/4 are broadly required for normal development and cell differentiation. *Mll4* knockout (KO) in mice leads to lethality around embryonic day (E) 9.5. Deletion of *Mll3* has a milder effect, as *Mll3* KO mice die at birth, suggesting that MLL4 plays a dominant role in embryonic development[4,7]. Previous work in mice has also shown that MLL3/4 are required for the development of tissues such as adipose, muscle, heart, B cells, T cells and mammary gland[4,11-14]. In humans, *MLL3/4* are frequently mutated in developmental diseases and cancers[15]. However, the functional role of MLL3/4 enzymatic activities in development and cell differentiation is poorly understood.

During mouse early embryonic development from implantation to gastrulation, the blastocyst undergoes an orchestrated series of lineage specification events and develops into the three germ layers (ectoderm, mesoderm and definitive endoderm), which contain progenitors of all fetal tissues[16]. Visceral

endoderm (VE), one major type of extraembryonic endoderm, is an epithelial layer of cells that surrounds the post-implantation embryo. Anterior VE migration induces early embryonic asymmetry by E6.5, when primitive streak (PS), the precursor of mesoderm and definitive endoderm, forms on the posterior side of the embryo[17]. Proper VE development is required for correct patterning of PS and early embryonic development[18,19]. Embryonic stem cells (ESCs) are self-renewing pluripotent cells that are derived from the inner cell mass of blastocysts[20]. ESCs can differentiate and organize into three-dimensional cavitated structures called embryoid bodies (EBs). EB differentiation of ESCs is a valuable *in vitro* model recapitulating many aspects of early embryonic development, during which the three germ layers form[21]. Under the treatment of defined factors, ESCs can also differentiate homogeneously into specific lineages such as neurons[22].

We reported previously that MLL3/4 are required for enhancer activation and ESC differentiation, but are largely dispensable for ESC identity maintenance[5]. A recent study showed that MLL3/4-catalyzed H3K4me1 is mostly dispensable for maintaining active enhancers in undifferentiated ESCs[9]. These findings motivated us to investigate the roles of MLL3/4 enzymatic activities and MLL3/4-mediated enhancer H3K4me1 during development and cell differentiation. Using MLL3/4 enzyme-dead single and double knockin mice generated by CRISPR/Cas9, we found that MLL3/4 enzymatic activities are partially redundant and are essential for early embryonic development. By knocking-in MLL4 enzyme-dead point mutation in *Mll3* KO ESCs, we observed that eliminating MLL3/4 enzymatic activities has little effects on ESC differentiation towards all three germ layers but results in defective cavitation and cardiomyogenesis during EB differentiation. The cavitation defect is likely the consequence of impaired VE induction. Finally, using EB differentiation and neural differentiation as model systems, we demonstrated that MLL3/4-catalyzed H3K4me1 is dispensable for enhancer activation during ESC differentiation.

## Results

### MLL3/4 enzymatic activities are essential for early embryonic development in mice

MLL3 and MLL4 are major enhancer H3K4me1 methyltransferases[4]. MLL3 Tyr 4792 (Y4792) and MLL4 Tyr5477 (Y5477) located in the catalytic SET domain are conserved throughout different histone methyltransferases (Figure 1a; Figure 1-S1) and are essential for enzymatic activity[9]. To investigate the roles of MLL3/4 enzymatic activities in mouse development, we introduced the enzyme-dead point mutations Y4792A and Y5477A to *Mll3* and *Mll4* gene loci, respectively, by injecting CRISPR/Cas9 components into wild type zygotes (Figure 1b). After germline transmission, mice heterozygous for *Mll3* KI and *Mll4* KI (*Mll3*$^{KI/+}$ and *Mll4*$^{KI/+}$) survived without any discernible phenotypes and were inbred to obtain homozygotes. While *Mll3*$^{-/-}$ mice display perinatal lethality[4], *Mll3*$^{KI/KI}$ mice survived to adulthood with reduced numbers (Figure 1c). Adult *Mll3*$^{KI/KI}$ mice were fertile and showed similar body weights as wild type or *Mll3*$^{KI/+}$ littermates (data not shown). *Mll4*$^{-/-}$ mice die around embryonic day (E) 9.5[4,7]. In contrast, *Mll4*$^{KI/KI}$ mice died around birth with no obvious morphological abnormalities (Figure 1d). These results suggest that MLL3 and MLL4 proteins play more crucial roles than their enzymatic activities in mouse development.

Since MLL3 and MLL4 are partially redundant *in vitro* and in adipose tissue development[4,10], we crossed *Mll3* KI and *Mll4* KI mice and obtained *Mll3*/*Mll4* double KI strains. Both *Mll3*$^{KI/+}$;*Mll4*$^{KI/+}$ and *Mll3*$^{KI/KI}$;*Mll4*$^{KI/+}$ mice survived to adulthood and were fertile. We set up self-crossing of *Mll3*$^{KI/+}$;*Mll4*$^{KI/+}$ mice but failed to recover homozygous *Mll3*/*Mll4* double KI (*Mll3*$^{KI/KI}$;*Mll4*$^{KI/KI}$) embryos at E14.5 (Figure 1e). To determine at which stage *Mll3*$^{KI/KI}$;*Mll4*$^{KI/KI}$ embryos die, we set up self-crossing of *Mll3*$^{KI/KI}$;*Mll4*$^{KI/+}$ mice. *Mll3*$^{KI/KI}$;*Mll4*$^{KI/KI}$ embryos were detected at E8.0 and E8.5 but not at any later stage (Figure 1f). All *Mll3*$^{KI/KI}$;*Mll4*$^{+/+}$ and the majority of *Mll3*$^{KI/KI}$;*Mll4*$^{KI/+}$ embryos displayed normal developmental morphology and exhibited characteristics of grooved neural plate and unturned early-somite-stage embryo at E8.0 and E8.5, respectively. However, *Mll3*$^{KI/KI}$;*Mll4*$^{KI/KI}$ embryos displayed delayed growth at E8.0; they morphologically resembled normal embryos around E5.5-6.5 with an egg cylinder-like structure. At E8.5, all *Mll3*$^{KI/KI}$;*Mll4*$^{KI/KI}$ embryos were much smaller and exhibited severe morphological defects (Figure 1g). Together, these data indicate that MLL3 and MLL4 enzymatic activities are partially redundant during

embryonic development, and that simultaneous elimination of both results in early embryonic lethality around E8.5.

**ESCs lacking MLL3/4 enzymatic activities maintain cell identity**

MLL4 is required for early embryonic development in mice while MLL3 is not[4]. To investigate the function of MLL3/4 enzymatic activities *in vitro*, we took advantage of our previously derived *Mll3*$^{-/-}$;*Mll4*$^{f/f}$ (f/f) ESCs to eliminate the potential redundancy between MLL3 and MLL4[5]. Using CRISPR/Cas9 gene editing, we knocked-in the enzyme-dead point mutation Y5477A to both alleles of *Mll4* in f/f cells and generated independent *Mll3*$^{-/-}$;*Mll4*$^{KI/KI}$ (KI, knockin) ES cell lines (Figure 2-S1a). We also re-generated *Mll3*/*Mll4* double knockout (KO) ESCs by transiently transfecting a Cre-expressing plasmid into f/f ESCs[5]. The genotypes of all cell lines were confirmed by genomic PCR (Figure 2-S1b-c), and the *Mll4* Y5477A point mutation in KI cell lines was confirmed by sequencing (Figure 2-S1d). Further validation of KI cells found no evidence of potential exonic off-target mutations (Figure 2-S1e-h). MLL4 was expressed at comparable levels in f/f and KI cells but absent in KO cells, and MLL3/4-associated protein UTX followed a similar pattern (Figure 2a). Other chromatin regulators including CBP, BRG1 and ARID1A did not exhibit observable changes among the panel of cells (Figure 2a). As expected, immunoblotting of histone extracts revealed similarly reduced levels of global H3K4me1 and H3K4me2, but not H3K4me3, in KI and KO ESCs (Figure 2b).

Next, we characterized the morphology and cell growth rate of KI ESCs. When cultured in 2i+LIF medium, neither KO nor KI ESCs could form regular dome-like colonies shown in f/f ESCs. Specifically, KI colonies were flat, and cells grew homogenously in monolayer (Figure 2c). In agreement with the previous report[5], f/f and KO ESCs displayed similar population doubling time. Surprisingly, KI ESCs grew faster and displayed ~20% decrease in population doubling time (Figure 2d). By RT-qPCR analysis of ESC identity genes, we observed that mRNA levels of *Nanog*, *Oct4*, *Rex1* and *Esrrb* were comparable among all cell lines (Figure 2e). Only the *Sox2* level decreased about two-fold in both KI and KO ESCs as reported previously[23]. In addition, both f/f and KI ESCs presented positive alkaline phosphatase (AP) staining (Figure

2c). Together, these data indicate that while MLL3/4 enzymatic activities are dispensable for maintaining ESC identity, KI ESCs show monolayer growth and increased proliferation.

**ESCs lacking MLL3/4 enzymatic activities are capable of differentiating towards the three germ layers, but show defects in EB cavitation and cardiomyogenesis**

We next investigated the roles of MLL3/4 enzymatic activities in embryoid body (EB) differentiation, which is an *in vitro* model that mimics early embryonic development[21]. ESCs were aggregated in hanging drops in the absence of LIF to form EBs, and the outgrowths were kept in culture for spontaneous differentiation (Figure 3-S1a). During EB differentiation, UTX was expressed at similar levels between f/f and KI cells (Figure 3-S1b), suggesting that the loss of enzymatic activity did not affect MLL4 level, which dictates the stability of UTX protein[4,24]. As expected, global H3K4me1 and, to a lesser extent, H3K4me2 decreased in KI cells during EB differentiation (Figure 3-S1c).

Consistent with the previous report[5], f/f ESCs formed large cystic EBs during differentiation whereas KO cells displayed severe defects (Figure 3a; Figure 3-S1d). Although EBs derived from KI ESCs exhibited similar sizes as f/f EBs throughout differentiation (Figure 3-S1d), none of KI EBs displayed cystic cavities at day 7 (D7) and day 10 (D10), when these structures were observed in the majority of f/f EBs (Figure 3a). By RT-qPCR analysis, we confirmed that during differentiation of f/f cells, expression of ESC identity genes *Nanog* and *Oct4* decreased dramatically, whereas germ layer markers for endoderm (*Gata4* and *Afp*), mesoderm (*T* and *Kdr*) and ectoderm (*Nes* and *Cdh2*) increased markedly (Figure 3b). In contrast to the severe defects observed in KO cells, KI cells did not exhibit appreciable impairment in ESC identity gene decommissioning or germ layer marker induction (Figure 3b). Consistently, NANOG protein decreased to undetectable levels during EB differentiation in both f/f and KI cells (Figure 3-S1b).

In addition to the cavitation defects during EB differentiation, KI ESCs also exhibited impaired capacity of cardiomyogenesis. After 8 days of differentiation, less than 20% of KI EBs contained contracting clusters, while 100% of f/f EBs displayed spontaneous contraction (Figure 3-S1e). Although the percentage of beating EBs derived from KI ESCs gradually rose to nearly 100% after day 12, KI EBs displayed obvious

morphologic changes compared to f/f EBs, evidenced by decreased beating frequency and smaller contracting clusters (Figure 3-S1f-g; Supplementary Videos). Correlated with the phenotype, RT-qPCR analysis revealed the induction of cardiac progenitor marker *Tbx5* and cardiac myofilament genes *Myh6*, *Myl2* and *Ttn* was profoundly compromised during differentiation of KI ESCs (Figure 3-S1h).

Next, we performed teratoma assays to validate the differentiation capacity of KI ESCs *in vivo*. Both f/f and KI ESCs developed into typical teratomas in which all three germ layers formed (Figure 3c). However, KI teratomas failed to form striated muscle-like tissues that were observed in f/f teratomas (Figure 3c). Together, these data indicate that while KI ESCs are generally capable of differentiating towards the three germ layers, they show specific defects in cavitation and cardiomyogenesis.

**Loss of MLL3/4 enzymatic activities impairs visceral endoderm induction in EB differentiation**

Sequential expression of marker genes during EB differentiation reflects successive developmental stages[21,25,26]. With an extraembryonic endoderm layer established on the surface, D3-D4 EBs transiently express primitive streak (PS) markers and are equivalent to the early gastrulation stage around E6.5. After D4, specialized cell lineages of the three germ layers develop (Figure 4a). Cavitation starts around D4-D5 and gives rise to large and cystic structures similar to the proamniotic cavity in embryos[21]. To further investigate the roles of MLL3/4 enzymatic activities during differentiation and the mechanism underlying the cavitation defects of KI EBs, we performed RNA-seq at ESC (D0) and EB (D4, D10) stages. Principal component analysis (PCA) revealed that KI cells, but not KO cells, had similar differentiation trajectories with f/f cells (Figure 4b).

Unsupervised K-means clustering of differentially expressed genes identified 12 clusters. Each cluster was associated with distinct GO terms (Figure 4c-d; Figure 4-S1a). Among the 12 clusters, only 4 (I, III, IV, IX) displayed evident discrepancies between f/f and KI cells at D4 or D10. Cluster I genes were markedly upregulated at D10 in f/f EBs but showed reduced expression levels in KI EBs. These genes were enriched for muscle and cardiovascular development terms, consistent with the cardiomyogenesis defects of KI EBs (Figure 4-S1a). Cluster III and IX genes were induced from D0 to D4 in f/f EBs with impaired

induction in KI EBs. Interestingly, these genes were enriched for early embryonic development terms such as gastrulation and anterior/posterior pattern specification (Figure 4d), which depend on the development of visceral endoderm (VE) and PS[17]. Gene set enrichment analysis (GSEA) using cell-type markers defined previously by single-cell RNA-seq in E6.5 embryos[27] revealed that strikingly, VE markers were significantly enriched in genes downregulated in KI D4 EBs, while markers of PS or nascent mesoderm (NM), an early derivative of PS, showed no enrichment (Figure 4e; Figure 4-S1b).

We also compared the expression of reported VE and PS genes between f/f and KI D4 EBs[17,28]. All 16 VE genes were expressed at lower levels in KI EBs, with 11 of them showing over two-fold decreases. In contrast, no PS genes changed more than two-fold (Figure 4f; Figure 4-S1c). BMP, Wnt and Nodal signaling pathways play key roles in early embryonic development[16]. Among components of these pathways, *Bmp2* and *Bmp4*, which encode two ligands critical for VE differentiation[29], decreased more than 2-fold in KI EBs at D4 (Figure 4-S1d). Together, these data indicate that MLL3/4 enzymatic activities are required for induction of VE, which is known to be required for cavitation during EB differentiation[30].

**MLL3/4 enzymatic activities are required for induction of visceral endoderm TFs GATA4/6 in the early phase of EB differentiation**

To understand how MLL3/4 methyltransferase activities regulate VE induction, we conducted ChIP-seq of MLL4, H3K4me1 and H3K27ac in ESCs and D4 EBs. We found that loss of enzymatic activity resulted in redistribution of MLL4 genomic occupancy in KI EBs without discernible changes of total binding intensity (Figure 5-S1). We identified 10,767 de novo MLL4$^{+}$ AEs that were present in D4 f/f or KI EBs but not in ESCs, and divided them into three groups based on the comparison of MLL4 binding intensities between f/f and KI EBs. Among the 10,767 de novo MLL4$^{+}$ AEs, 14% (1,504; Group I), 34% (3,610; Group II) and 52% (5,653; Group III) showed increased, unchanged and decreased levels of MLL4 occupancies in KI EBs compared to f/f EBs, respectively (Figure 5a). Among them, only Group III was overrepresented in VE marker-associated AEs (Figure 5b). Motifs of ESC master TFs NANOG and OCT4 were enriched in Group I; motifs of T-box factors EOMES and MGA and Wnt pathway effector TCF4 were enriched in Group II; and

motifs of GATA family TFs were enriched in Group III (Figure 5c). Consistent with changes in MLL4 binding, TFs linked to Group I, II or III exhibited increased, unchanged or decreased expression levels in KI EBs compared to f/f EBs (Figure 5d), suggesting the notion that differentially expressed TFs contribute to the redistribution of MLL4 genomic binding.

GATA4 and GATA6 are lineage-determining TFs essential for VE development in mouse embryos[31,32]. Motifs of GATA family TFs including GATA4/6 were highly enriched in de novo MLL4$^{+}$ AEs associated with VE markers but not PS or NM markers (Figure 5-S2). In addition, GATA6 binding sites in extraembryonic endoderm (GSE69323)[33] colocalized with Group III de novo MLL4$^{+}$ AEs around representative VE genes, whose expression levels decreased in KI D4 EBs (Figure 5e). These observations indicate that MLL3/4 enzymatic activities are required for induction of VE transcription factors GATA4/6 in the early phase of EB differentiation.

**MLL3/4-catalyzed H3K4me1 is dispensable for enhancer activation during early EB differentiation**

We reported previously that MLL3/4 proteins are required for de novo enhancer activation in D4 EBs[5]. To test if this requirement is mediated by MLL3/4-catalyzed H3K4me1, we examined distributions of H3K4me1, H3K27ac and chromatin accessibility on Group II de novo MLL4$^{+}$ AEs where f/f and KI EBs displayed similar levels of MLL4 binding. We observed dramatic induction of H3K4me1, H3K27ac and chromatin accessibility from ESC to D4 EB in f/f cells. In KI EBs, despite no gain of H3K4me1, Group II AEs showed comparable induction of H3K27ac and chromatin accessibility as in f/f EBs (Figure 6a-c; Figure 6-S1). Using ChIP-seq data from KO cells[5], we further confirmed the requirement of MLL3/4 proteins for the activation of Group II de novo MLL4$^{+}$ AEs. Representative genes near Group II AEs were induced comparably in f/f and KI but not KO EBs (Figure 6-S2). On Group I and III AEs, induction patterns of H3K27ac and chromatin accessibility in f/f and KI EBs were consistent with those of MLL4 but not H3K4me1 (Figure 6a-b). Together,

these data indicate that while MLL3/4 proteins are required for enhancer activation during early EB differentiation, MLL3/4-catalyzed H3K4me1 is dispensable.

**MLL3/4-catalyzed H3K4me1 is dispensable for activation of super-enhancers during neural differentiation**

We also investigated the role of MLL3/4-catalyzed H3K4me1 in enhancer activation during lineage-specific differentiation. Following the retinoid acid (RA)-induced neural differentiation protocol[22] (Figure 7-S1a), we observed that ESCs differentiated homogeneously into neural progenitor cells (NPCs) around day 8 (D8) with the characteristics of spindle-like shape and neural rosette. After day 10, NPCs gradually differentiated into mature glutamatergic neurons until day 16 (D16), when a dense neuritic network formed. Interestingly, no morphological differences could be observed between f/f and KI cells during RA-induced neural differentiation (Figure 7a; Figure 7-S1b). RNA-seq analysis in NPCs and neurons revealed that only induced genes with comparable expression levels between f/f and KI cells were functionally associated with nervous system development (Figure 7b-c; Figure 7-S1c-d). NPC markers *Nes*, *Sox3*, *Pax6* and *Pax7* and neuron markers *NeuN*, *Tau*, *Ntrk2* and *App* were induced comparably in f/f and KI cells (Figure 7-S1e-f). These results indicate that MLL3/4 enzymatic activities are dispensable for cell-identity gene induction during neural differentiation.

Super-enhancers (SEs) are clusters of AEs bound by lineage-specific master TFs and drive high-level expression of cell-identity genes[34]. SOX3 is a master TF in NPCs[35]. We identified 254 SEs in f/f NPCs using stitched SOX3 peaks[36] ranked by H3K27ac signal intensity. 97.6% (248) of them were MLL4$^{+}$ and associated with NPC identity genes such as *Nes*, *Notch1* and *Ptn* (Figure 7d). GO analysis of SE-associated genes identified nervous system development and neuron differentiation as top terms (Figure 7-S1g). While induction of H3K4me1 signals on NPC SEs displayed severe impairment in KI NPCs, signal intensities of MLL4 and H3K27ac did not show significant differences between f/f and KI NPCs. Expression levels of SE-associated genes were also induced comparably (Figure 7e-g; Figure 7-S1h). Together, these data indicate that MLL3/4-catalyzed H3K4me1 is dispensable for SE activation during neural differentiation.

## Discussion

Using the CRISPR/Cas9 system to knockin enzyme-dead point mutations into MLL3/4 in mice or ESCs, we have examined roles of MLL3/4 enzymatic activities in embryonic development and ESC differentiation, as well as MLL3/4-catalyzed H3K4me1 in enhancer activation. We show that MLL3/4 carry both enzymatic activity-dependent and -independent functions during mouse development, and that simultaneous elimination of MLL3/4 enzymatic activities leads to embryonic lethality around E8.5. In *Mll3* KO ESCs, inactivating MLL4 enzymatic activity has little effects on the maintenance of cell identity but leads to faster proliferation and monolayer cell growth. During EB differentiation, ESCs lacking MLL3/4 enzymatic activities maintain the capacity of differentiating towards the three germ layers but show defective EB cavitation, which is due to, at least in part, the impaired induction of visceral endoderm. Finally, our data reveal that MLL3/4-catalyzed H3K4me1 is dispensable for enhancer activation during early EB differentiation and super-enhancer activation in lineage-specific neural differentiation (Figure 7h).

In mice, all six Set1-like H3K4 methyltransferases are essential for normal development[37]. MLL3 is required for lung maturation and *Mll3*$^{-/-}$ embryos exhibit perinatal lethality. *Mll4*$^{-/-}$ embryos die around E9.5 and display a defect in anterior VE migration precedes gastrulation[4,7]. The phenotypes of *Mll3*$^{KI/KI}$ and *Mll4*$^{KI/KI}$ mice are much alleviated. *Mll3*$^{KI/KI}$ mice survive to adulthood with a lower proportion and *Mll4*$^{KI/KI}$ mice die around birth without obvious phenotypes, indicating that MLL3/4 proteins play a more dominant role than their enzymatic activities during mouse development (Figure 1h). Given the partial redundancy between MLL3 and MLL4 during adipogenesis[4,10], we generated homozygous *Mll3*/*Mll4* double KI embryos. Strikingly, *Mll3*$^{KI/KI}$;*Mll4*$^{KI/KI}$ embryos display severe developmental retardation at E8.0 and die around E8.5, indicating enzymatic activities of MLL3 and MLL4 are partially redundant and together indispensable for early embryonic development. This is the first *in vivo* evidence of functional redundancy between enzymatic activities of Set1-like methyltransferases. Interestingly, a previous study reported that *Drosophila* embryos expressing catalytically deficient Trr, the only homolog of MLL3 and MLL4 in *Drosophila*, produce fertile adults with only minor abnormalities[38]. Such a distinct dependence of embryonic development in mice and in *Drosophila* suggests a more critical role of MLL3/4 enzymatic activities in mammals.

The cavitation of EBs is a complex process regulated by VE as well as factors involved in autophagy and apoptosis[30,39-42]. During EB differentiation of ESCs lacking MLL3/4 enzymatic activities, VE induction

is impaired while autophagy or apoptosis factors known to regulate cavitation show normal levels of expression (Figure 4-S1e), which indicate that the EB cavitation defects after inactivating MLL3/4 enzymatic activities is attributed, at least in part, to impaired induction of VE. During EB differentiation, the damaged VE induction likely contributes to the defective cardiomyogenesis of MLL3/4 enzyme-dead ESCs as well, because VE-secreted factors promote cardiomyogenesis of ESCs[43]. In addition, the impairment of VE induction may also lead to the early lethality of *Mll3*$^{KI/KI}$;*Mll4*$^{KI/KI}$ embryos. This possibility is supported by at least two observations. First, E8.0 *Mll3*$^{KI/KI}$;*Mll4*$^{KI/KI}$ embryos morphologically resemble normal embryos at E5.5-6.5, indicating development of *Mll3*$^{KI/KI}$;*Mll4*$^{KI/KI}$ embryos is blocked at the stage when anterior VE migration takes place to establish early embryonic asymmetry[17]. Second, MLL3/4 enzyme-dead ESCs are capable of decommissioning ESC identity genes and differentiating towards all three germ layers during EB differentiation and teratoma formation, suggesting that eliminating MLL3/4 enzymatic activities may affect primarily the development of extraembryonic tissues, such as VE. Interestingly, the defect in anterior VE migration has been reported in *Mll4* null embryos[7].

In undifferentiated ESCs, MLL3/4 proteins but not MLL3/4-catalyzed H3K4me1 are essential for maintaining AEs and target gene transcription[9]. Using *Mll3*/*Mll4* double KO ESCs, our previous study showed that MLL3/4-regulated p300 recruitment and enhancer activation during EB differentiation even occur on H3K4me1-primed enhancers[5]. In this study, we generated *Mll3*$^{-/-}$;*Mll4*$^{KI/KI}$ ESCs and, for the first time, directly addressed the question whether MLL3/4-mediated enhancer H3K4me1 is required for enhancer function during cell fate transition. Using two model systems, we found that MLL3/4-catalyzed H3K4me1 is dispensable either for enhancer activation during EB differentiation or for SE activation during lineage-specific neural differentiation. Consistently, MLL3/4 regulate developmental gene induction largely independent of enhancer H3K4me1. Interestingly, during EB differentiation, enzyme-dead point mutation results in significant redistribution of de novo MLL4$^{+}$ AEs without discernible changes of MLL4 total binding intensity, which partially explains the transcriptional alteration and phenotypic changes after eliminating MLL3/4 enzymatic activities. These observations suggest that MLL3/4 enzymatic activities regulate ESC differentiation through mechanisms other than MLL3/4-catalyzed H3K4me1 on enhancers. One possibility is that MLL3/4 regulate ESC differentiation through non-histone substrates, such as lineage-determining or signal activated TFs. Indeed, a recent publication reported that P53 is a potential substrate of Set1-like

methyltransferases including MLL3/4[44]. MLL3/4-catalyzed H3K4me1 has been reported to be required for enhancer-promoter interaction and gene expression at the *Sox2* locus in ESCs, due to H3K4me1-dependent recruitment of Cohesin complex to enhancers[23]. While our results cast doubt on the role of H3K4me1-mediated enhancer activation during ESC differentiation in general, it could well be that H3K4me1-dependent recruitment functions at specific loci such as *Sox2*.

## Materials and Methods

### Generation of *Mll3*/*Mll4* knockin mice

The MLL3 Y4792A and MLL4 Y5477A knockin mouse lines were generated using CRISPR/Cas9 system modified from Wang et al.[45]. Briefly, sgRNA sequences cutting near genomic loci of MLL3 Y4792 (5'-ACTATGGTCATCGAGTACAT-3') and MLL4 Y5477 (5'-ACGATGGTCATCGAGTACAT-3') were synthesized by Thermo Fisher's custom *in vitro* transcription service. The Tyrosine to Alanine point mutations were introduced using single-strand Ultramer DNA Oligonucleotide (IDT). Resulting donor oligonucleotides are as follows: MLL3 Y4792A, 5'-TTCCTTCCACTTAGGGACTGGGCCTGTATGCTGCTAGAGACATTGAAAAACACACTATGGTCATCGAG***GCT***ATTGGAACAATTATTCGAAATGAGGTTGCAAACCGGAAGGAGAAGCTTTATGAGTCTCAGGTACTGTAT-3'; MLL4Y5477A, 5'-CGCGTATCCAGGGCCTCGGCCTCTATGCAGCCAAGGACCTGGAGAAGCACACGATGGTCATCGAG***GCT***ATCGGCACCATCATTCGCAATGAGGTGGCCAATCGGCGGGAGAAAATCTATGAGGAGCAGGTACTGTGGG-3', in which the inserted mutations are highlighted in italic and bold. To generate each knockin mouse line, sgRNA sequences (20 ng/μL) and its corresponding donor oligonucleotide (100 ng/μL) were co-microinjected with Cas9 nickase mRNA (50 ng/μL, Trilink Biotechnologies) into the cytoplasm of zygotes collected from B6D2F1/J mice (JAX #100006). Injected embryos were cultured in M16 medium (MilliporeSigma) overnight in a 37°C incubator with 6% $CO_2$. The next morning, 2-cell-stage embryos were implanted into the oviducts of pseudopregnant surrogate mothers. Offspring were genotyped by PCR and confirmed by Sanger Sequencing. All mouse experiments were performed in accordance with the NIH Guide for the Care and Use of Laboratory Animals and approved by the Animal Care and Use Committee of NIDDK, NIH.

### Generation of *Mll4* knockin ES cells

The MLL4 Y5477A knockin ES cell lines were generated using the CRISPR/Cas9 system modified from Ran et al.[46]. Briefly, a sgRNA sequence cutting near MLL4 Y5477 (5'-ACGATGGTCATCGAGTACAT-3') was cloned into lentiCRISPR v2 plasmid (Addgene #52961). The Tyrosine to Alanine point mutation was introduced using single-strand Ultramer DNA Oligonucleotide (IDT). The resulting donor oligonucleotide is

as follows: 5'-TCGGCCTCTATGCAGCCAAGGACCTGGAGAAGCACACGATGGTCATCGAG***GCT***ATCGGCACCATCATTCGCAATGAGGTGGCCAATCGGCGGGAGAAAATCTA-3', in which the inserted mutations are highlighted in italic and bold. To generate knockin ES cells, the plasmid containing sgRNA (10 µg) and the donor oligonucleotide (400 pmol) were co-transfected into 1 x $10^6$ *Mll3*$^{-/-}$;*Mll4*$^{f/f}$ mouse ES cells by Lipofectamine 3000 (Thermofisher). 24 hours later, cells were selected with 1.5 µg/mL puromycin for 3 days. After recovery, 1 x $10^4$ single cells were seeded into a 6 cm dish. Colonies derived from single cells were picked 1 week later and expanded for genotyping. ~1.1 kb PCR products covering the targeted region were subsequently screened by Sanger sequencing to map the expected mutation. Several potential exonic off-target regions were also checked and no indel was detected. Finally, the stability of the mutated protein was confirmed by immunoblotting.

**ES cell culture**

Mouse ES cells were cultured as described in Koehler and Hashino[47]. Briefly, cells were maintained on dishes coated with 0.1% gelatin (Millipore) in ESC culture medium, a serum-free 2i+LIF medium containing 47.75% Advanced DMEM/F-12 (Gibco), 47.75% Neurobasal (Gibco), 0.5% N-2 supplement (Gibco), 1% B-27 supplement minus Vitamin A (Gibco), 1% GlutaMax (Gibco), 1% Penicillin-Streptomycin (Corning), 1% EmbryoMax 2-mercaptoethanol (MilliporeSigma), and supplemented with 1 µM MEK inhibitor PD0325901 (Selleckchem), 3 µM GSK3 inhibitor CHIR99021 HCl (Selleckchem) and 1,000 units/mL mouse LIF protein (MilliporeSigma).

For Alkaline Phosphatase (AP) staining, 4 x $10^3$ single cells were seeded into one well of a 24-well plate. 4 days later, AP staining was done using Alkaline Phosphatase Detection Kit (MilliporeSigma) following the manufacturer's instructions.

**ES cell differentiation**

Embryoid body (EB) differentiation was performed according to the hanging drop method modified from Cao et al.[48] (Figure 3-S1a). Briefly, cells were suspended into 1 × $10^5$ /mL in indicated differentiation medium and aliquoted into 20 µL drops on the lid of tissue culture dishes. EBs were harvested 2 days later,

transferred into ultra-low-attached dishes (Corning) and further cultured in suspension to the indicated time points. The medium used in EB differentiation from day 0 to day 2 is modified from ESC culture medium. It contains no LIF, 1/4 concentration of both inhibitors and is supplemented with 2% ESC-qualified FBS (Gibco). After day 2, the two inhibitors were removed, and medium was changed every other day. The diameter of EB was measured using ImageJ software (NIH).

To monitor the cardiomyogenesis during EB differentiation, day 2 EBs harvested from hanging drops were plated into gelatinized 24-well plates with one EB per well. Spontaneous contracting clusters could be observed as early as day 6. The percentage of beating EBs in each 24-well plate was calculated at indicated time points.

Neural differentiation protocol was modified from Bibel et al.[22] (Figure 7-S1a). Briefly, neural lineage aggregates were generated following EB differentiation method in ultra-low-attached dishes described above from day 0 to day 8 except that after day 4, 5 µM retinoic acid (MilliporeSigma) was added to the culture. At day 8, aggregates were dissociated with trypsin and neural progenitor cells (NPCs) were plated on PDL/laminin-coated dishes in N-2 medium (Advanced DMEM/F-12 supplemented with 1% N-2 supplement, 1% 2 mM GlutaMax, 1% 2-mercaptoethanol and 50 µg/mL BSA). The N-2 medium was changed after 2 hours and again after 1 day. After 2 days, the medium was replaced by the B-27 medium (Neurobasal medium supplemented with 2% B-27 serum free (Gibco), 1% 2 mM GlutaMax and 1% 2-mercaptoethanol) to induce neurogenesis. The B-27 medium was changed every other day. Neural differentiation was performed until day 16.

**Teratoma formation assay**

Approximately 2 x $10^5$ ES cells were mixed with 50% volume of Matrigel Matrix (Corning) and injected subcutaneously into two sides of an immunocompromised mouse (NSG mouse, JAX #005557). 25 days later, the teratomas were dissected out, fixed in 10% neutral buffered formalin and subjected to histological analysis with H&E staining following standard protocols.

**Immunoblotting**

Nuclear proteins were extracted as described previously[10]. Briefly, cells were collected and resuspended in cell lysis buffer (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl and 0.1% NP-40) supplemented with protease inhibitors (Roche) for 20 min to swell cellular membrane. After centrifugation at 2,000x*g*, nuclei were extracted in nuclear lysis buffer (20 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 420 mM NaCl, 0.2 mM EDTA and 25% glycerol) supplemented with protease inhibitors. Nuclear extracts were separated using 3–8% Tris-Acetate gradient gel (Invitrogen).

Histone extraction protocol was modified from Shechter et al.[49]. Briefly, cells were collected and resuspended in hypotonic lysis buffer (10 mM Tris-HCl pH 8.0, 1 mM KCl, 1.5 mM $MgCl_2$ and 1 mM DTT) supplemented with protease inhibitors for 10 min to swell cellular membrane. After centrifugation at 10,000x*g*, histone fraction was extracted with 0.4 N HCl and neutralized to pH 7.0. Histone extracts were separated using home-made 12% SDS-PAGE gel.

Primary antibodies used in immunoblotting were indicated as follows. Anti-MLL3#3[8], anti-MLL4#3[8], and anti-UTX#2[50] were home-made. Anti-CBP (A700-010) and anti-RBBP5 (A300-109A) were from Bethyl Laboratories. Anti-ARID1A (sc-32761X) and anti-BRG1 (sc-10768 X) were from Santa Cruz Biotechnology. Anti-CBP (7389S) was from Cell Signaling Technology. Anti-H3K4me1 (13-0040) was from EpiCypher. Anti-H3K4me2 (ab7766), anti-H3 (ab1791), and anti-H3K27ac (ab4729) were from Abcam. Anti-H3K4me3 (17-614) was from MilliporeSigma.

**RNA isolation and RT-qPCR analysis**

Total RNA was extracted by TRIzol (Life Technologies) and reverse transcribed using ProtoScript II first-strand cDNA synthesis kit (NEB) following the manufacturer's instructions. Quantitative PCR was performed in duplicate with Luna Universal qPCR Master Mix (NEB) using QuantStudio 5 Real-Time PCR System (Thermo Fisher). RT-qPCR data were normalized using *Gapdh* and were presented as means ± SD. Primer sequences used in RT-qPCR are listed below.

*Mll4*: forward 5'-GCTATCACCCGTACTGTGTCAACA-3', reverse 5'-CACACACGATACACTCCACACAA-3'

*Nanog*: forward 5'-CACCCACCCATGCTAGTCTT-3', reverse 5'-ACCCTCAAACTCCTGGTCCT-3'

*Oct4*: forward 5'-CGGAAGAGAAAGCGAACTAGC-3', reverse 5'-ATTGGCGATGTGAGTGATCTG-3'

*Sox2*: forward 5'-GGAAAGGGTTCTTGCTGGGT-3', reverse 5'-ACGAAAACGGTCTTGCCAGT-3'

*Rex1*: forward 5'-GTTCGTCCATCTAAAAAGGG-3', reverse 5'-TAGTCCATTTCTCTAATGCCC-3'

*Esrrb*: forward 5'-AAAGCCATTGACTAAGATCG-3', reverse 5'-AATTCACAGAGAGTGGTCAG-3'

*Gata4*: forward 5'-CACCCCAATCTCGATATGTTTGA-3', reverse 5'-GGTTGATGCCGTTCATCTTGT-3'

*Afp*: forward 5'-TTTCCAGAACCTGCCGAGAG-3', reverse 5'-GACAGAATGGCTGGGGCATA-3'

*T*: forward 5'-GCTTCAAGGAGCTAACTAACGAG-3', reverse 5'-CCAGCAAGAAAGAGTACATGGC-3'

*Kdr*: forward 5'-CAAACCTCAATGTGTCTCTTTGC-3', reverse 5'-AGAGTAAAGCCTATCTCGCTGT-3'

*Nes*: forward 5'-CTCTTGGCTTTCCTGACCCC-3', reverse 5'-AGGCTGTCACAGGAGTCTCA-3'

*Cdh2*: forward 5'-CAGGGTGGACGTCATTGTAG-3', reverse 5'-AGGGTCTCCACCACTGATTC-3'

*Gapdh*: forward 5'-AATGTGTCCGTCGTGGATCTGA-3', reverse 5'-GATGCCTGCTTCACCACCTTCT-3'

**RNA-seq library preparation**

1 µg total RNA was used for mRNA purification with NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, E7490). From purified mRNA, libraries were constructed using NEBNext Ultra™ II RNA Library Prep Kit for Illumina (NEB, E7770) following the manufacturer's instructions, and were sequenced on Illumina HiSeq 4000 and NovaSeq 6000.

**ChIP and ChIP-seq library preparation**

Chromatin immunoprecipitation (ChIP) and ChIP-Seq were performed as described previously[5]. Briefly, cells were cross-linked with 1% formaldehyde for 10 min and quenched by 125 mM glycine for 10 min. Fixed cells were swelled in the lysis buffer containing 5mM PIPES pH 7.5, 85mM KCl, 0.5% NP-40 and protease inhibitors, incubated on ice for 20 min, and centrifuged at 500x*g* for 5 min at 4°C. Nuclei were resuspended with cold TE buffer (10mM Tris-HCl pH8.0, 1mM EDTA), and subjected to sonication. Sheared chromatin was clarified by centrifugation at 13,000x*g* for 10 min at 4°C. The supernatant was transferred to a new tube and further supplemented with 150mM NaCl, 1% Triton X-100, 0.1% sodium deoxycholate and protease inhibitors. 2% of the mixture was set aside as input. For each ChIP, chromatin from 1 x $10^7$ cells were mixed with 20 ng spike-in chromatin (Active Motif, #53083) and incubated with 4-8 µg primary target antibody and 2 µg spike-in antibody (Active Motif, #61686) overnight at 4°C. Then, ChIP samples were precipitated by 50 µL prewashed Protein A Dynabeads (ThermoFisher). Beads were washed twice

with 1 mL RIPA buffer, once with 1 mL RIPA buffer containing 500mM NaCl, twice with 1 mL LiCl buffer and once with 1 mL TE buffer, and then eluted with 100 μL fresh elution buffer (0.1M $NaHCO_3$, 1% SDS). ChIP samples and input were incubated with Proteinase K (NEB) at 65°C overnight to reverse formaldehyde crosslinking, and then purified using QIAquick PCR purification kit (Qiagen).

For ChIP-Seq, the entire ChIP DNA and 300 ng input DNA were used to construct libraries using NEBNext Ultra™ II DNA Library Prep kit for Illumina (NEB, E7645) following the manufacturer's instructions. The final libraries were sequenced on Illumina HiSeq 3000 and HiSeq 4000.

**ATAC-seq library preparation**

Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-Seq) was performed as described[51]. Briefly, for each reaction, $5 \times 10^4$ cells were freshly collected, washed and swelled in 50 μL cold lysis buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% NP-40, 0.1% Tween-20, and 0.01% digitonin) for 3 min. Nuclei were collected by centrifugation at 500×*g* for 10 min at 4°C, and then resuspended in 50 μL of transposition reaction buffer containing 25 μL 2x Tagment DNA buffer, 2.5 μL transposase (100 nM final, Illumina), 0.5 μL 1% digitonin, 0.5 μL 10% Tween-20, 16.5 μL PBS, and 5 μL $H_2O$. The reaction was mixed for 30 min at 37°C and subjected to purification using the MinElute Reaction Cleanup Kit (Qiagen) according to the manufacturer's instructions. Purified DNA was amplified for library construction with PCR using Nextera i5 common adaptor and i7 index adaptors (Illumina). The final libraries were sequenced on Illumina HiSeq 4000.

**Computational analysis**

NGS data processing

For RNA-seq, raw sequencing data were aligned to the mouse genome mm9 using STAR software[52] (v2.7.5). Reads on exons defined by UCSC gene annotation system were collected to calculate Transcripts Per Million (TPM) as a measure of gene expression level, and genes with TPM > 3 were regarded as expressed.

For ChIP-seq, raw sequencing data were aligned to the mouse genome mm9 and the drosophila genome dm6 using Bowtie2[53] (v2.3.2). To identify ChIP-enriched regions, SICER[54] (v1.1) was used. For

ChIP-Seq of MLL4, the window size of 50 bp, the gap size of 50 bp, and the false discovery rate (FDR) threshold of $10^{-10}$ were used. For ChIP-Seq of histone modifications (H3K4me1 and H3K27ac), the window size of 200 bp, the gap size of 200 bp, and the FDR threshold of $10^{-3}$ were used. Reads on indicated regions were collected to calculate Reads Per Kilobase Million (RPKM) as a measure of signal intensity.

For ATAC-seq, raw ATAC-Seq reads were processed using Kudaje lab's ataqc pipeline (https://github.com/kundajelab/atac_dnase_pipelines) (v0.3.4) which includes adapter trimming, aligning to mouse mm9 genome by Bowtie2, and peak calling by MACS2[55]. For downstream analysis, filtered reads that were retained after removing unmapped reads, duplicates and mitochondrial reads were used.

RNA-seq analysis

Differentially expressed genes were determined using DESeq2[56] (v1.20.0) in R/Bioconductor with a cutoff of 2.5-fold ($p < 0.05$). PCA graph and scatter plots were drawn using ggplot2[57] (v3.2.1) in R. Gene ontology (GO) analysis was done using DAVID[58] with the whole mouse genome as background.

For comparison of gene expression levels during EB differentiation (Figure 4-S1b,e; Figure 5c), duplicates of KI #27 and KI #48 were combined as KI quadruplicates to compare with f/f duplicates. For comparison of gene expression levels during neural differentiation (Figure 7-S1c-d), single data of KI #27 and KI #48 were combined as KI duplicates to compare with f/f duplicates.

For clustering analysis, we collected genes expressed in either f/f, KI #27 or KI #48 at any of the three time points during EB differentiation. Genes differentially expressed between any two time points in either cell line were clustered into 12 groups (Cluster I-XII) using the K-means clustering (ward.D2). Heat map (Figure 4c) of clustered gene expression in terms of z-score was drawn using pheatmap[59] (v1.0.12) in R. Line chart (Figure 4d; Figure 4-S1a) was drawn from the average expression of genes in each cluster in terms of z-score.

Gene set enrichment analysis (GSEA) was performed using GSEA software[60] (v4.0.3), and cell-type markers in E6.5 embryos with $p < 0.05$[27] were used as signature databases.

Genomic distribution of ChIP-seq peaks

To define regulatory regions, a combination of genomic coordinates and histone modification ChIP-Seq data were used. Promoter regions were defined as transcription start sites ± 1 kb. Promoter-distal regions were further overlapped with H3K27ac$^+$ regions to define active enhancers (AEs). MLL4$^+$ AEs were defined as AEs that are overlapping with MLL4 peaks at least 1 bp. All MLL4$^+$ AEs existed in D4 EBs but not in ESCs were regarded as de novo MLL4$^+$ AEs. We associated AEs to the proximal genes within ±100 kb.

To compare MLL4 genomic distributions between f/f and KI D4 EBs, the union of de novo MLL4$^+$ AEs was divided into three groups based on MLL4 binding intensities in KI cells compared to f/f cells with a cutoff of 2-fold. Heat map matrices were generated using in-house scripts with 50 bp resolution and visualized in R. Average profiles were plotted using the number of ChIP-seq reads (normalized to the size of each library) in 5 bp bins from the center of each regulatory element to 2 kb on both sides.

Motif analysis

To find enriched TF motifs in given genomic regions identified by ChIP-Seq, we utilized the SeqPos motif tool in the Cistrome toolbox[61]. For ChIP-Seq data with more than 5,000 regions, top 5,000 significant regions were used. For ChIP-Seq data with small numbers, all regions were used.

Analysis of super-enhancer

We used Rank Ordering of Super-Enhancers[62] (ROSE) with default parameters to identify super-enhancers (SEs). To identify SEs in NPCs, we stitched SOX3 binding sites (GSE33059)[36] in NPC AEs and used H3K27ac signal intensity for ranking. We associated SEs to the proximal genes within ±100 kb.

Box plots and genomic profile visualization

Box plots were drawn with GraphPad Prism 8 software using the Tukey method. In Figure 7e-f, ChIP-seq signal intensities of MLL4, H3K4me1 and H3K27ac on NPC SEs, and expression levels of NPC SE-associated genes in f/f and KI cells were plotted. Wilcoxon signed-rank test (two-sided) was used to determine statistical differences in signal enrichment and gene expression levels between f/f and KI NPCs.

To visualize data from ChIP-seq, ATAC-seq and RNA-seq in the Integrative Genomics Viewer[63] (IGV) (v2.4.4), reads were collected and converted to wiggle (wig) format-profiles using in-house script.

**Acknowledgement**

We thank Nhien Tran and Ruth Kopyto for assistance in genotyping, Hualong Yan and Todd S. Macfarlan for suggestions on ESC culture and gene editing, NHLBI DNA Sequencing and Genomics Core and UCSD IGM Genomics Center for next-generation sequencing, NIH HPC group for high-performance computing. This work was supported by the Intramural Research Program of NIDDK, NIH to K.G.

**Author Contributions**

Conceptualization, G.X. and K.G.; Methodology, G.X., J.-E.L., W.P. and C.L.; Investigation, G.X., Y.-K.P. and Y.J.; Software, Formal Analysis, and Data Curation, J.-E.L., G.X. and W.P.; Writing – Original Draft, G.X. and K.M.; Writing – Review & Editing, G.X., J.-E.L., W.P. and K.G.; Project Administration and Funding Acquisition, K.G.

**Declaration of Interests**

The authors declare no competing interests.

**Data availability**

All data sets described in the paper have been deposited in NCBI Gene Expression Omnibus under accession number GSE154475.

**Code availability**

In-house generated computational codes are available upon request.

## References


1. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol Cell* **49**, 825-37 (2013).
2. Creyghton, M.P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-6 (2010).
3. Jin, Q. *et al.* Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *EMBO J* **30**, 249-62 (2011).
4. Lee, J.E. *et al.* H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *Elife* **2**, e01503 (2013).
5. Wang, C. *et al.* Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. *Proc Natl Acad Sci U S A* **113**, 11871-11876 (2016).
6. Lai, B. *et al.* MLL3/MLL4 are required for CBP/p300 binding on enhancers and super-enhancer formation in brown adipogenesis. *Nucleic Acids Res* **45**, 6388-6403 (2017).
7. Ashokkumar, D. *et al.* MLL4 is required after implantation whereas MLL3 becomes essential during late gestation. *Development* (2020).
8. Cho, Y.W. *et al.* PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *J Biol Chem* **282**, 20395-406 (2007).
9. Dorighi, K.M. *et al.* Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Mol Cell* **66**, 568-576 e4 (2017).
10. Jang, Y. *et al.* H3.3K4M destabilizes enhancer H3K4 methyltransferases MLL3/MLL4 and impairs adipose tissue development. *Nucleic Acids Res* **47**, 607-620 (2019).
11. Zhang, J. *et al.* Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. *Nat Med* **21**, 1190-8 (2015).
12. Ang, S.Y. *et al.* KMT2D regulates specific programs in heart development via histone H3 lysine 4 di-methylation. *Development* **143**, 810-21 (2016).
13. Zhang, Z. *et al.* Mammary-Stem-Cell-Based Somatic Mouse Models Reveal Breast Cancer Drivers Causing Cell Fate Dysregulation. *Cell Rep* **16**, 3146-3156 (2016).
14. Placek, K. *et al.* MLL4 prepares the enhancer landscape for Foxp3 induction via chromatin looping. *Nat Immunol* **18**, 1035-1045 (2017).
15. Froimchuk, E., Jang, Y. & Ge, K. Histone H3 lysine 4 methyltransferase KMT2D. *Gene* **627**, 337-342 (2017).
16. Tam, P.P. & Loebel, D.A. Gene function in mouse embryogenesis: get set for gastrulation. *Nat Rev Genet* **8**, 368-81 (2007).
17. Rossant, J. & Tam, P.P. Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. *Development* **136**, 701-13 (2009).
18. Stern, C.D. & Downs, K.M. The hypoblast (visceral endoderm): an evo-devo perspective. *Development* **139**, 1059-69 (2012).
19. Stuckey, D.W., Di Gregorio, A., Clements, M. & Rodriguez, T.A. Correct patterning of the primitive streak requires the anterior visceral endoderm. *PLoS One* **6**, e17620 (2011).
20. Keller, G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev* **19**, 1129-55 (2005).
21. Leahy, A., Xiong, J.W., Kuhnert, F. & Stuhlmann, H. Use of developmental marker genes to define temporal and spatial patterns of differentiation during embryoid body formation. *J Exp Zool* **284**, 67-81 (1999).
22. Bibel, M., Richter, J., Lacroix, E. & Barde, Y.A. Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nat Protoc* **2**, 1034-43 (2007).
23. Yan, J. *et al.* Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers. *Cell Res* **28**, 387 (2018).

24. Kato, H. *et al.* Cancer-derived UTX TPR mutations G137V and D336G impair interaction with MLL3/4 complexes and affect UTX subcellular localization. *Oncogene* (2020).
25. Hirst, C.E. *et al.* Transcriptional profiling of mouse and human ES cells identifies SLAIN1, a novel stem cell gene. *Dev Biol* **293**, 90-103 (2006).
26. Gloss, B.S. *et al.* High resolution temporal transcriptomics of mouse embryoid body development reveals complex expression dynamics of coding and noncoding loci. *Sci Rep* **7**, 6731 (2017).
27. Pijuan-Sala, B. *et al.* A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490-495 (2019).
28. Paca, A. *et al.* BMP signaling induces visceral endoderm differentiation of XEN cells and parietal endoderm. *Dev Biol* **361**, 90-102 (2012).
29. Coucouvanis, E. & Martin, G.R. BMP signaling plays a role in visceral endoderm differentiation and cavitation in the early mouse embryo. *Development* **126**, 535-46 (1999).
30. Coucouvanis, E. & Martin, G.R. Signals for death and survival: a two-step mechanism for cavitation in the vertebrate embryo. *Cell* **83**, 279-87 (1995).
31. Narita, N., Bielinska, M. & Wilson, D.B. Wild-type endoderm abrogates the ventral developmental defects associated with GATA-4 deficiency in the mouse. *Dev Biol* **189**, 270-4 (1997).
32. Morrisey, E.E. *et al.* GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev* **12**, 3579-90 (1998).
33. Wamaitha, S.E. *et al.* Gata6 potently initiates reprograming of pluripotent and differentiated cells to extraembryonic endoderm stem cells. *Genes Dev* **29**, 1239-55 (2015).
34. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-47 (2013).
35. Bylund, M., Andersson, E., Novitch, B.G. & Muhr, J. Vertebrate neurogenesis is counteracted by Sox1-3 activity. *Nat Neurosci* **6**, 1162-8 (2003).
36. Bergsland, M. *et al.* Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev* **25**, 2453-64 (2011).
37. Crump, N.T. & Milne, T.A. Why are so many MLL lysine methyltransferases required for normal mammalian development? *Cell Mol Life Sci* **76**, 2885-2898 (2019).
38. Rickels, R. *et al.* Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. *Nat Genet* **49**, 1647-1653 (2017).
39. Qu, X. *et al.* Autophagy gene-dependent clearance of apoptotic cells during embryonic development. *Cell* **128**, 931-46 (2007).
40. He, X. *et al.* Rac1 is essential for basement membrane-dependent epiblast survival. *Mol Cell Biol* **30**, 3569-81 (2010).
41. Qi, Y. *et al.* Bnip3 and AIF cooperate to induce apoptosis and cavitation during epithelial morphogenesis. *J Cell Biol* **198**, 103-14 (2012).
42. Qi, Y. *et al.* PTEN induces apoptosis and cavitation via HIF-2-dependent Bnip3 upregulation during epithelial lumen formation. *Cell Death Differ* **22**, 875-84 (2015).
43. Wang, J. *et al.* MicroRNA-125b/Lin28 pathway contributes to the mesendodermal fate decision of embryonic stem cells. *Stem Cells Dev* **21**, 1524-37 (2012).
44. Li, Y. *et al.* Crystal Structure of MLL2 Complex Guides the Identification of a Methylation Site on P53 Catalyzed by KMT2 Family Methyltransferases. *Structure* (2020).
45. Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910-8 (2013).
46. Ran, F.A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308 (2013).
47. Koehler, K.R. & Hashino, E. 3D mouse embryonic stem cell culture for generating inner ear organoids. *Nat Protoc* **9**, 1229-44 (2014).

48. Cao, N. *et al.* In vitro differentiation of rat embryonic stem cells into functional cardiomyocytes. *Cell Res* **21**, 1316-31 (2011).
49. Shechter, D., Dormann, H.L., Allis, C.D. & Hake, S.B. Extraction, purification and analysis of histones. *Nat Protoc* **2**, 1445-57 (2007).
50. Hong, S. *et al.* Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. *Proc Natl Acad Sci U S A* **104**, 18439-44 (2007).
51. Corces, M.R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959-962 (2017).
52. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
53. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-9 (2012).
54. Zang, C. *et al.* A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952-8 (2009).
55. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
56. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
57. Wickham, H., Sievert, C. & Springer International, P. *ggplot2 : elegant graphics for data analysis*, (2016).
58. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
59. Kolde, R. Pheatmap: pretty heatmaps. *R package version* **1**(2012).
60. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).
61. He, H.H. *et al.* Nucleosome dynamics define transcriptional enhancers. *Nat Genet* **42**, 343-7 (2010).
62. Whyte, W.A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-19 (2013).
63. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-6 (2011).

## Figure legends

**Figure 1. MLL3/4 enzymatic activities are essential for early embryonic development in mice**

(**a**) Schematic diagram of mouse MLL3 and MLL4 proteins. To eliminate the methyltransferase activities, tyrosine (Y) residues (highlighted in red) are mutated to alanine (A). PHD: Plant Homeotic Domain, HMG: High Mobility Group, FYRC/N: FY-Rich C/N-terminal, SET: Su(var)3-9, Enhancer-of-zeste and Trithorax.

(**b**) Diagram of generating enzyme-dead MLL3/MLL4 knockin (KI) mice by Cas9 nickase. sgRNA: single guide RNA, nCas9: Cas9 nickase, ssDNA: single-strand DNA.

(**c**) Genotype of progeny from crossing between *Mll3*$^{KI/+}$ mice at postnatal day 21 (P21).

(**d**) *Mll4*$^{KI/KI}$ mice died perinatally. Genotypes of progeny from crossing between *Mll4*$^{KI/+}$ mice at embryonic day 18.5 (E18.5), P0.5 and P21 are shown. The numbers of dead pups are indicated in parentheses.

(**e**) Genotype of progeny from crossing between *Mll3*$^{KI/+}$;*Mll4*$^{KI/+}$ mice at E14.5.

(**f**) *Mll3*$^{KI/KI}$;*Mll4*$^{KI/KI}$ mice died around E8.5. Genotypes of progeny from crossing between *Mll3*$^{KI/KI}$;*Mll4*$^{KI/+}$ mice at E8.0, E8.5, E10.5 and E14.5 are shown. The numbers of morphologically abnormal embryos are indicated in parentheses.

(**g**) Representative images of E8.0 and E8.5 embryos with indicated genotypes.

(**h**) Summary of phenotypes of *Mll3*$^{-/-}$, *Mll3*$^{KI/KI}$, *Mll4*$^{-/-}$, *Mll4*$^{KI/KI}$, and *Mll3*$^{KI/KI}$;*Mll4*$^{KI/KI}$ mice. Phenotypes of *Mll3*$^{-/-}$ and *Mll4*$^{-/-}$ mice have previously been reported[4].

**Figure 1-S1. Alignment of SET domains in different histone methyltransferases**

The highly conserved tyrosines (Y), Y4792 in MLL3 and Y5477 in MLL4, are mutated to alanines (A) in mice and/or ESCs in this study.

**Figure 2. ESCs lacking MLL3/4 enzymatic activities maintain cell identity**

(**a-b**) Enzyme-dead Y5477A mutation reduces global levels of H3K4me1 and H3K4me2 without affecting MLL4 protein stability. MLL3/4 enzyme-dead (*Mll3*$^{-/-}$;*Mll4*$^{KI/KI}$, KI) ESCs were generated as described in Figure 2-S1. Nuclear extracts (**a**) or histone extracts (**b**) prepared from wild-type (WT), *Mll3*$^{-/-}$;*Mll4*$^{f/f}$ (f/f), *Mll3*/*Mll4* double knockout (KO) and KI ESCs were analyzed with immunoblotting using indicated antibodies. The asterisk indicates a non-specific band.

(**c-d**) KI ESCs grow in monolayer and proliferate faster than f/f ESCs. (**c**) Representative phase contrast microscopic (*upper*) and AP staining (*lower*) images. Scale bar = 100 μm. (**d**) Population doubling time. Data are presented as means ± SD ($n$ = 5).

(**e**) MLL3/4 enzymatic activities are dispensable for maintaining ESC identity. *Mll4* and ESC identity genes were analyzed by RT-qPCR ($n$ = 2).

**Figure 2-S1. Generation of ESCs lacking MLL3/4 enzymatic activities**

(**a**) Diagram of generating ESCs harboring enzyme-dead mutant (Y5477A) MLL4. Knockin was done in *Mll3*$^{-/-}$;*Mll4*$^{f/f}$ ESCs.

(**b**) Schematic representation of *Mll4* flox, knockout (KO) and knockin (KI) alleles. In flox and KI alleles, two loxP sites were inserted in the intron before exon 16 and the intron after exon 19; in the KI allele, the Y5477A mutation is located in exon 52. The locations of PCR genotyping primers P1, P2, P3, WF, MF and R are indicated by arrows. The MF primer is specific to the KI allele.

(**c**) PCR genotyping using primer pairs indicated on the right. Sizes of PCR products are indicated on the left.

(**d-h**) Chromatograms of genomic PCR sequencing from two independent KI ES cell lines. Trace files display sequences around the Y5477A mutation (**d**) or sequences around potential exonic off-target regions of *Dups4* (**e**), *Setd1b* (**f**), *Ccdc80* (**g**) and *Fignl1* (**h**). In the reference sequences, sgRNA-defined target regions and PAM sites are highlighted in yellow and cyan, respectively.

**Figure 3. ESCs lacking MLL3/4 enzymatic activities are capable of differentiating towards the three germ layers but show defects in EB cavitation**

(**a-b**) MLL3/4 enzyme-dead ESCs show defects in cavitation during EB differentiation. f/f, KO and KI ESCs were induced to differentiate into embryoid bodies (EBs). (**a**) Representative microscopic images at day 2 (D2), day 4 (D4), day 7 (D7) and day 10 (D10) of EB differentiation. Representative cystic cavities are marked with red arrows. Scale bar = 250 μm. (**b**) RT-qPCR analysis of ESC identity genes (*Nanog*, *Oct4*), endoderm markers (*Gata4*, *Afp*), mesoderm markers (*T*, *Kdr*) and ectoderm markers (*Nes*, *Cdh2*) at indicated time points.

(**c**) MLL3/4 enzyme-dead ESCs are capable of differentiating towards all three germ layers in teratoma assays. Representative histological sections of teratomas are shown. C, cartilage; G, glandular epithelium; M, muscle; N, neural epithelium. Scale bar = 200 μm.

**Figure 3-S1. ESCs lacking MLL3/4 enzymatic activities show defects in cardiomyogenesis**

(**a**) Diagram of EB differentiation.

(**b-c**) Immunoblotting of f/f and KI cells at D0 (Day 0), D4 and D10 of EB differentiation. Nuclear extracts (**b**) or histone extracts (**c**) were analyzed using indicated antibodies.

(**d**) The diameters of EBs at indicated time points are presented as floating bars. Each bar indicates min to max and the horizontal line inside represents the mean. For each time point, more than 20 EBs were measured per group.

(**e-h**) Loss of MLL3/4 enzymatic activities impairs cardiomyogenesis during EB differentiation. (**e**) The percentage of beating EBs at indicated time points. The results are from 3 independent experiments with 24 EBs per group. (**f**) Representative microscopic images of attached EBs at day 12 (D12). Scale bar = 250 μm. (**g**) The beating frequencies of EBs at D12 are presented as dot plots. Horizontal lines represent mean values. 24 EBs were measured per group. Statistical significance was determined by the two-tailed unpaired *t*-test. (**h**) RT-qPCR analysis of cardiac progenitor marker (*Tbx5*) and cardiac myofilament genes (*Myh6*, *Myl2*, *Ttn*) at indicated time points.

**Figure 4. Loss of MLL3/4 enzymatic activities impairs visceral endoderm induction in EB differentiation**

f/f and KI ESCs were induced to EB differentiation for 10 days. D4 and D10 EBs were collected for RNA-seq.

(**a**) Schematic diagram of ESC (D0) and EB (D4, D10) stages during EB differentiation. At D4, EBs are covered by the extraembryonic endoderm layer and contain differentiated cells that transiently express PS markers. From D4 to D10, cavitation takes place and the three germ layers develop.

(**b**) Principal component analysis (PCA) of RNA-seq data from cells of indicated genotypes at D0, D4 and D10 during EB differentiation.

(**c-d**) Characterization of the dynamic transcriptome. Differentially expressed genes during EB differentiation from f/f and KI cells were clustered according to their expression patterns. (**c**) Heat map of clustered gene expression. (**d**) Temporal profiles of expression levels in terms of z-score (*left*) and GO analysis (*right*) of genes in Clusters III and IX. The number of genes in each group is indicated in parenthesis. Top 5 significant GO terms are displayed. Terms associated with early embryonic development are highlighted in blue.

(**e-f**) Impaired expression of visceral endoderm (VE) but not primitive streak (PS) markers in KI D4 EBs. (**e**) Gene set enrichment analysis (GSEA) of expression profiles in D4 EBs using cell-type markers in E6.5 embryos[27]. NES, normalized enrichment score. (**f**) Expression fold changes of reported VE and PS genes between KI and f/f D4 EBs.

**Figure 4-S1. Loss of MLL3/4 enzymatic activities impairs visceral endoderm induction in EB differentiation**

(**a**) Temporal profiles of expression levels in terms of z-score (*left*) and GO analysis (*right*) of genes in indicated Clusters. The number of genes in each group is indicated in parenthesis. Top 4 significant GO terms are displayed. Terms associated with cardiomyogenesis are highlighted in blue.

(**b**) GSEA of expression profiles in D4 EBs using nascent mesoderm (NM) markers in E6.5 embryos[27]. NES, normalized enrichment score.

(**c**) Expression levels of representative VE genes. Data from RNA-seq are presented as dot plots with median lines (f/f: *n* = 2; KI: *n* = 4). Statistical significance was determined by the two-tailed unpaired *t*-test.

(**d**) Expression fold changes between KI and f/f D4 EBs for genes encoding components of BMP, Wnt and Nodal signaling pathways.

(**e**) Expression levels of genes reported to be important for EB cavitation. Data from RNA-seq are presented as dot plots with median lines (f/f: *n* = 2; KI: *n* = 4).

**Figure 5. MLL3/4 enzymatic activities are required for induction of visceral endoderm TFs GATA4/6 in the early phase of EB differentiation**

f/f and KI D4 EBs were collected for ChIP-seq.

(**a**) MLL4$^+$ AEs in f/f or KI cells at ESC and D4 EB are depicted by the Venn diagram. 10,767 de novo MLL4$^+$ AEs highlighted in red were divided into three groups based on the changes of MLL4 binding intensities from f/f to KI EBs: increased (Group I), unchanged (Group II) and decreased (Group III).

(**b**) Group III de novo MLL4$^+$ AEs are exclusively overrepresented in AEs associated with VE markers.

(**c**) Motif analysis of three groups of de novo MLL4$^+$ AEs in D4 EBs.

(**d**) Expression levels of representative transcription factors highlighted in bold in Figure 5c. Data from RNA-seq are presented as dot plots with median lines (f/f: *n* = 2; KI: *n* = 4). Statistical significance was determined by the two-tailed unpaired *t*-test.

(**e**) ChIP-seq profiles of MLL4, H3K4me1 and H3K27ac, ATAC-seq and RNA-seq profiles in f/f and KI cells are displayed on representative VE gene loci. Group III de novo MLL4$^+$ AEs are highlighted in shades. GATA6 binding sites in extraembryonic endoderm (GSE69323)[36] are presented as red arrows.

**Figure 5-S1. Genomic binding profiles of MLL4 in D4 EBs**

f/f and KI D4 EBs were collected for ChIP-seq and 34,772 MLL4$^+$ regions were identified in f/f or KI ESCs. MLL4$^+$ distal regions or promoters were splitted based on the MLL4 binding intensities in KI ESCs compared to f/f ESCs.

(**a**) Heat maps of 34,772 MLL4$^+$ regions identified by ChIP-seq in f/f or KI D4 EBs.

(**b**) Enzyme-dead mutation does not affect MLL4 total binding intensities in D4 EBs. Average profiles of MLL4 genomic binding on MLL4$^+$ regions in f/f and KI EBs are shown.

**Figure 5-S2. GATA family TFs are highly enriched in de novo MLL4$^+$ AEs associated with visceral endoderm markers**

(**a-c**) Motif analysis of de novo MLL4$^+$ AEs associated with cell-type markers in E6.5 embryos. 89, 95 and 41 de novo MLL4$^+$ AEs were annotated to markers in E6.5 embryos for VE (**a**), PS (**b**) and NM (**c**), respectively. GATA family TFs are highlighted in bold.

**Figure 6. MLL3/4-catalyzed H3K4me1 is dispensable for enhancer activation during early EB differentiation**

f/f and KI D4 EBs were collected for ChIP-seq and ATAC-seq.

(**a-b**) Heat maps (**a**) and average profiles (**b**) of MLL4 genomic bindings, H3K4me1 and H3K27ac enrichments and ATAC-seq signals on the three groups of de novo MLL4$^+$ AEs.

(**c**) ChIP-seq profiles of MLL4, H3K4me1 and H3K27ac, ATAC-seq and RNA-seq profiles in f/f and KI cells are displayed on *Fgf8* locus. Group II de novo MLL4$^+$ AEs are highlighted in shades.

**Figure 6-S1. MLL3/4-catalyzed H3K4me1 is dispensable for enhancer activation during early EB differentiation**

ChIP-seq profiles of MLL4, H3K4me1 and H3K27ac, ATAC-seq and RNA-seq profiles in f/f and KI cells are displayed on *Wnt3*, *Mixl1* and *Evx1* loci. Group II de novo MLL4$^+$ AEs are highlighted in shades.

**Figure 6-S2. MLL3/4 proteins, but not MLL3/4-catalyzed H3K4me1 are required for enhancer activation during early EB differentiation**

ChIP-seq data were obtained from GSE50534[5].

(**a-b**) Heat maps (**a**) and average profiles (**b**) of MLL4 genomic bindings, H3K4me1 and H3K27ac enrichments in f/f and KO cells on Group II de novo MLL4$^+$ AEs identified in Figure 5a.

(**c-f**) ChIP-seq profiles of MLL4, H3K4me1 and H3K27ac in f/f and KO cells (**c**, **e**), and RNA-seq profiles in f/f, KO and KI cells (**d**, **f**) are displayed on *Fgf8* (**c-d**), *Wnt3*, *Mixl1* and *Evx1* (**e-f**) loci. The Group II de novo MLL4$^{+}$ AEs are highlighted in shades.

**Figure 7. MLL3/4-catalyzed H3K4me1 is dispensable for activation of super-enhancers during neural differentiation**

f/f and KI ESCs were induced to differentiate into neural progenitor cells (NPCs) and neurons sequentially. D8 NPCs and D16 neurons were collected for RNA-seq. D8 NPCs were collected for ChIP-seq.

(**a-c**) MLL3/4 enzymatic activities are dispensable for differentiation into NPCs. (**a**) Representative microscopic images of NPCs (day 8, *upper*) and neurons (day 16, *lower*). Scale bar = 25 μm. (**b**) Expression levels of 2,526 genes induced from ESC to NPC were presented as scatter plots. Green, black and red dots indicate genes expressed at higher, similar and lower levels in KI NPCs compared to f/f NPCs, respectively. Pearson correlation coefficient (*R*) and *P* value are shown. 2.5-fold cutoff was used to identify differentially expressed genes. (**c**) GO analysis of three gene groups in (B). Terms associated with neural differentiation are highlighted in purple.

(**d-f**) MLL3/4-catalyzed H3K4me1 is dispensable for super-enhancer activation during neural differentiation. (**d**) NPC super-enhancers (SEs) were identified by stitching SOX3 peaks from GSE33059[36] in NPC AEs. SEs were ranked by H3K27ac signal intensities. (**e**) MLL4, H3K4me1 and H3K27ac intensities in ESCs and NPCs on NPC SEs are presented in box plots. (**f**) Expression levels in ESCs and NPCs of genes associated with NPC SEs are presented in box plots. Statistical significance was determined by the two-sided Wilcoxon signed-rank test. Box plot elements: center line, median; box limits, lower and upper quartiles; whiskers, calculated using the Tukey method and outliers are not shown.

(**g**) ChIP-seq profiles of MLL4, H3K4me1 and H3K27ac, and RNA-seq profiles in f/f and KI cells are displayed on *Nes* locus. NPC SEs are indicated by purple bars and de novo MLL4$^{+}$ enhancers within SEs are highlighted in shades.

(**h**) Model on regulation of ESC differentiation by MLL3/4 methyltransferase activities. LDTFs: lineage-determining transcription factors.

**Figure 7-S1. MLL3/4-catalyzed H3K4me1 is dispensable for activation of super-enhancers during neural differentiation**

(**a**) Diagram of neural differentiation.

(**b**) Representative microscope images of NPCs (day 9). Scale bar = 25 μm. White arrows indicate neural rosettes.

(**c-d**) MLL3/4 enzymatic activities are dispensable for differentiation into neurons. (**c**) Expression levels of 3,833 genes induced from ESC to neuron were presented as scatter plots. Green, black and red dots indicate genes expressed at higher, similar and lower levels in KI neurons compared to f/f neurons, respectively. Pearson correlation coefficient (*R*) and *P* value are shown. 2.5-fold cutoff was used to identify differentially expressed genes. (**d**) GO analysis of three gene groups in (E). Terms associated with neural differentiation are highlighted in purple.

(**e-f**) Expression levels of NPC (**e**) and neuron (**f**) markers. Data from RNA-seq are presented as dot plots with median lines (f/f: $n = 2$; KI: $n = 2$).

(**g**) GO analysis of genes associated with NPC SEs. Terms associated with neural differentiation are highlighted in purple.

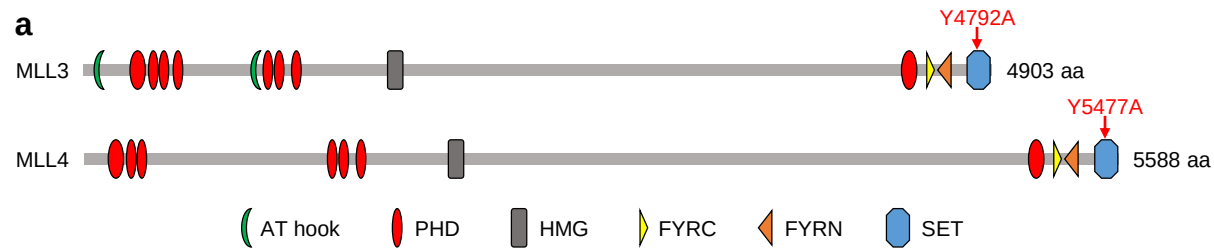
(**h**) ChIP-seq profiles of MLL4, H3K4me1 and H3K27ac, and RNA-seq profiles in f/f and KI cells are displayed on representative gene loci. NPC SEs are indicated by purple bars and de novo MLL4$^{+}$ enhancers within SEs are highlighted in shades.

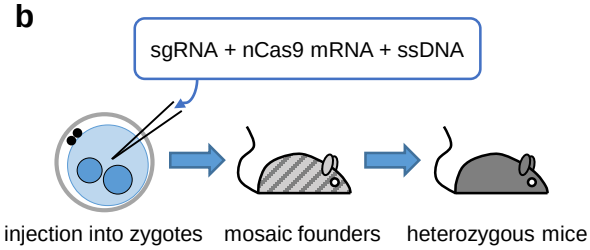
**Supplementary Information**

Supplementary Video 1. Representative day 12 beating EB derived from f/f ESCs during EB differentiation

Supplementary Video 2. Representative day 12 beating EB derived from KI #27 ESCs during EB differentiation

Supplementary Video 3. Representative day 12 beating EB derived from KI #48 ESCs during EB differentiation




**c**

| *Mll3*$^{KI/+}$ x *Mll3*$^{KI/+}$ | |
|---|---|
| Genotype | P21 |
| *Mll3*$^{+/+}$ | 30 |
| *Mll3*$^{KI/+}$ | 69 |
| *Mll3*$^{KI/KI}$ | 19 |
| Total | 118 |

**d**

| *Mll4*$^{KI/+}$ x *Mll4*$^{KI/+}$ | | | |
|---|---|---|---|
| Genotype | E18.5 | P0.5 | P21 |
| *Mll4*$^{+/+}$ | 19 | 18 | 26 |
| *Mll4*$^{KI/+}$ | 46 | 27 | 53 |
| *Mll4*$^{KI/KI}$ | 17 | 1 (1) | 0 |
| Total | 82 | 46 | 59 |

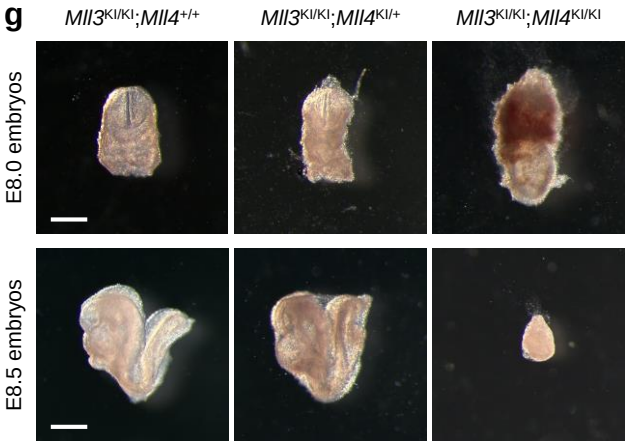
**e**

| *Mll3*$^{KI/+}$;*Mll4*$^{KI/+}$ x *Mll3*$^{KI/+}$;*Mll4*$^{KI/+}$ | | |
|---|---|---|
| Genotype (E14.5) | Observed number | Expected number |
| *Mll3*$^{+/+}$;*Mll4*$^{+/+}$ | 18 | 16 |
| *Mll3*$^{+/+}$;*Mll4*$^{KI/+}$ | 36 | 31 |
| *Mll3*$^{+/+}$;*Mll4*$^{KI/KI}$ | 13 | 16 |
| *Mll3*$^{KI/+}$;*Mll4*$^{+/+}$ | 46 | 31 |
| *Mll3*$^{KI/+}$;*Mll4*$^{KI/+}$ | 72 | 61 |
| *Mll3*$^{KI/+}$;*Mll4*$^{KI/KI}$ | 11 | 31 |
| *Mll3*$^{KI/KI}$;*Mll4*$^{+/+}$ | 18 | 16 |
| *Mll3*$^{KI/KI}$;*Mll4*$^{KI/+}$ | 35 | 31 |
| *Mll3*$^{KI/KI}$;*Mll4*$^{KI/KI}$ | 0 | 16 |
| Total | 249 | 249 |

**f**

| *Mll3*$^{KI/KI}$;*Mll4*$^{KI/+}$ x *Mll3*$^{KI/KI}$;*Mll4*$^{KI/+}$ | | | | |
|---|---|---|---|---|
| Genotype | E8.0 | E8.5 | E10.5 | E14.5 |
| *Mll3*$^{KI/KI}$;*Mll4*$^{+/+}$ | 4 | 4 | 8 | 6 |
| *Mll3*$^{KI/KI}$;*Mll4*$^{KI/+}$ | 11 (2) | 14 (2) | 13 | 6 |
| *Mll3*$^{KI/KI}$;*Mll4*$^{KI/KI}$ | 4 (4) | 6 (6) | 0 | 0 |
| Total | 19 | 24 | 21 | 12 |
| Resorbed | 3 | 2 | 11 | 0 |

**g**




**h**

| Summary | |
|---|---|
| Genotype | Phenotype |
| *Mll3*$^{-/-}$ | Perinatal lethal |
| *Mll3*$^{KI/KI}$ | Survived to adulthood, fertile |
| *Mll4*$^{-/-}$ | Embryonic lethal ~E9.5 |
| *Mll4*$^{KI/KI}$ | Perinatal lethal |
| *Mll3*$^{KI/KI}$;*Mll4*$^{KI/KI}$ | Embryonic lethal ~E8.5 |

**Figure 1. MLL3/4 enzymatic activities are essential for early embryonic development in mice**

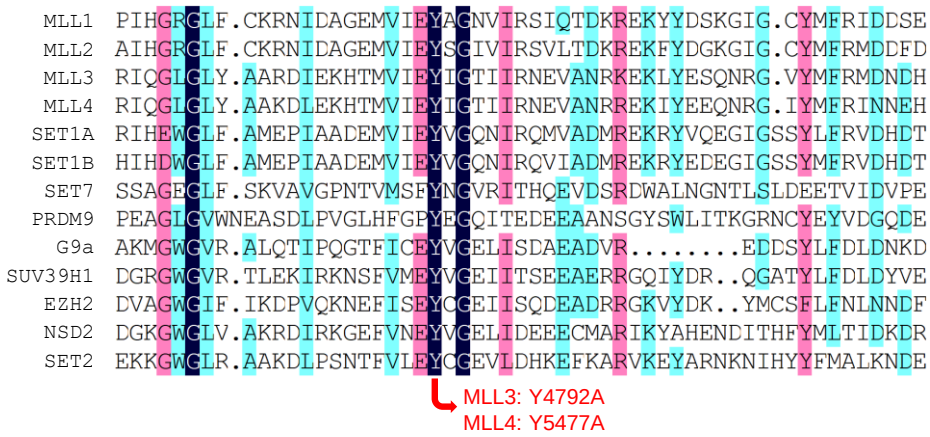



**Figure 1-S1. Alignment of SET domains in different histone methyltransferases**

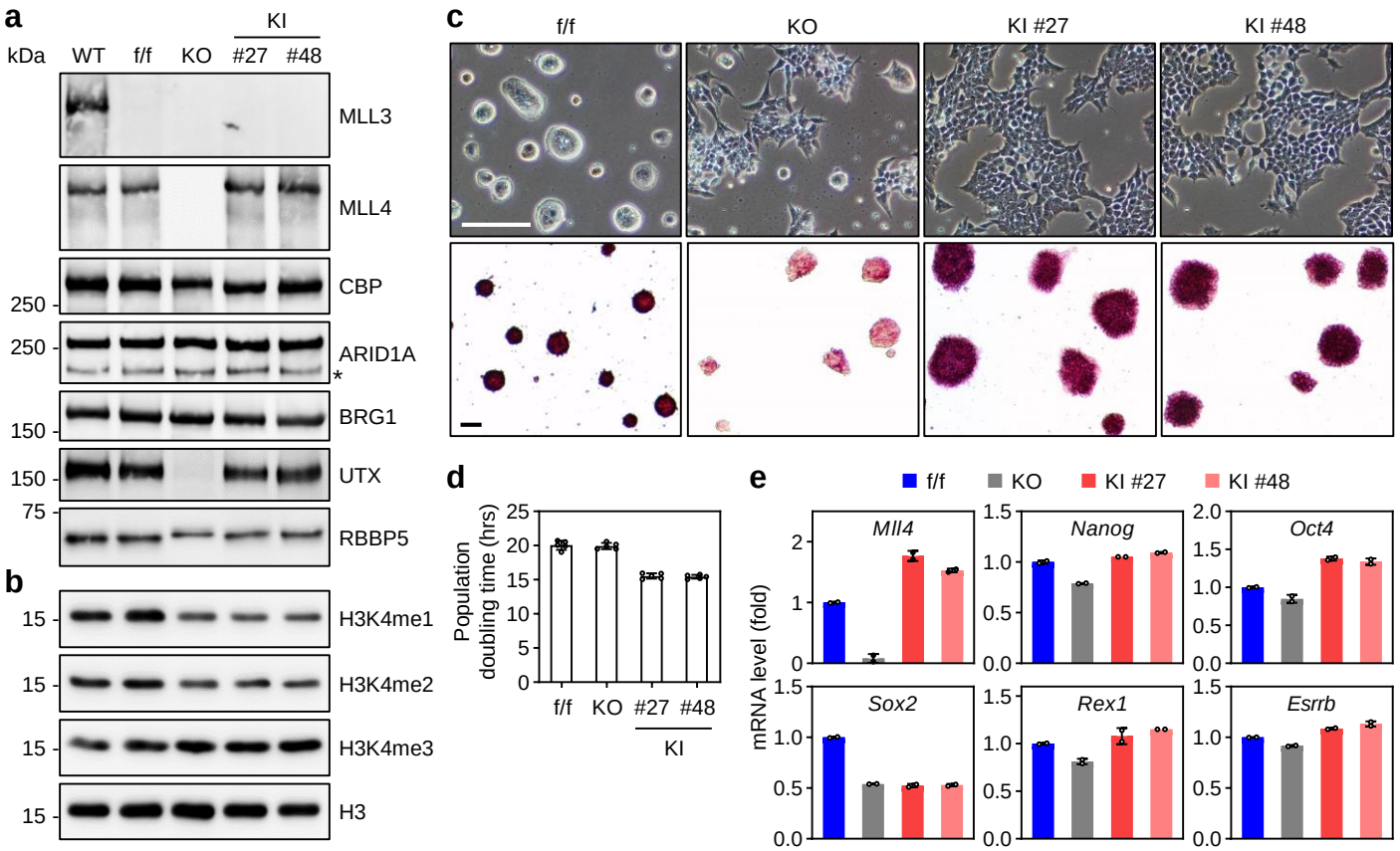



**Figure 2. ESCs lacking MLL3/4 enzymatic activities maintain cell identity**

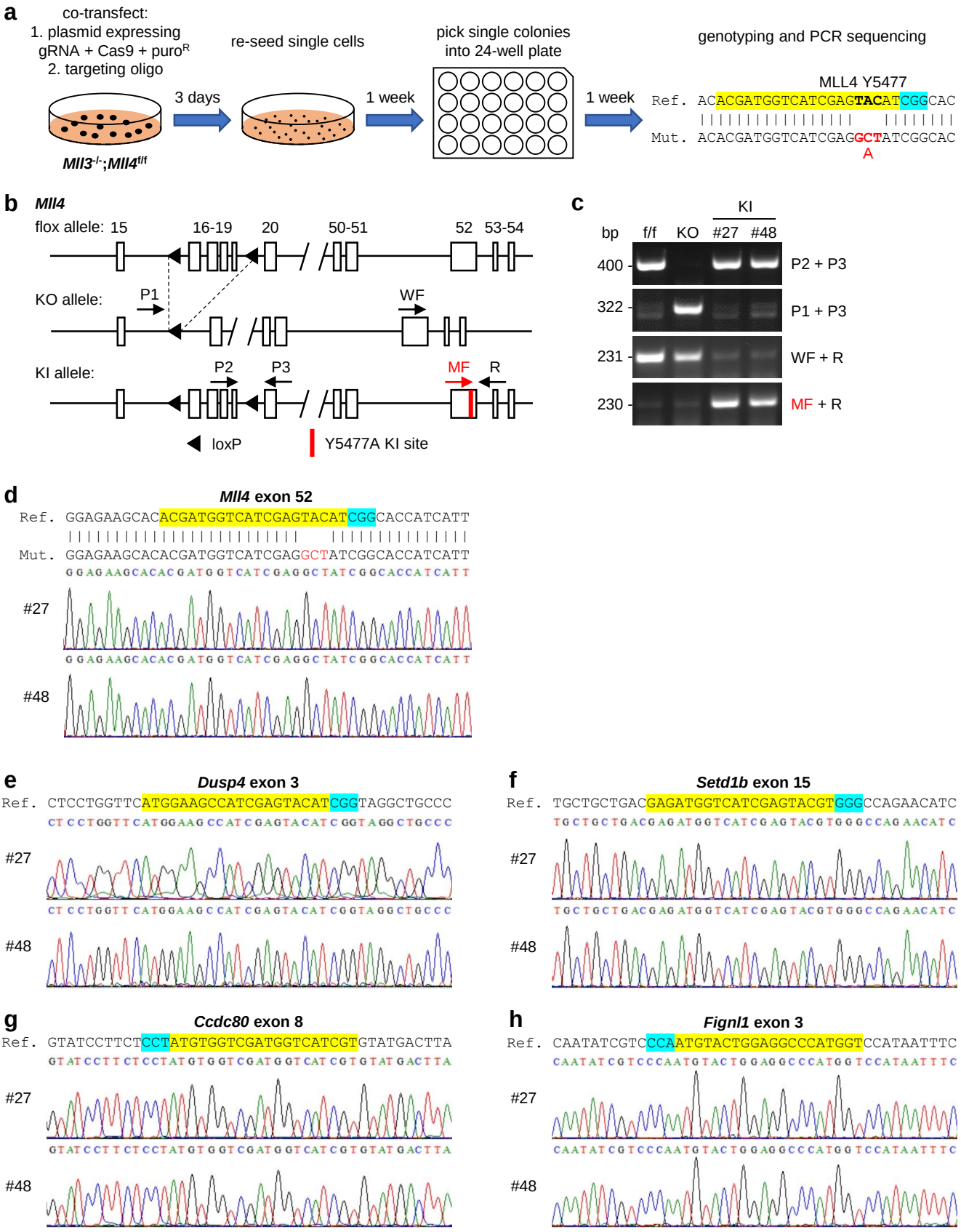



**Figure 2-S1. Generation of ESCs lacking MLL3/4 enzymatic activities**

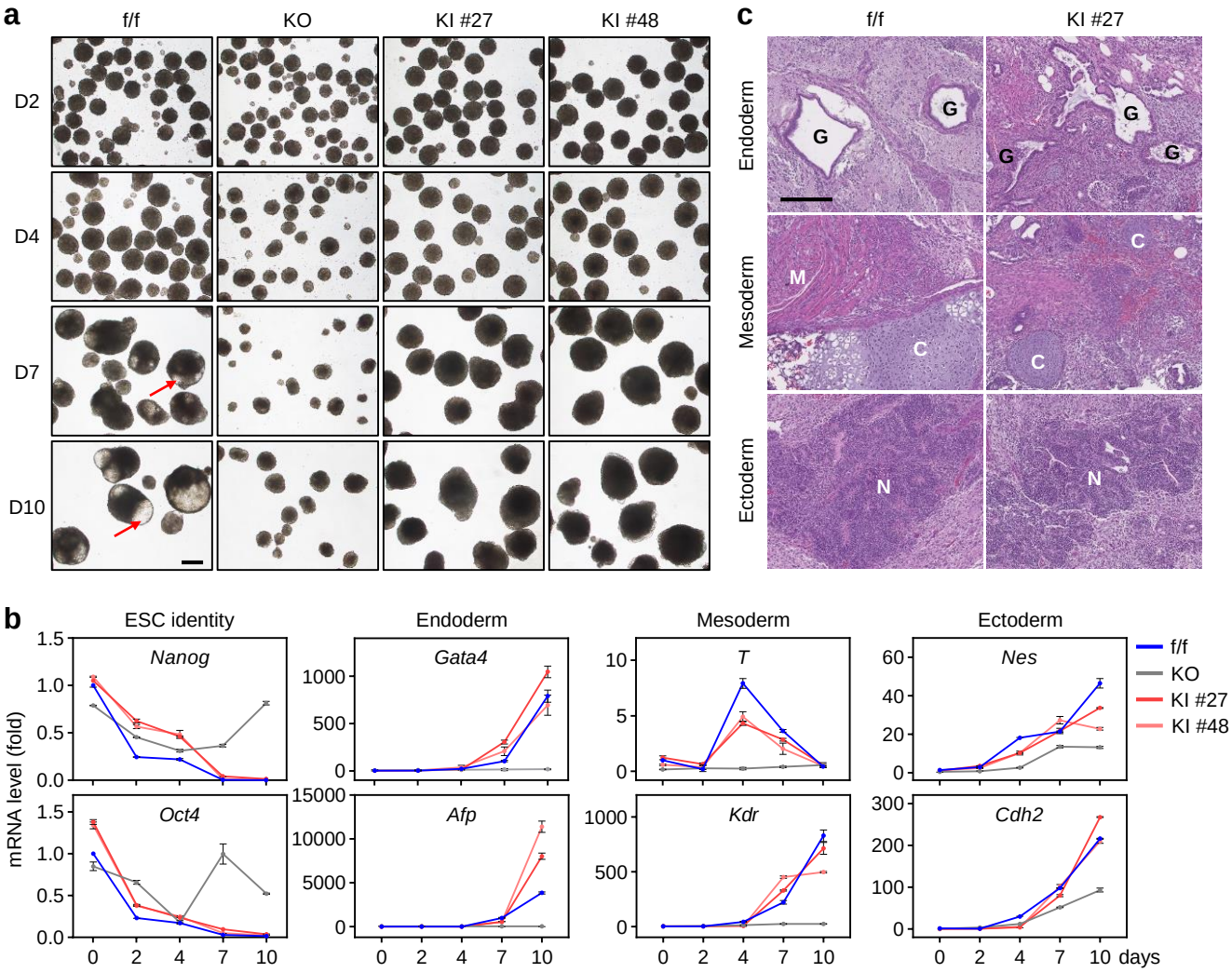



**Figure 3. ESCs lacking MLL3/4 enzymatic activities are capable of differentiating towards the three germ layers but show defects in EB cavitation**

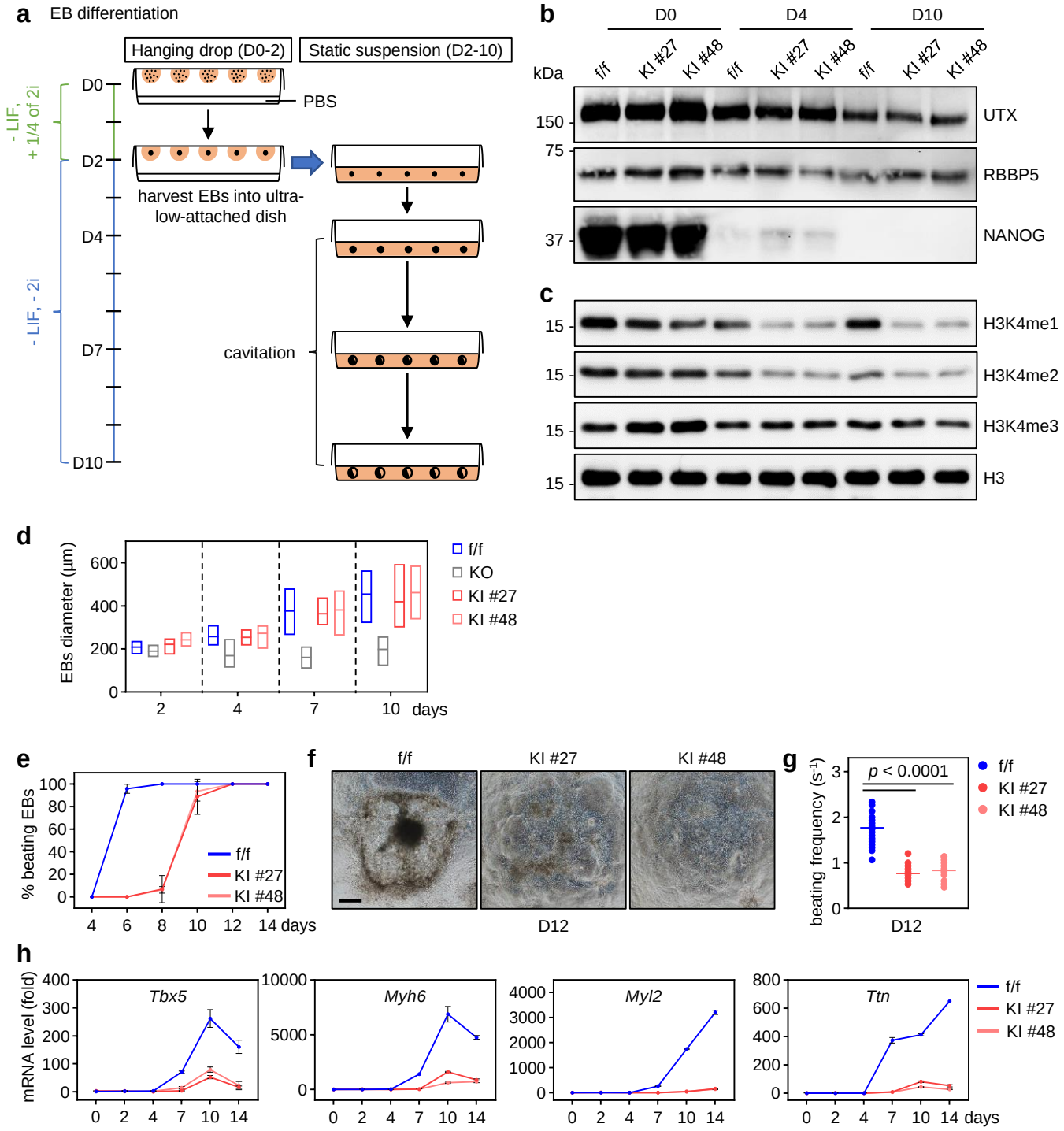



**Figure 3-S1. ESCs lacking MLL3/4 enzymatic activities show defects in cardiomyogenesis**

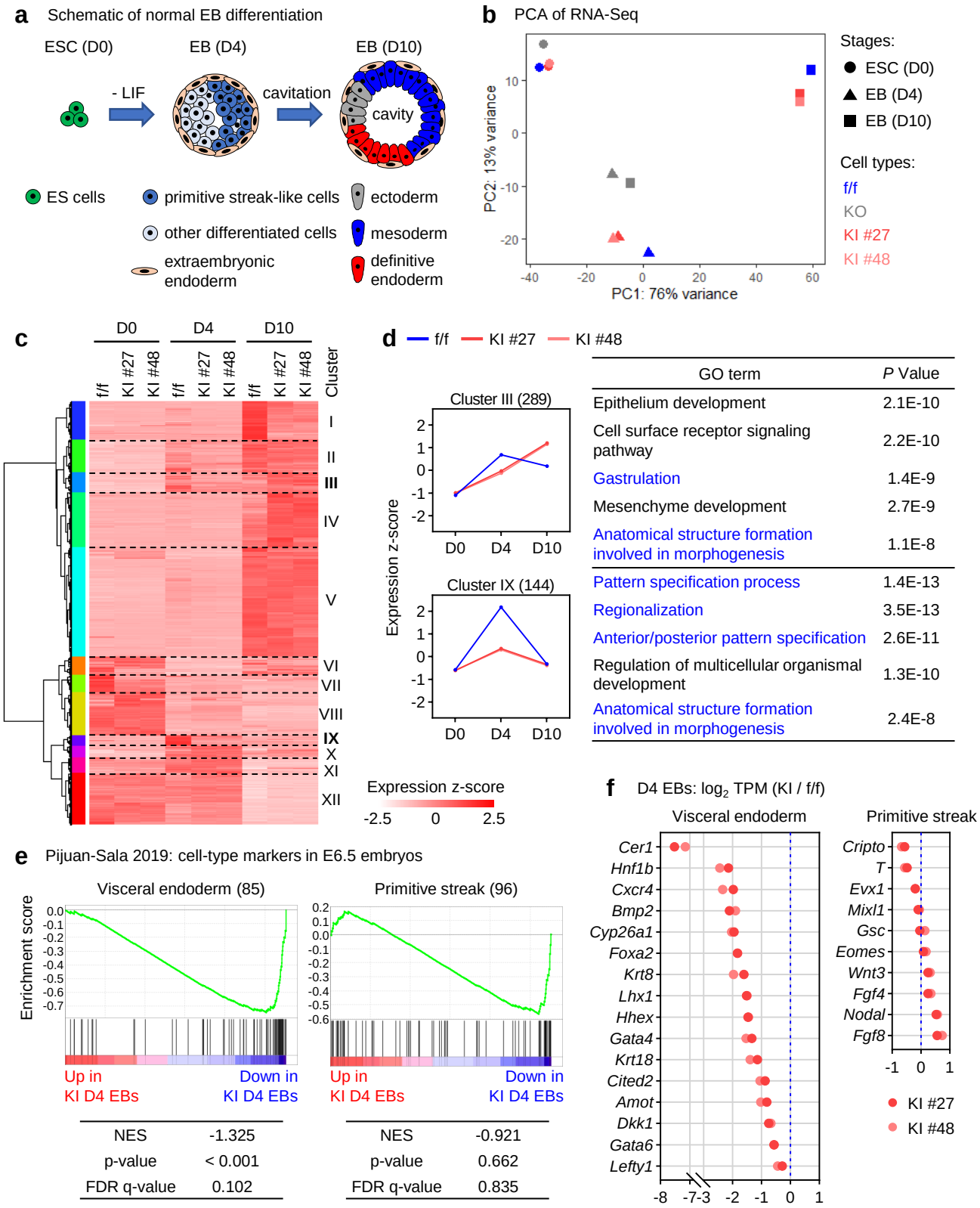



**Figure 4. Loss of MLL3/4 enzymatic activities impairs visceral endoderm induction in EB differentiation**

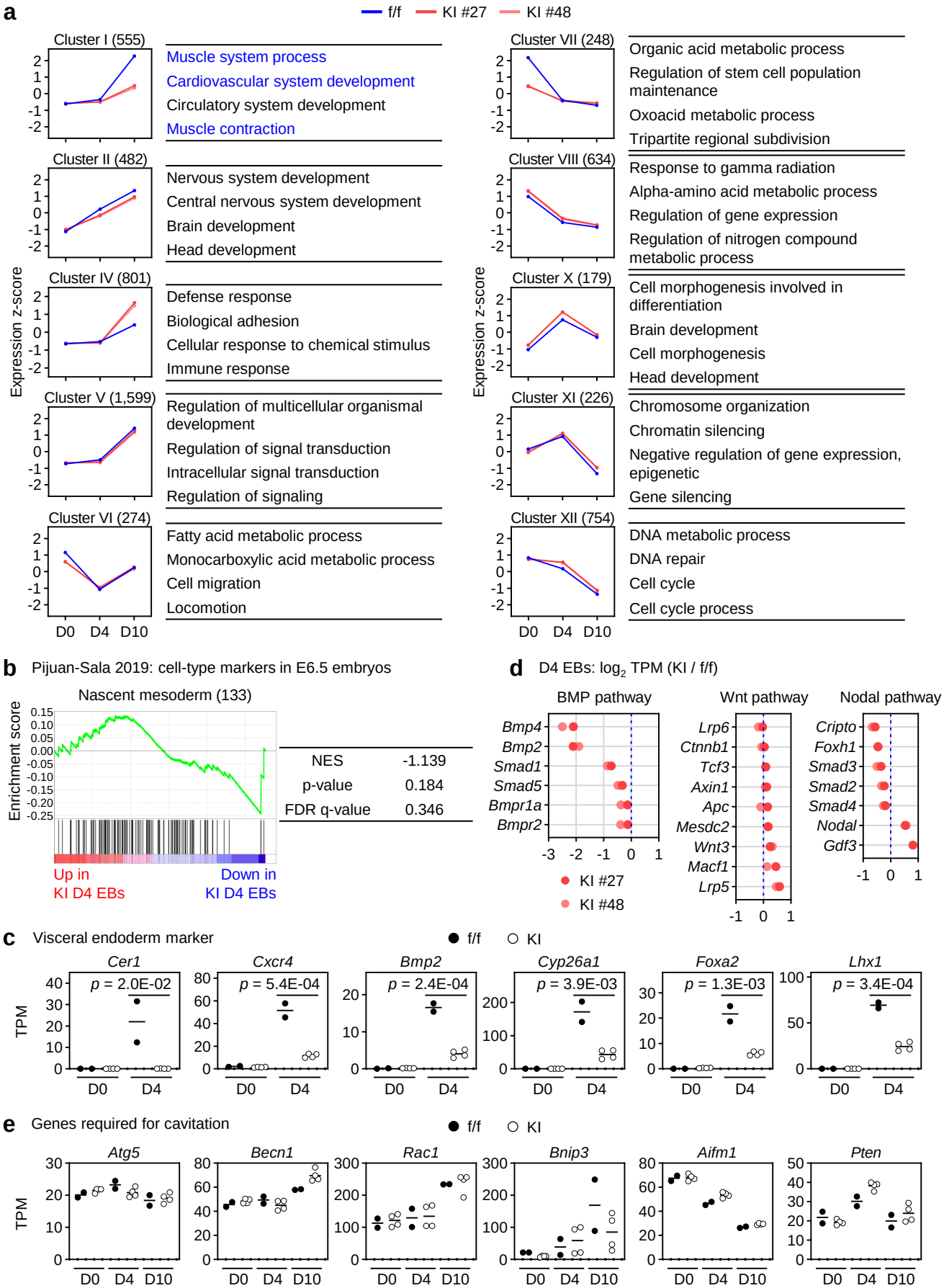
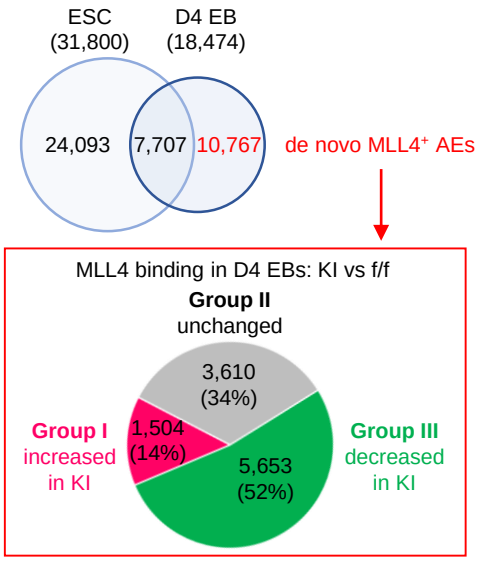
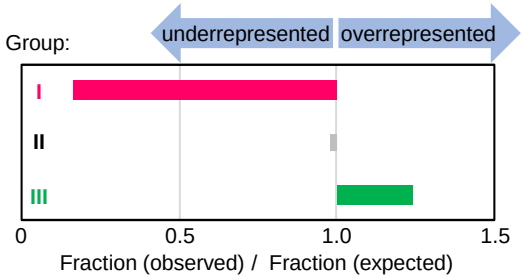



**Figure 4-S1. Loss of MLL3/4 enzymatic activities impairs visceral endoderm induction in EB differentiation**

**a** MLL4⁺ AEs in f/f or KI cells



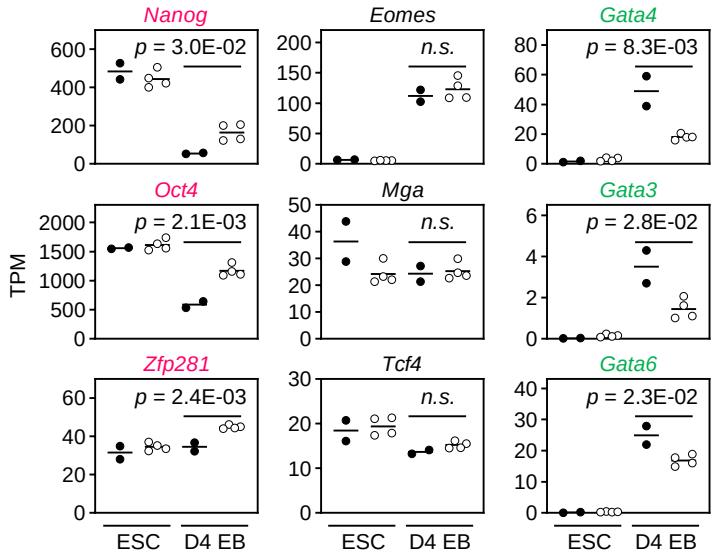

**b** Distribution of 89 de novo MLL4⁺ AEs associated with visceral endoderm markers in E6.5 embryos




**c** Motif analysis: de novo MLL4⁺ AEs in D4 EBs

| Group I | | Group II | | Group III | |
|---|---|---|---|---|---|
| Motif | *P* value | Motif | *P* value | Motif | *P* value |
| ZIC3 | 8E-26 | ZIC3 | 4E-125 | ZIC3 | 5E-98 |
| **NANOG** | 4E-23 | TBX2 | 1E-61 | CBFA2T2 | 6E-82 |
| ZFP219 | 3E-21 | **EOMES** | 2E-58 | POLR2B | 2E-80 |
| **OCT4** | 7E-20 | **MGA** | 1E-43 | **GATA4** | 1E-77 |
| **ZFP281** | 1E-19 | POLR2B | 3E-36 | **GATA3** | 4E-60 |
| BCL11A | 2E-19 | **TCF4** | 1E-33 | LMO2 | 4E-69 |
| SMAD2/3 | 6E-18 | CBFA2T2 | 1E-33 | **GATA6** | 6E-58 |

**d** RNA-seq: TFs linked to de novo MLL4⁺ AEs in 3 groups ● f/f ○ KI




**e**

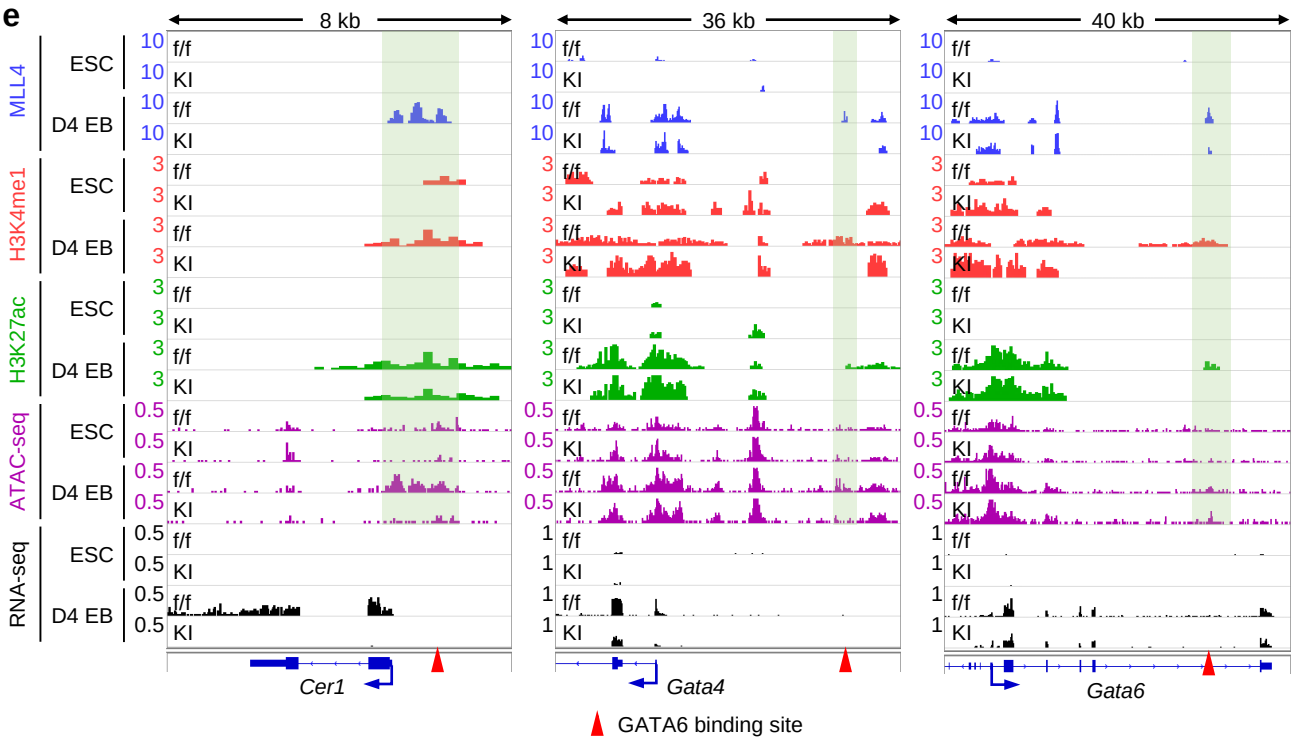



**Figure 5. MLL3/4 enzymatic activities are required for induction of visceral endoderm TFs GATA4/6 in the early phase of EB differentiation**

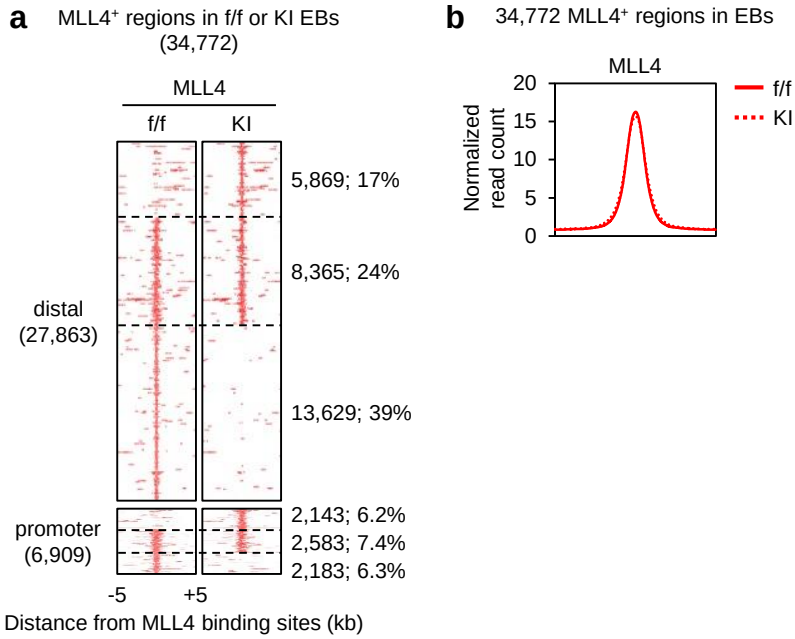
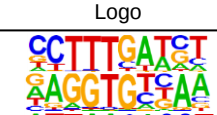
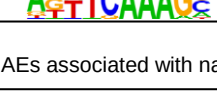



**Figure 5-S1. Genomic binding profiles of MLL4 in D4 EBs**

**a** Motif analysis: 89 de novo MLL4$^{+}$ AEs associated with visceral endoderm markers in E6.5 embryos

| Motif | Logo | *P* value |
|---|---|---|
| **GATA6** | | 1E-7 |
| EOMES |  | 1E-6 |
| **GATA3** | | 1E-6 |
| **GATA2** |  | 1E-6 |
| **GATA1** | | 1E-5 |
| ZFP281 |  | 1E-5 |
| **GATA4** |  | 1E-5 |

**b** Motif analysis: 95 de novo MLL4$^{+}$ AEs associated with primitive streak markers in E6.5 embryos

| Motif | Logo | *P* value |
|---|---|---|
| LEF |  | 1E-5 |
| TBX6 |  | 1E-4 |
| EOMES | | 1E-4 |
| ZIC3 |  | 1E-4 |
| TCF7 | | 1E-4 |
| FOXO3 |  | 1E-4 |
| TCF3 |  | 1E-3 |

**c** Motif analysis: 41 de novo MLL4$^{+}$ AEs associated with nascent mesoderm markers in E6.5 embryos

| Motif | Logo | *P* value |
|---|---|---|
| ZIC2 | | 1E-3 |
| LHX1 |  | 1E-3 |
| PHOX2A | | 1E-3 |

**Figure 5-S2. GATA family TFs are highly enriched in de novo MLL4$^{+}$ AEs associated with visceral endoderm markers**

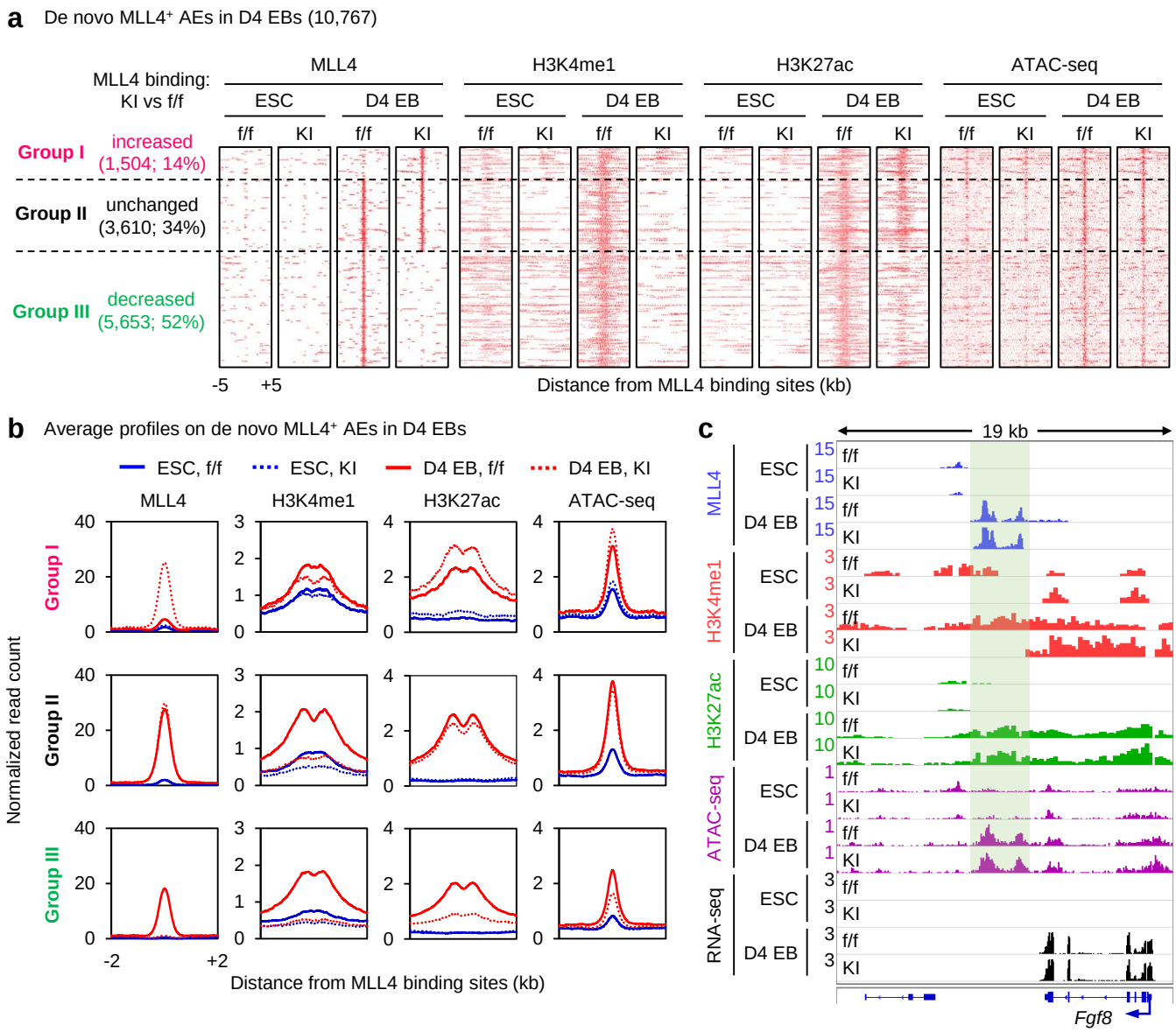



**Figure 6. MLL3/4-catalyzed H3K4me1 is dispensable for enhancer activation during early EB differentiation**

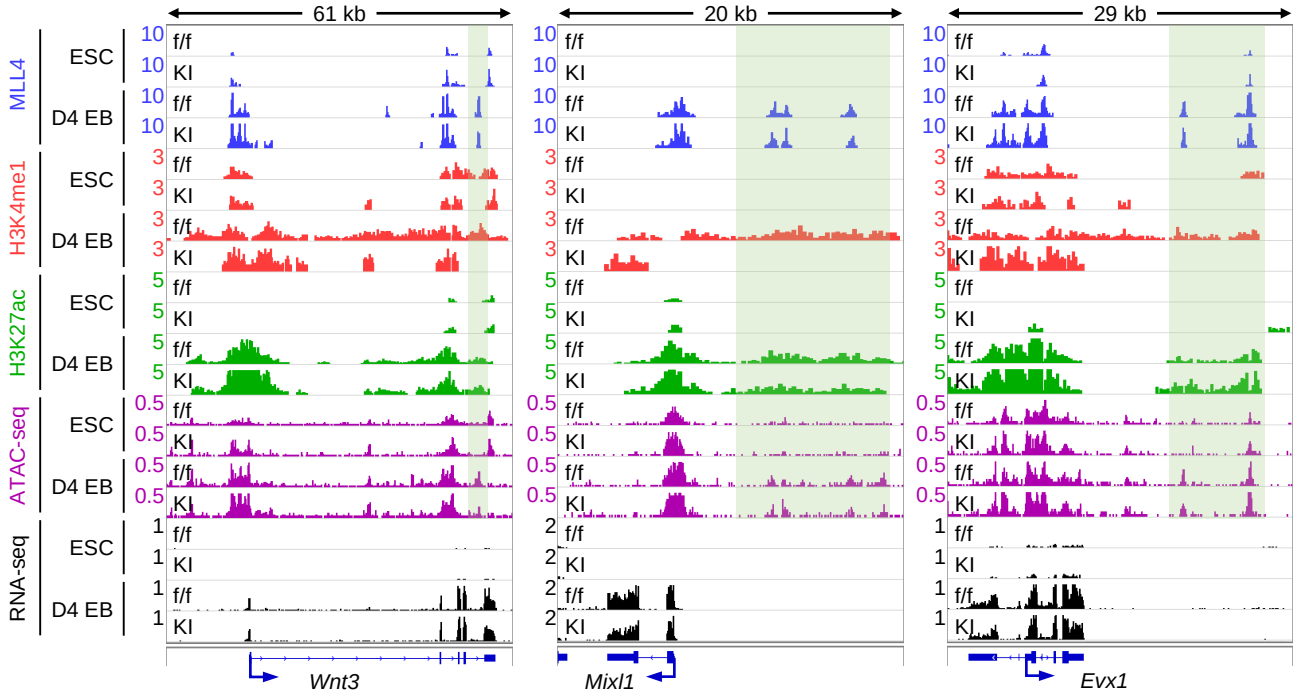



**Figure 6-S1. MLL3/4-catalyzed H3K4me1 is dispensable for enhancer activation during early EB differentiation**

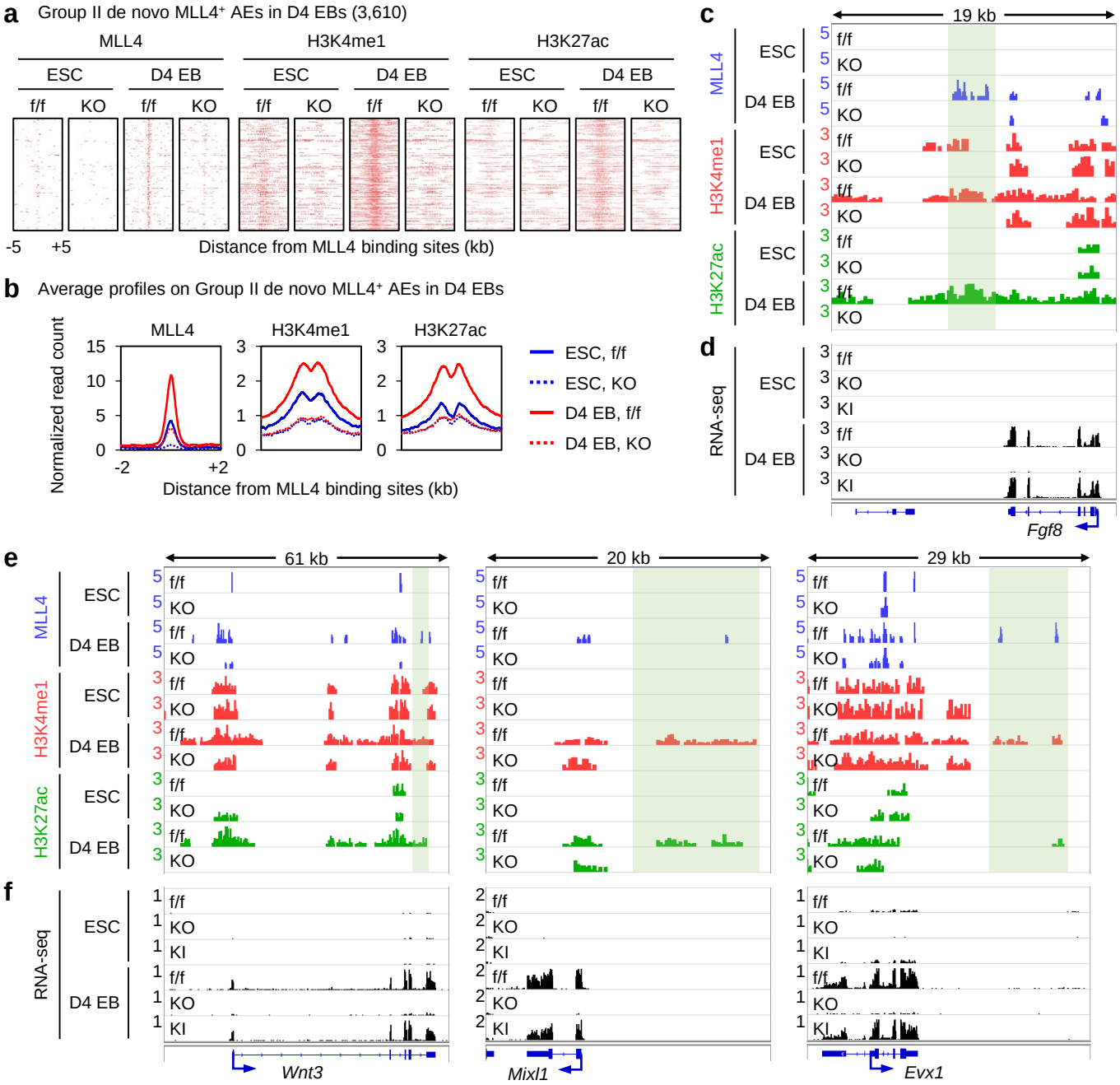
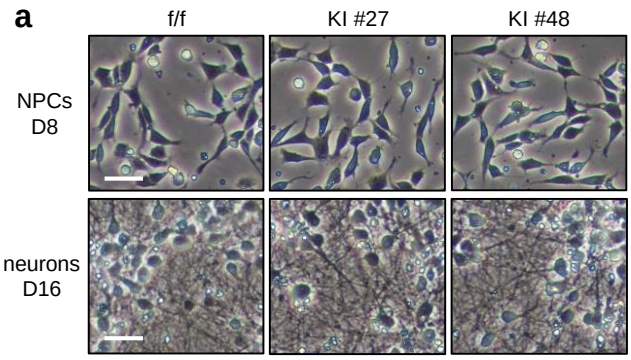
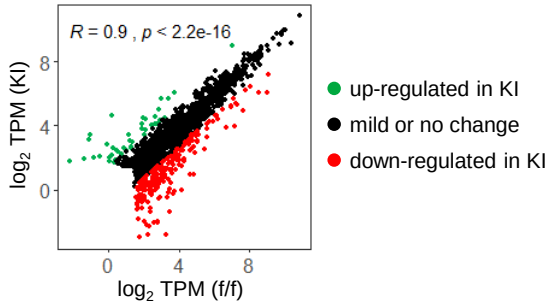


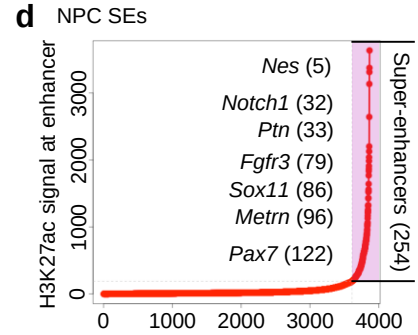

**Figure 6-S2. MLL3/4 proteins, but not MLL3/4-catalyzed H3K4me1 are required for enhancer activation during early EB differentiation**

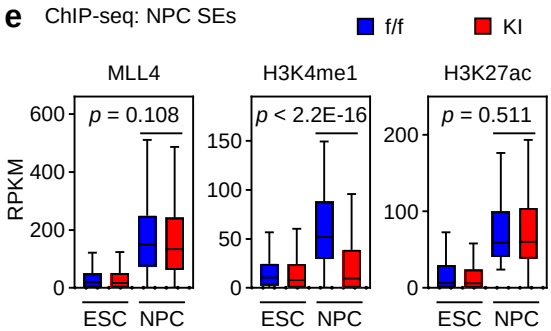


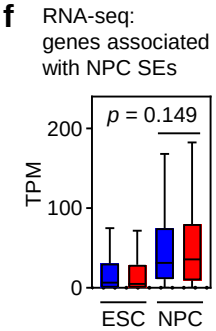

c GO analysis: 2,526 genes induced from ESC to NPC

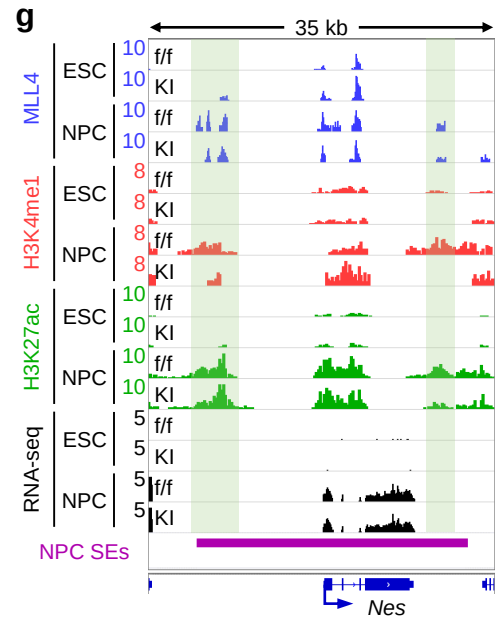
| Gene group | GO term | *P* Value |
|---|---|---|
| up-regulated in KI (42) | Regionalization | 8.7E-10 |
| | Pattern specification process | 1.8E-8 |
| | Anterior/posterior pattern specification | 2.6E-8 |
| | Embryonic morphogenesis | 3.1E-8 |
| | Organ morphogenesis | 1.1E-7 |
| mild or no change (2,283) | Nervous system development | 6.5E-79 |
| | Neurogenesis | 4.1E-64 |
| | Generation of neurons | 5.7E-64 |
| | Neuron differentiation | 1.4E-59 |
| | Neuron development | 6.6E-45 |
| down-regulated by in (201) | Blood vessel development | 1.0E-15 |
| | Vasculature development | 3.3E-15 |
| | Blood vessel morphogenesis | 4.1E-15 |
| | Organ morphogenesis | 4.5E-15 |
| | Circulatory system development | 8.6E-14 |



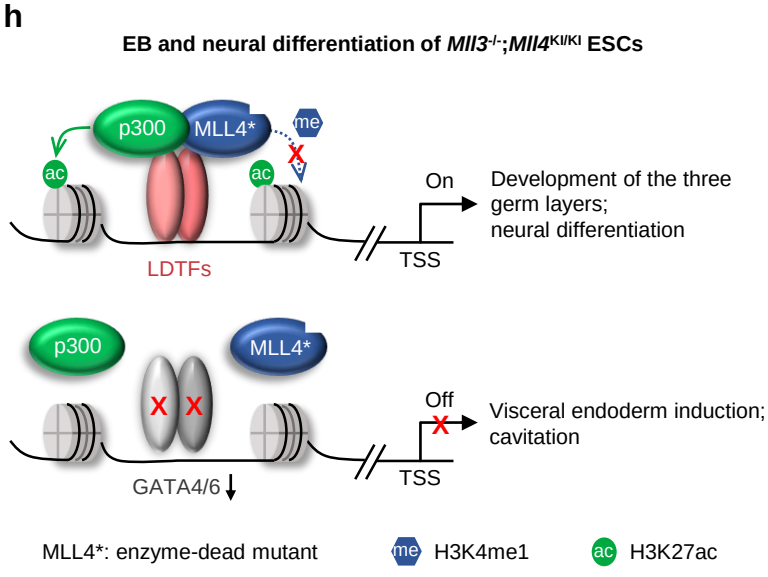
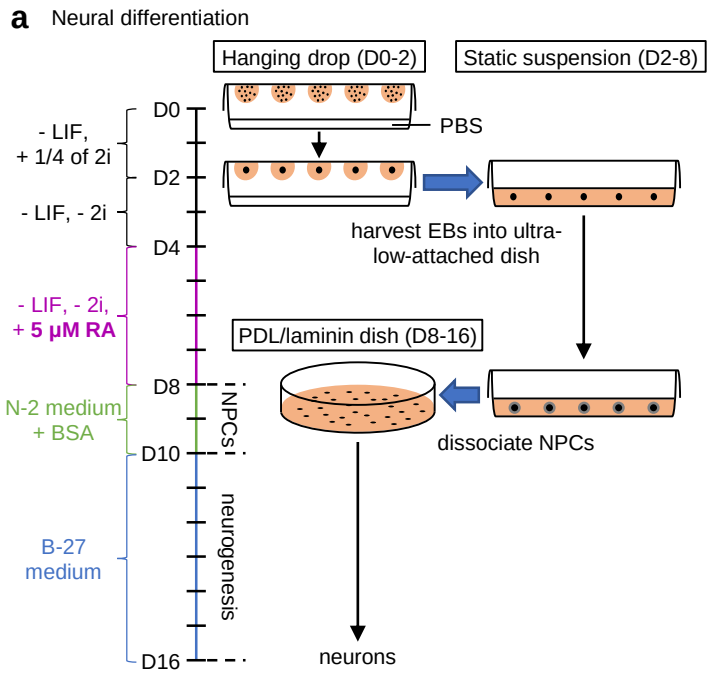
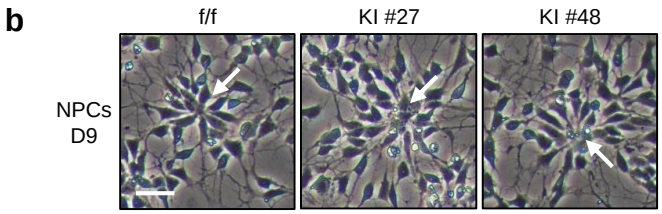
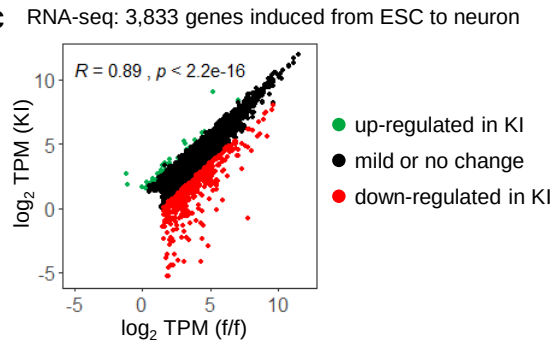




**Figure 7. MLL3/4-catalyzed H3K4me1 is dispensable for activation of super-enhancers during neural differentiation**

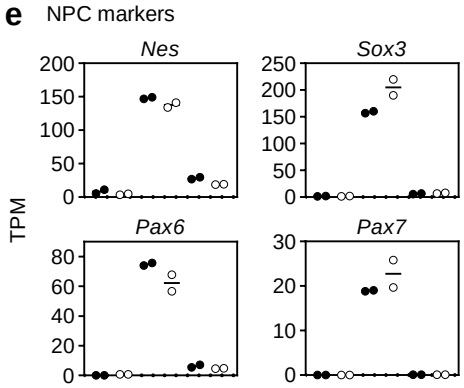


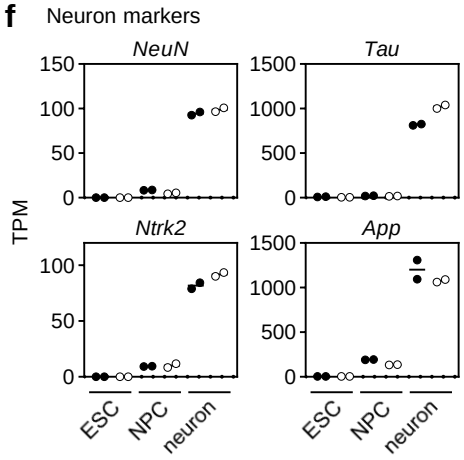




**d** GO analysis: 3,833 genes induced from ESC to neuron

| Gene group | GO term | *P* Value |
|---|---|---|
| up-regulated in KI (35) | No significant term | |
| mild or no change (3,431) | Nervous system development | 5.7E-115 |
| | Generation of neurons | 2.1E-91 |
| | Neurogenesis | 8.8E-91 |
| | Neuron differentiation | 3.8E-89 |
| | Neuron development | 6.5E-85 |
| down-regulated in KI (367) | Circulatory system development | 1.5E-20 |
| | Cardiovascular system Development | 1.5E-20 |
| | Vasculature development | 3.9E-20 |
| | Blood vessel development | 6.0E-20 |
| | Organ morphogenesis | 2.3E-18 |




**g** GO analysis: genes associated with NPC SEs

| GO term | *P* Value |
|---|---|
| Nervous system development | 5.30E-19 |
| Generation of neurons | 1.80E-18 |
| Neurogenesis | 1.80E-18 |
| Central nervous system development | 8.80E-16 |
| Positive regulation of RNA metabolic process | 1.00E-15 |
| Neuron differentiation | 2.30E-15 |

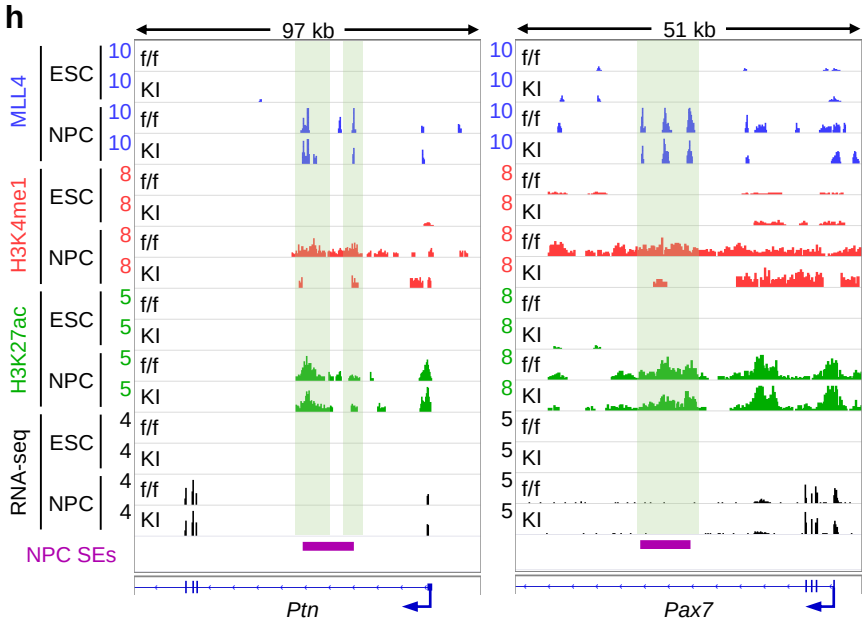



**Figure 7-S1. MLL3/4-catalyzed H3K4me1 is dispensable for activation of super-enhancers during neural differentiation**