



Uncoupling histone H3K4 trimethylation from developmental gene expression via an equilibrium of COMPASS, Polycomb and DNA methylation

Delphine Douillet^{1,2}, Christie C. Sze^{1,2}, Caila Ryan^{1,2}, Andrea Piunti^{1,2}, Avani P. Shah^{1,2}, Michal Ugarenko^{1,2}, Stacy A. Marshall^{1,2}, Emily J. Rendleman^{1,2}, Didi Zha^{1,2}, Kathryn A. Helmin³, Zibo Zhao^{1,2}, Kaixiang Cao^{1,2}, Marc A. Morgan^{1,2}, Benjamin D. Singer^{1,2,3}, Elizabeth T. Bartom^{1,2}, Edwin R. Smith^{1,2} and Ali Shilatifard^{1,2}

The COMPASS protein family catalyzes histone H3 Lys 4 (H3K4) methylation and its members are essential for regulating gene expression. MLL2/COMPASS methylates H3K4 on many developmental genes and bivalent clusters. To understand MLL2-dependent transcriptional regulation, we performed a CRISPR-based screen with an MLL2-dependent gene as a reporter in mouse embryonic stem cells. We found that MLL2 functions in gene expression by protecting developmental genes from repression via repelling PRC2 and DNA methylation machineries. Accordingly, repression in the absence of MLL2 is relieved by inhibition of PRC2 and DNA methyltransferases. Furthermore, DNA demethylation on such loci leads to reactivation of MLL2-dependent genes not only by removing DNA methylation but also by opening up previously CpG methylated regions for PRC2 recruitment, diluting PRC2 at Polycomb-repressed genes. These findings reveal how the context and function of these three epigenetic modifiers of chromatin can orchestrate transcriptional decisions and demonstrate that prevention of active repression by the context of the enzyme and not H3K4 trimethylation underlies transcriptional regulation on MLL2/COMPASS targets.

Post-translational modifications of chromatin play diverse roles in eukaryotic cell function¹. One of these post-translational marks of histones in chromatin, the H3K4 trimethyl mark (H3K4me3), is a broadly evolutionarily conserved modification that has been extensively studied². In the budding yeast *Saccharomyces cerevisiae*, H3K4 methylation is performed by Set1/COMPASS, a complex containing the lysine methyltransferase (KMT) Set1 and several accessory proteins^{3–7}. In *Drosophila melanogaster*, three variants of COMPASS have been reported, each with a distinct KMT: dSet1, Trithorax (Trx) or Trithorax-related protein (Trr)⁸. Mammals have six forms of COMPASS with two close relatives each for the *Drosophila* enzymes: SET1A (SETD1A) and SET1B (SETD1B), related to dSet1; MLL1 (KMT2A) and MLL2 (KMT2B), related to Trx; and MLL3 (KMT2C) and MLL4 (KMT2D), related to Trr³. Each of these COMPASS family members appears to have unique properties and exhibits both redundant and non-redundant functions. In mammalian cells, SET1A/B reportedly implement most of the H3K4me3 in cells^{9,10}, while MLL3/4 implement monomethylation of H3K4 at enhancers^{11–13}. MLL1/2 implement H3K4me2 and H3K4me3 at developmental genes^{14,15}. MLL1 is required for hematopoiesis and its gene is frequently translocated in leukemia^{16–18}, while MLL2 is required for early development¹⁹ and heterozygous variants in the *MLL2* gene are causative of childhood dystonia^{20,21}.

The extent to which the overall biological function of COMPASS is related to its KMT activity has been re-evaluated during the past few years. Indeed, the SET domain of the SET1A protein has been shown to be dispensable for mouse embryonic stem cell (mESC)

proliferation and self-renewal²². Similarly, H3K4 monomethylation catalyzed by Trr and mammalian MLL3/4 COMPASS-like proteins at enhancers was shown to be dispensable for enhancer activation²³ and for development²⁴; however, Trr and MLL3/4 proteins are required for viability and development²⁴. Therefore, the role of H3K4 methylation and COMPASS family in regulating transcription and development remains a complex issue that requires further molecular investigations.

Our previous studies demonstrated that although MLL2 is required broadly for the implementation of H3K4me3 at bivalent genes in mESCs^{14,25}, surprisingly the expression of only a few genes depended on the H3K4 methylation function of MLL2 (ref. ¹⁴). In the present study, we employed a genome-wide genetic screening approach to identify factors required for the proper regulation of transcription of these MLL2-dependent target genes to understand the molecular basis for this epigenetic process. In so doing, we endogenously tagged *Magohb* (the most downregulated gene in the absence of MLL2)^{14,19} and performed a CRISPR–Cas9-mediated screen for regulators of *Magohb* expression in the presence or absence of MLL2. On the basis of this molecular screen, we found that loss of SET1A/B COMPASS leads to upregulation of ~1,200 genes downregulated as the result of MLL2 loss. Our mechanistic studies demonstrated that in the absence of MLL2, these loci are repressed either by DNA methylation and/or H3K27me3. The effect of SET1A/B COMPASS loss in upregulation of these genes is due to its positive regulation of DNA methyltransferase gene transcription, which in turn can influence H3K27me3. Therefore, our study

¹Simpson Querrey Center for Epigenetics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. ²Department of Biochemistry and Molecular Genetics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. ³Division of Pulmonary and Critical Care, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. e-mail: ASH@Northwestern.edu

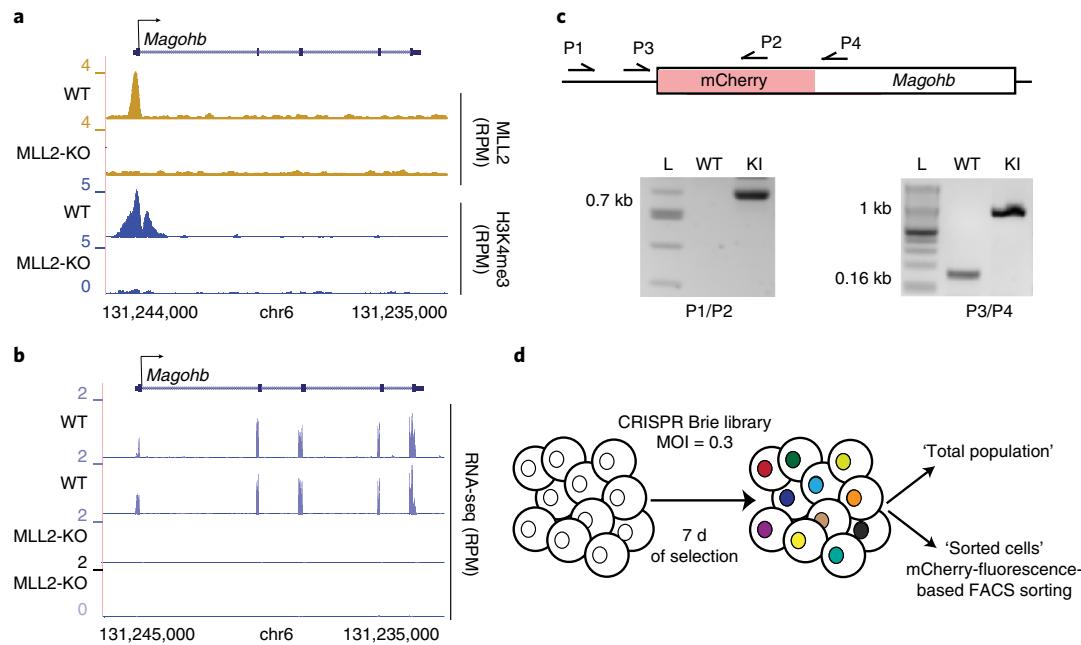


Fig. 1 | Generation of endogenously tagged mCherry-Magohb mESCs. **a**, ChIP-seq tracks of MLL2 and H3K4me3 occupancy at the *Magohb* locus in WT and MLL2-KO mESCs. The experiment was repeated two times independently with similar results. RPM, reads per million. **b**, RNA-seq tracks for *Magohb* in WT and MLL2-KO mESCs. Two biological replicates are displayed. $n=2$ independent biological replicates. **c**, Top: strategy for integration of a 3xFLAG-mCherry tag at the N terminus of *Magohb*. The primers (P1 to P4) used for genotyping are shown as arrows. Bottom: PCR genotyping showing the WT and a homozygous mCherry-Magohb knock-in (KI) clone (L, ladder). **d**, A diagram of the CRISPR screen.

demonstrates the existence of a molecular interplay among the COMPASS family, H3K27me3 and DNA methylation for the proper regulation of MLL2-dependent developmental target genes.

Results

Genome-wide CRISPR screen to identify factors required for MLL2-dependent *Magohb* expression. Chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) for MLL2 and H3K4me3 demonstrated that MLL2 binds to the *Magohb* promoter, and that H3K4me3 is strongly reduced at this locus after knocking out MLL2 (MLL2 KO; Fig. 1a), in accordance with previously reported data^{14,19}. RNA sequencing (RNA-seq) analysis confirmed that *Magohb* expression was lost in MLL2-KO mESCs compared to wild-type (WT) mESCs (Fig. 1b and Extended Data Fig. 1a). Since *Magohb* transcription was fully dependent on MLL2 binding to its promoter, we chose this gene to create a reporter to study MLL2-dependent transcriptional regulation.

An endogenous reporter was generated in WT and MLL2-KO mESCs by inserting a 3xFlag-mCherry sequence at the amino terminus of *Magohb* via CRISPR-Cas9-mediated genome editing (Fig. 1c). To ensure our reporter was functional, we complemented MLL2-KO mESCs using a MLL2-expressing vector and could effectively rescue the mCherry-MAGOHB signal (Extended Data Fig. 2a). In contrast, knockdown of MLL2 in WT mESCs strongly decreased the mCherry-MAGOHB signal (Extended Data Fig. 2b), confirming that our fusion gene maintained its MLL2-dependent expression.

We subsequently performed a CRISPR screen using our generated reporter cell line by isolating populations of cells on the basis of their mCherry fluorescence by FACS (Fig. 1d). Briefly, Cas9-expressing cells were selected (Extended Data Fig. 2c) and transduced with the Brie CRISPR-KO library²⁶ at a multiplicity of infection (MOI) rate of 0.3 and a library representation of at least 400 cells infected per single-guide RNA (sgRNA). After 7 d of puromycin selection, a fraction of the population was collected to assess

the presort sgRNA library coverage ('Total population'). The rest of the population was sorted ('Sorted cells') to isolate cells carrying a mCherry signal lower than that of the negative control for the WT cells or higher than that of the negative control for MLL2-KO cells (Extended Data Fig. 2d). Genomic DNA was extracted and sgRNA abundance was measured by deep sequencing. Each sgRNA with a $\log_2[\text{fold change}] (\log_2[\text{FC}]) > 1$ in the 'Sorted cells' versus the 'Total population' was considered as an enriched sgRNA, and the total number of enriched sgRNAs per gene was calculated. As expected, *Magohb* and *Mll2* were identified as top hits in the cells sorted for lower MAGOHB expression in the WT mESCs (Extended Data Fig. 3a and Supplementary Table 1) demonstrating the specificity of our screen. Moreover, MAGOH and MAGOHB were previously identified as being paralog dependent, as only the simultaneous depletion of MAGOH and MAGOHB was shown to impair viability^{27,28}. In accordance, sgRNAs targeting *Magoh* were depleted from our CRISPR screen performed in MLL2-KO mESCs that do not express MAGOHB, further validating our approach (Supplementary Table 2).

Knockdowns of SET1A/B COMPASS members rescue *Magohb* expression in MLL2-KO cells. To identify repressors of *Magohb* expression in the absence of MLL2, we examined the candidates from the screen performed in MLL2-KO mESCs and identified eight potential hits that met our stringent threshold (four out of four sgRNAs enriched; Fig. 2a and Supplementary Table 2). To individually test each candidate, the strongest scoring sgRNA per gene was used to infect MLL2-KO *Magohb*-knock-in cells and was then selected for 7 d, before assessing their effect on *Magohb* expression by western blot and reverse-transcription quantitative PCR (RT-qPCR) (Fig. 2b,c and Extended Data Fig. 3b). Note that we found only one out of the eight hits as impacting on *Magohb* expression, suggesting that the seven remaining potential hits were false positives or that the sgRNAs used in these cases were ineffective at fully depleting their target.

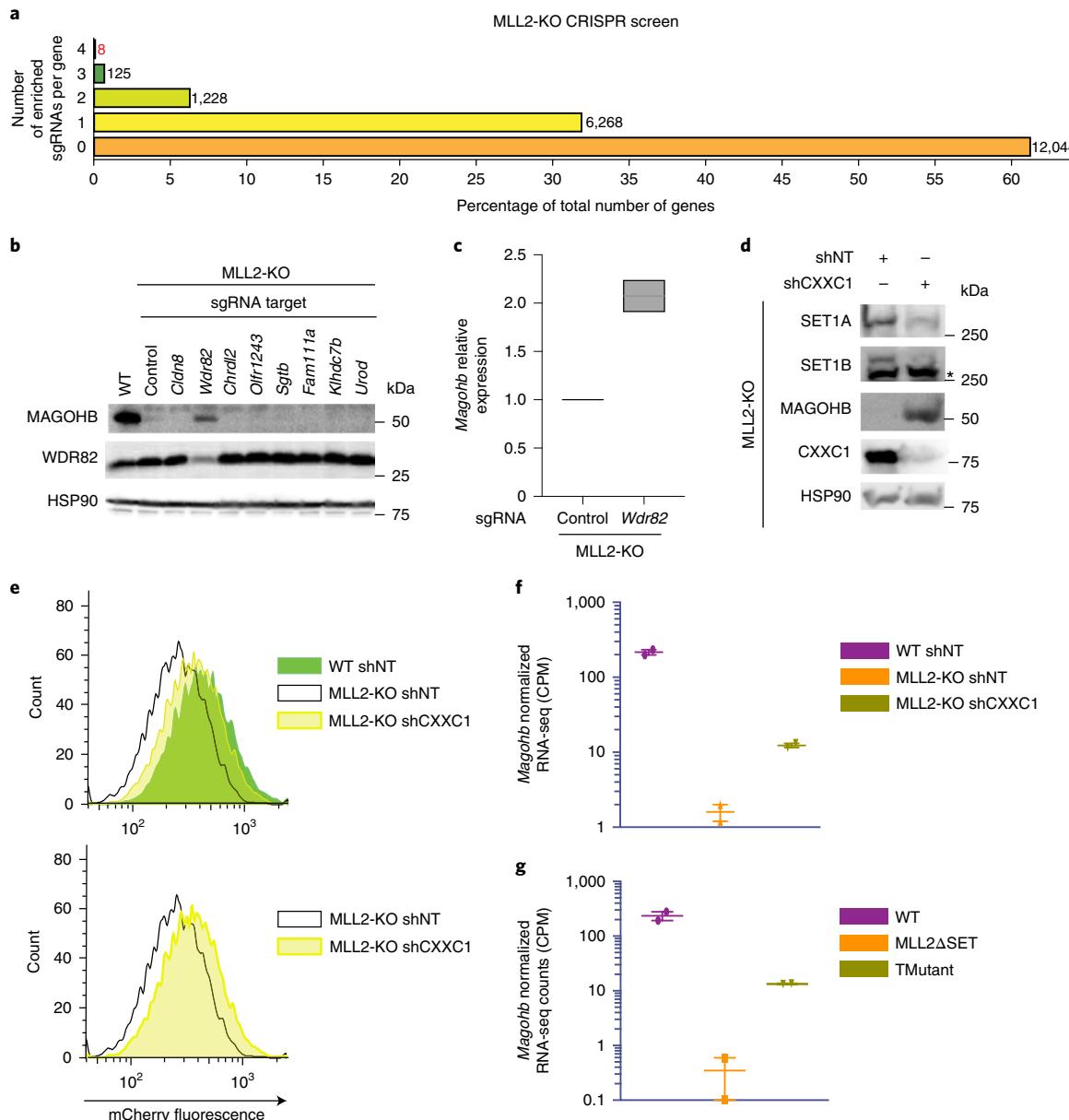


Fig. 2 | Knockdowns of SET1A/B COMPASS complex members rescue *Magohb* expression in MLL2-KO cells. **a**, CRISPR screen results showing the number of enriched sgRNAs in the 'Sorted cells' compared to the 'Total population' per gene. The majority of genes fall into the category '0 sgRNAs enriched out of 4'. **b**, A western blot of the MAGOHB, WDR82 and HSP90 (loading control) protein levels following sgRNA targeting of each of the eight candidates. The experiment was repeated two times independently with similar results. **c**, The MAGOHB expression level as assessed by reverse-transcription quantitative PCR in MLL2-KO mESCs expressing a control sgRNA (Control) or a sgRNA targeting *Wdr82* ($n=2$). The box indicates the standard deviation of the mean. The line across the box depicts the mean value. **d**, A western blot of the SET1A, SET1B, FLAG-MAGOHB, CXXC1 and HSP90 (loading control) protein levels in MLL2-KO mESCs expressing shNT or shCXXC1. The experiment was repeated two times independently with similar results. **e**, Flow cytometry on mCherry-*Magohb*-knock-in WT or MLL2-KO mESCs expressing shNT or shCXXC1. The experiment was repeated two times independently with similar results. **f**, An interleaved scatter plot of *Magohb* normalized RNA-seq counts in WT mESCs expressing shNT, MLL2-KO mESCs expressing shNT or shCXXC1 ($n=2$). **g**, An interleaved scatter plot of *Magohb* normalized RNA-seq counts in WT mESCs, MLL2ΔSET mESCs and TMutant mESCs ($n=2$). In **f** and **g**, individual values are represented as single dots. The middle line represents the median. The top and bottom lines depict the standard error of the mean. CPM, counts per million. Uncropped gels are available as source data.

Strikingly, *Wdr82* knockdown led to a substantial rescue of *Magohb* expression in MLL2-KO mESCs at the protein level (Fig. 2b) and the RNA level (Fig. 2c and Extended Data Fig. 3b). WDR82 is primarily found in two complexes: SET1A/B COMPASS implicated in H3K4me3 deposition^{2,9,29}; and the PNUITS/PP1 complex implicated in transcription termination of active enhancers³⁰. To pinpoint the specific function of WDR82 implicated in *Magohb*

transcriptional regulation, we targeted individual members of the SET1A/B COMPASS or PNUITS/PP1 complex via CRISPR sgRNAs. We found that sgRNAs targeting *Cxxc1* or *Set1a* and *Set1b* rescued *Magohb* expression in MLL2-KO cells (Extended Data Fig. 3c), suggesting that SET1A/B COMPASS members, rather than the PNUITS/PP1 complex, are implicated in *Magohb* transcription inhibition in the absence of MLL2.

While WDR82 is reported to be a subunit of two major complexes, CXXC1 has mainly been described as a subunit of SET1A/B COMPASS³¹, and thus, to target SET1A/B COMPASS specifically in the following experiments, we used a short hairpin RNA (shRNA) strategy against *Cxxc1* (shCXXC1; Fig. 2d). Following CXXC1 knockdown, both SET1A and SET1B protein levels were largely decreased, similar to previous findings after WDR82 knockdown³⁰, while in contrast, the MAGOHB protein level was considerably increased. Consistently, by flow cytometry, the population of MLL2-KO cells exhibited a shift in mCherry-MAGOHB fluorescence signal following CXXC1 knockdown (Fig. 2e), and *Magohb* expression rescue was recapitulated in our RNA-seq experiments (Fig. 2f). Finally, we validated the direct implication of SET1A/B in *Magohb* transcriptional inhibition by using mESCs knocked out for SET1B and deleted for the SET domain of SET1A and MLL2, referred to as 'TMutant' hereafter, previously generated in our laboratory (C.C.S. et al., manuscript in preparation). TMutant mESCs exhibited a marked increase of *Magohb* transcription compared to MLL2ΔSET cells, akin to effects observed in CXXC1 knockdown (Fig. 2g). Taken together, our results identify an underlying balance within the COMPASS family, where SET1A/B are required for inhibition of *Magohb* transcription in the absence of MLL2 in mESCs.

MLL2-KO transcriptional defects are rescued by SET1A/B COMPASS knockdown. We next extended our analysis genome wide to examine whether the transcriptional behavior of *Magohb* was specific or shared by a whole set of genes. We focused on genes bound by MLL2 after partitioning MLL2 and H3K4me3 ChIP-seq data into clusters, with 'cluster 1' containing loci bound by MLL2 and where H3K4me3 was strongly decreased in MLL2-KO compared to WT mESCs, representing 2,021 associated genes (Fig. 3a). Our analysis revealed that the majority of cluster-1-associated genes exhibited a strongly reduced transcription in MLL2-KO compared to WT mESCs (Fig. 3b and Extended Data Fig. 4a). Strikingly, the reduction in expression of cluster-1-associated genes was significantly rescued following CXXC1 knockdown for more than 82% of cluster-1-associated genes (Fig. 3b and Extended Data Fig. 4a,b). This indicates that the regulatory balance between MLL2 and SET1A/B complexes does not apply only to *Magohb*, but rather is shared among approximately 1,200 MLL2-bound genes. Notably, while *Magohb* expression was only partially rescued, the expression of a majority of genes in cluster 1 was fully rescued following CXXC1 knockdown (Extended Data Fig. 4b,c and Extended Data Fig. 1b), indicating that *Magohb* does not fully exemplify the bulk of MLL2-dependent genes and potentially explaining the low rate of confirmed hits from our CRISPR screen. Finally, our precision nuclear run-on sequencing (PRO-seq)³² analysis, which measures engaged RNA polymerase, indicated that the effect on gene expression could be imputed to active transcription since our results complemented our RNA-seq data (Fig. 3c). Note that the expression of cluster-2-associated genes is globally stable in all tested conditions, indicating that effects of CXXC1 knockdown are specific to cluster 1 and not a reflection of a global change in transcription.

We then analyzed the TMutant mESCs and observed that gene expression of cluster-1 genes was significantly rescued compared to that of MLL2ΔSET mESCs (Fig. 3d and Extended Data Figs. 1c and 4c,d), in agreement with the CXXC1-knockdown effects discussed earlier, further indicating that SET1A/B COMPASS negatively impact the transcription of cluster-1 genes in MLL2-KO mESCs. To further validate that the effect of CXXC1 knockdown was due to SET1A/B COMPASS activity loss, we performed epistasis experiments by knocking down CXXC1 in the TMutant cells. While TMutant cells with a shRNA control (shNT) rescued the expression of cluster-1 genes, the knockdown of CXXC1 on top of the TMutant cells did not have a global additive effect (Fig. 3d and Extended Data Fig. 4d). This indicates that the rescue of MLL2-dependent

transcription was specifically due to SET1A/B COMPASS function. Gene Ontology (GO) analysis of cluster-1 genes that were at least partially rescued by CXXC1 knockdown showed enrichment for genes involved in 'protein localization to cell periphery' and 'regulation of neuron projection development' (Fig. 3e). To conclude, our studies demonstrate that in the absence of MLL2, SET1A/B COMPASS are necessary to maintain the transcriptional inhibition of ~1,200 MLL2-dependent genes.

Transcription in the absence of H3K4me3 in MLL2-KO CXXC1-knockdown cells. Previous studies demonstrated that CXXC1 is required for deposition of H3K4me3 at many promoters genome wide¹⁰. Therefore, we decided to further analyze H3K4me3 changes in our conditions. Unlike for cluster 2, where the H3K4me3 level was globally unchanged between WT and MLL2-KO cells, there was a substantial decrease of H3K4me3 levels in cluster-1 gene promoters (Fig. 4a,b and Extended Data Fig. 1d,e). This pattern was expected since cluster-1-associated genes are downregulated in MLL2-KO mESCs and reduction of transcription is generally correlated with a decrease of H3K4me3 levels around the transcription start site (TSS)³³. However, H3K4me3 remained low in cluster 1 following CXXC1 knockdown, despite the strong rescue in gene expression (Figs. 3b-d and 4a,b). Thus, in MLL2-KO shCXXC1 cells, we observed an uncoupling of H3K4me3 and transcription. We conclude that increasing H3K4me3 is not required to increase transcription under these conditions, thereby challenging the existing notion of an instructional role of H3K4me3 for transcription, at least on these loci.

MLL2/COMPASS functions to counteract DNA methylation inhibition of transcription. We next addressed the question of how CXXC1 knockdown rescues MLL2-dependent transcriptional loss in MLL2-KO mESCs. Since CXXC1-KO mESCs were previously shown to exhibit reduced DNA methyltransferase 1 (DNMT1) expression^{34,35}, we investigated the impact of CXXC1 knockdown on DNA methylation in the context of MLL2-KO mESCs. We found that *Dnmt1* transcription was strongly reduced following CXXC1 knockdown in MLL2-KO mESCs (Fig. 5a and Extended Data Fig. 1f). Interestingly, *Dnmt1* expression was correlated with SET1A (subunit of COMPASS) binding at this locus (Fig. 5a): a strong SET1A peak was observed in WT and MLL2-KO mESCs, while this SET1A peak was highly reduced following CXXC1 knockdown at the *Dnmt1* promoter. Immunoblotting showed that reduced *Dnmt1* transcription was reflected at the protein level (Extended Data Fig. 5a). Notably, *Dnmt3a* expression was decreased in MLL2-KO mESCs compared to WT mESCs, although its expression appears to be unrelated to SET1A expression (Extended Data Figs. 1g and 5a,b). Overall, these data indicate that SET1A is necessary for *Dnmt1* transcription and suggest that this regulation could be through direct binding of SET1A/COMPASS to the *Dnmt1* promoter.

Next, to determine whether the reduction in DNMT1 expression was sufficient to recapitulate CXXC1 knockdown, we directly targeted DNMT1 in MLL2-KO mESCs (Extended Data Fig. 5c). Our genome-wide analyses indicated that DNMT1 knockdown resulted in a partial rescue of the expression of cluster-1 genes (Fig. 5b and Extended Data Fig. 5c). DNMT1 transfers methyl groups to cytosine nucleotides and its KO has been shown to significantly reduce DNA methylation³⁶. Thus, we then directly targeted DNA methylation using a widely used DNA-demethylating agent³⁷, the cytosine nucleoside analog 5-aza-2'-deoxycytidine (5dAza), and observed a significant rescue of cluster-1 transcription in MLL2-KO mESCs (Fig. 5b and Extended Data Fig. 5c). Notably, 5dAza treatment led to a higher rescue than the knockdown of DNMT1, indicating either that the knockdown efficiency of DNMT1 was not sufficient to recapitulate CXXC1 knockdown or that other actors regulating DNA methylation were implicated.

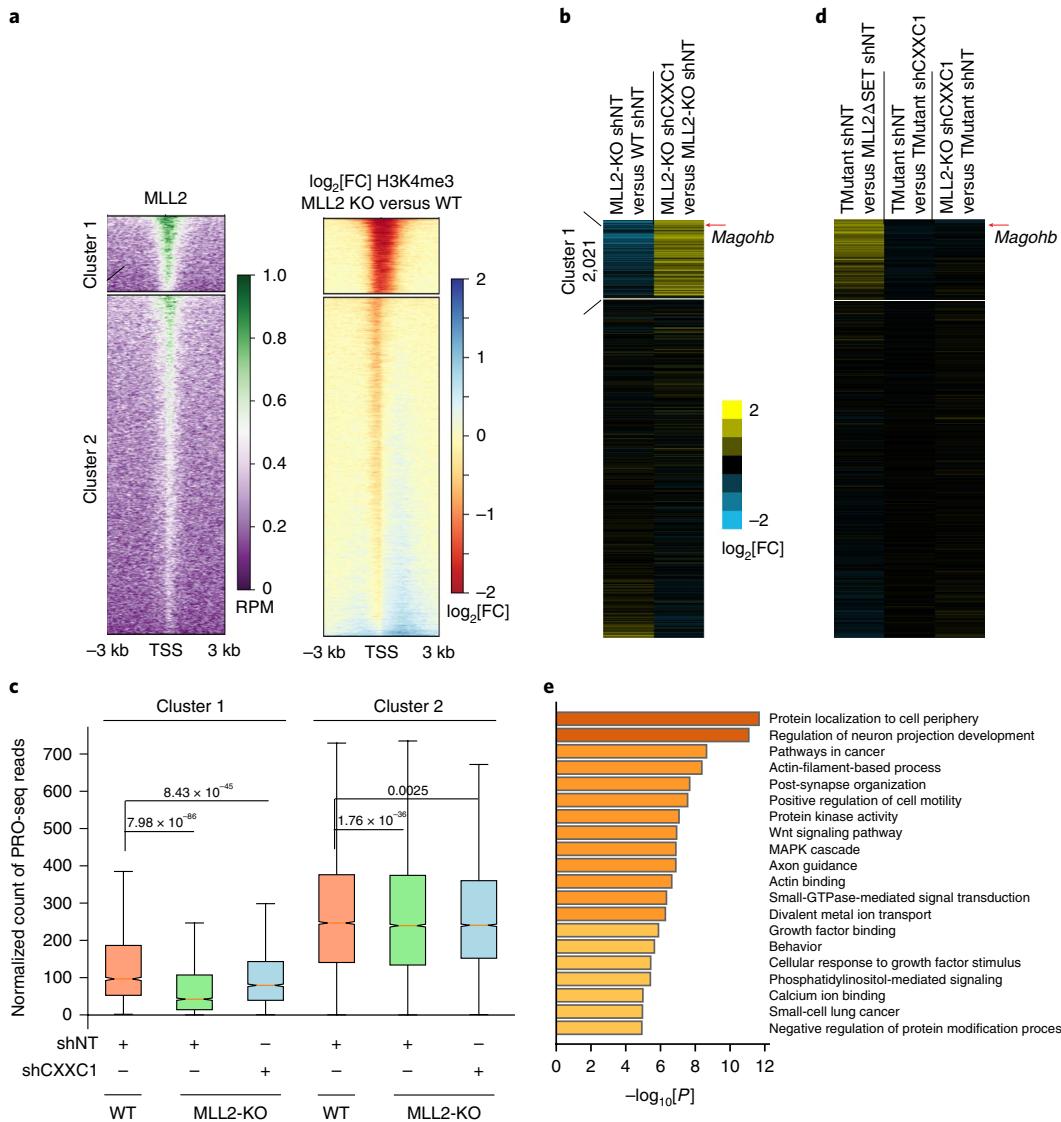


Fig. 3 | MLL2-KO transcriptional defects are rescued by SET1A/B knockdown. **a**, Left panel: MLL2 occupancy in WT cells in reads per million (RPM). Note that the ChIP-seq signal found in MLL2-KO cells has been subtracted for analysis. The profile is centered on the TSS (± 3 kb). Right panel: $\log_2[\text{FC}]$ of the H3K4me3 ChIP-seq signal in MLL2-KO mESCs compared to WT mESCs ($n=2$). k -means clustering was used to partition all genes with RNA polymerase II signal in WT mESCs by their H3K4me3 $\log_2[\text{FC}]$ following MLL2 KO and by their MLL2 occupancy. **b**, Heatmaps showing the corresponding $\log_2[\text{FC}]$ in gene expression in MLL2-KO shNT mESCs compared to WT shNT mESCs and MLL2-KO shCXXC1 mESCs compared to MLL2-KO shNT mESCs. Cluster 1 is highlighted. **c**, Box plots quantifying the PRO-seq signal around the TSS (± 3 kb). The box plots indicate the median (middle line), the third and first quartiles (box) and the first and fourth quartiles (error bars). The P values were computed using the Wilcoxon test (two-sided); cluster 1, $n=2,357$; cluster 2, $n=10,605$. The experiment was repeated two times independently with similar results. **d**, Heatmaps showing the corresponding $\log_2[\text{FC}]$ in gene expression in TMutant shNT mESCs versus MLL2 Δ SET shNT mESCs, TMutant shNT mESCs versus TMutant shCXXC1 mESCs, and MLL2-KO shCXXC1 mESCs versus TMutant shNT mESCs. **e**, GO terms for genes at least partially rescued following CXXC1 knockdown in MLL2-KO mESCs compared to MLL2-KO mESCs. The P values were computed using Metascape, which utilizes the hypergeometric test and Benjamini-Hochberg P -value correction algorithm; $n=942$.

To further investigate the relationship between MLL2-dependent transcription and DNA methylation, we performed modified reduced representation bisulfite sequencing (mRRBS) to map CpG methylation genome wide³⁸. Our analysis demonstrated that DNA methylation was slightly decreased for cluster-2 genes in MLL2-KO compared to WT mESCs, presumably reflecting the reduced DNMT3A expression in these cells (Extended Data Fig. 5e). Interestingly, we found that cluster-1 promoter regions exhibited two distinct patterns of DNA methylation changes, with one sub-cluster (41 genes including *Magohb*) exhibiting a strong increase of DNA methylation following MLL2 KO (more than 25% increase of CpG methylation) that was rescued by CXXC1 knockdown (Fig. 5c).

These results suggested that, for these genes, loss of DNA methylation was sufficient to restore MLL2-dependent transcription. To test this hypothesis, we next performed targeted DNA demethylation recruitment using a dCas9-Tet1 catalytic domain (CD) fusion³⁹. In MLL2-KO mESCs, targeting Tet1CD to the *Magohb* promoter led to a marked increase of MAGOHB expression, while targeting a dead Tet1CD to the promoter of *Magohb* or targeting a region 5 kilobases (kb) upstream of the *Magohb* promoter did not affect MAGOHB expression, as assessed by flow cytometry (Fig. 5d and Extended Data Fig. 5f). These results confirmed a direct repressive role of DNA methylation for approximately 40 genes, including *Magohb*, in MLL2-KO cells that can be rescued by CXXC1

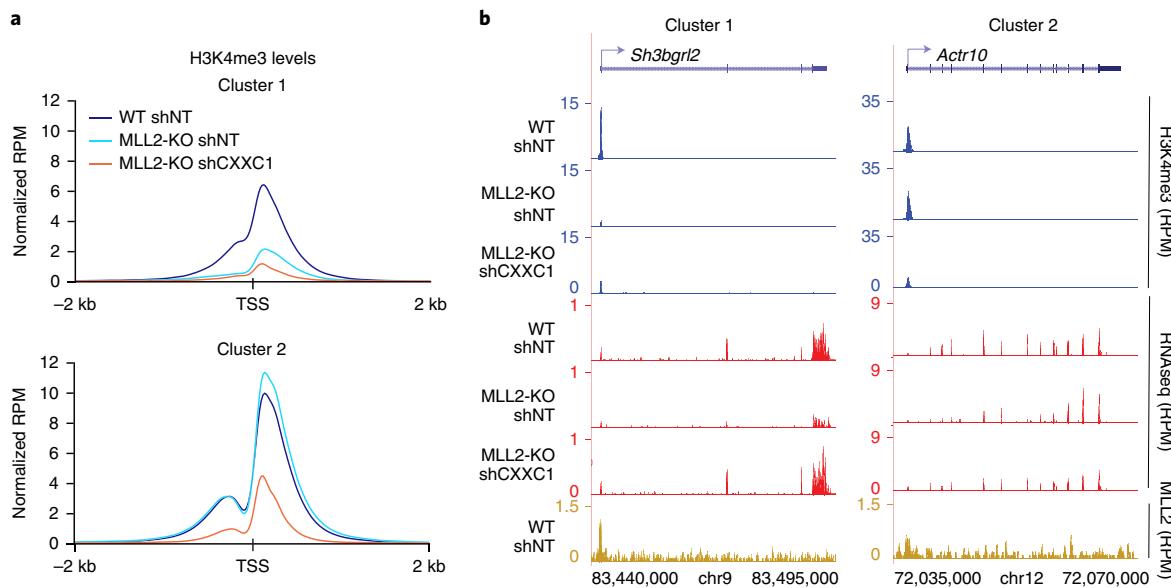


Fig. 4 | H3K4me3 levels at MLL2-dependent genes are not rescued by CXXC1 knockdown. **a**, Metaplots comparing the H3K4me3 levels at the clusters defined in Fig. 3a. The plots are centered on the TSS (± 2 kb). The experiment was repeated two times independently with similar results. **b**, H3K4me3 ChIP-seq, RNA-seq and MLL2 ChIP-seq track examples for each cluster in reads per million.

knockdown. Notably, the rest of the cluster-1 genes, representing the vast majority of the cluster, exhibited low levels of DNA methylation on their promoter regions and their DNA methylation levels remained globally constant among all three conditions, suggesting the existence of a third actor implicated in the repression of these MLL2-dependent genes in MLL2-KO cells.

DNA methylation and H3K27me3 deposition are counteracted by MLL2. We hypothesized that H3K27me3, deposited by Polycomb repressive complex 2 (PRC2) and classically linked to transcriptional silencing⁴⁰, might be at the basis of this repression. Our H3K27me3 ChIP-seq analyses demonstrated that the H3K27me3 signal was indeed increased around the TSS of MLL2-dependent genes following MLL2 KO (Fig. 6a), while in contrast, the H3K27me3 signal was rescued following CXXC1 knockdown. To test whether this effect was directly associated with gene expression changes, we next deleted the *Suz12* gene (which encodes one of the four core proteins of PRC2) in MLL2-KO mESCs to generate double-mutant MLL2- and SUZ12-KO cells that lose H3K27me3 (Fig. 6b). Our RNA-seq analyses revealed that SUZ12 KO was sufficient to significantly rescue the loss of gene expression in the MLL2-KO cells in all of the tested MLL2- and SUZ12-KO clones (Fig. 6c,d and Extended Data Figs. 1h and 6a). When comparing the effect of CXXC1 knockdown, 5dAza treatment and SUZ12 KO on MLL2-dependent gene rescue in MLL2-KO cells, we observed a large overlap among the genes at least partially rescued (Fig. 6e).

These data point toward two different mechanisms of regulation for MLL2-dependent genes: for 2% of MLL2-dependent genes, DNA methylation directly represses genes expression in the absence of MLL2; in contrast, for the vast majority of MLL2-dependent genes (98%), it is H3K27me3 that represses those genes in the absence of MLL2. Unexpectedly, this larger set of genes repressed by H3K27me3 in the absence of MLL2 can also be rescued by DNA methylation removal. We hypothesized that this overlap is due to the fact that DNA methylation impacts on H3K27me3 deposition, as previously reported⁴¹, and thus, explains why either removal of H3K27me3 or 5dAza treatment can rescue most of the MLL2-dependent gene expression in MLL2-KO mESCs.

An MLL2, H3K27me3 and DNA methylation balance regulates developmental gene expression. To test our model proposed above, we next assessed the effect of 5dAza treatment on H3K27me3 deposition (Fig. 7a,b and Extended Data Fig. 6b). First, we confirmed the efficiency of our 5dAza treatment on DNA methylation reduction by mRRBS experiments: 5dAza treatment led to a significant decrease of DNA methylation for both subclusters of MLL2-dependent genes (Fig. 7a). In fact, for the genes with a more than 25% increase in DNA methylation (as defined in Fig. 5c; 2% of MLL2-dependent genes), we observed up to a 4.5-fold reduction of DNA methylation following 5dAza treatment, confirming that DNA methylation directly regulates the expression of those genes. For the second subcluster (98% of MLL2-dependent genes), we observed no major changes in DNA methylation in all conditions tested. In fact, our H3K27me3 ChIP-seq analyses showed an increase of H3K27me3 levels in MLL2-KO mESCs compared to WT mESCs, while in contrast, 5dAza treatment led to a substantial decrease of H3K27me3 signal (Fig. 7b and Extended Data Fig. 6b). These results confirmed that altering DNA methylation affects H3K27me3 deposition. Taken together, our results establish that for most MLL2-dependent genes, removal of H3K27me3 or 5dAza treatment can restore MLL2-dependent gene expression in the absence of MLL2 through direct loss of H3K27me3 or through indirect dilution of H3K27me3.

Next, we explored the conservation of our proposed molecular mechanism of MLL2-dependent gene regulation during differentiation. WT, MLL2-KO, MLL2-KO 5dAza-treated, and double-mutant MLL2- and SUZ12-KO mESCs were differentiated into embryoid bodies (EBs) for 6d and RNA was collected for sequencing. Our analysis demonstrated that up to 434 genes were downregulated in MLL2-KO versus WT EBs, and rescued in both MLL2-KO 5dAza-treated and double-mutant MLL2- and SUZ12-KO EB conditions (Fig. 7c). The fact that DNA methylation reduction or loss of H3K27me3 both led to this rescue indicated that our presented molecular mechanism for MLL2-dependent transcriptional regulation, implicating COMPASS, H3K27me3 and DNA methylation, was at least partially conserved during differentiation. Moreover, 5dAza treatment led to a larger rescue than SUZ12 KO, in accordance with our model where 5dAza treatment affects DNA methylated sites as well

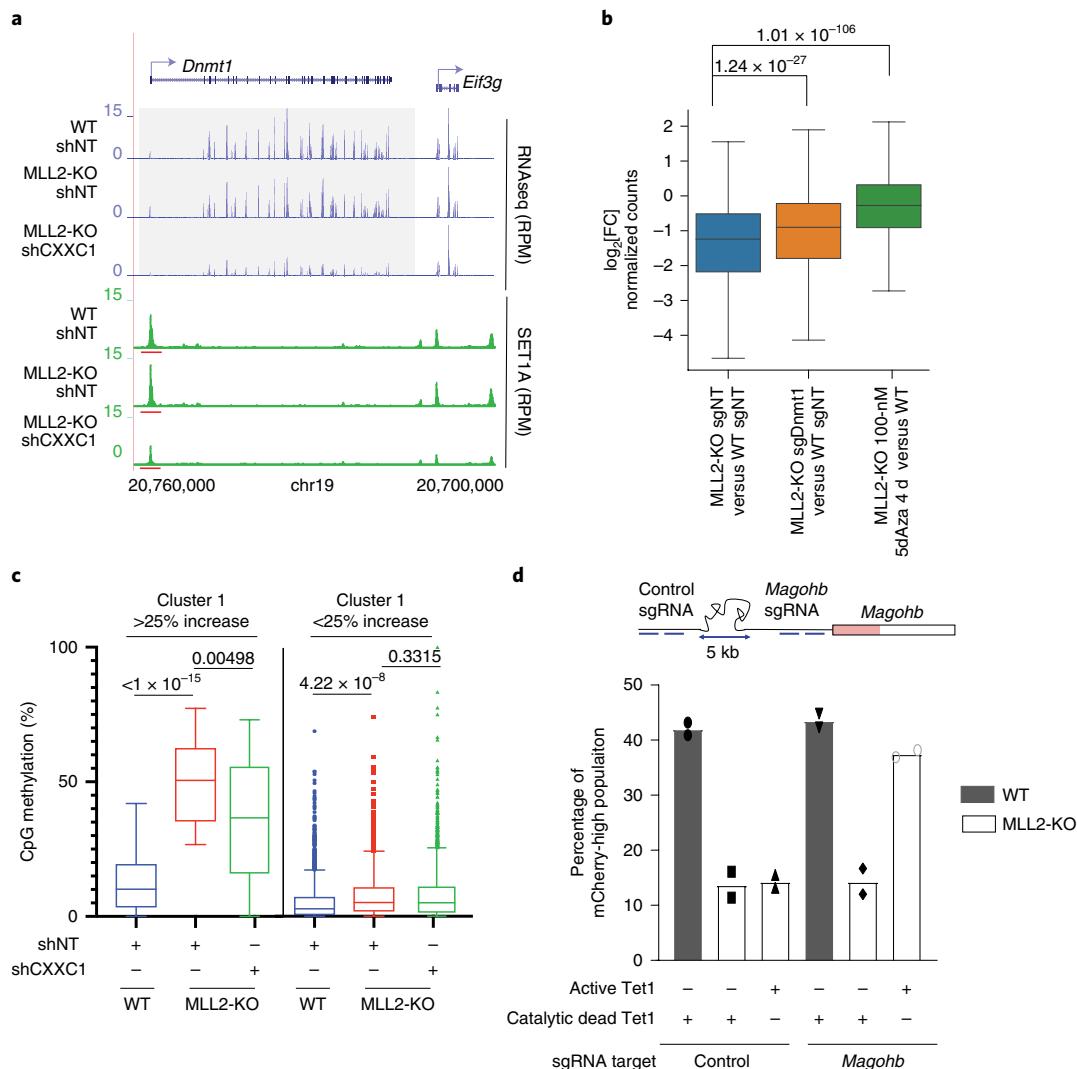


Fig. 5 | Inhibition of DNA methylation is sufficient to restore cluster-1 gene expression in MLL2-KO mESCs. **a**, RNA-seq and SET1A ChIP-seq tracks for *Dnmt1*. The experiment was repeated two times independently with similar results. **b**, Box-plot analysis of cluster-1 gene expression in MLL2-KO versus WT mESCs, MLL2-KO mESCs expressing a sgRNA targeting *Dnmt1* versus WT mESCs expressing a control sgRNA, and MLL2-KO mESCs treated for 4 d with 100-nM 5dAza versus WT mESCs. The box plots indicate the median (middle line), the third and first quartiles (box) and the first and fourth quartiles (error bars). The *P* values were computed using the Wilcoxon test (two-sided), $n=2,021$. **c**, A box-and-whisker plot quantifying the changes in CpG methylation around the TSS ($\pm 1\text{kb}$) for cluster 1, which was subclustered into genes with a more or a less than 25% increase in DNA methylation in MLL2-KO versus WT cells ($n=3$ biological replicates). The *P* values were calculated using a Brown-Forsythe and Welch analysis of variance (ANOVA) test assuming unequal variances with the Dunnett T3 correction for multiple comparisons (two-sided). **d**, The percentage of mCherry-high cells was calculated by flow cytometric analyses 2 d post infection and is shown as the mean percentage of mCherry-high cells of two biological replicates.

as H3K27me3 deposition, while SUZ12 KO affects only H3K27me3 sites. Finally, our gene set enrichment analysis (GSEA) on this specific set of genes, comparing MLL2-KO to WT EBs, showed that genes associated with spermatogenesis were among the most downregulated genes (Fig. 7d), in accordance with previously reported data⁴². Interestingly, 5dAza treatment or SUZ12 KO was sufficient to rescue that same gene set associated with spermatogenesis (Fig. 7d).

In summary, our results demonstrate the existence of a molecular interplay among DNA methylation, H3K27me3 and MLL2-dependent function in transcription. In the absence of MLL2, SET1A/B are required to inhibit the transcription of a cluster of ~1,200 genes normally bound by MLL2. This inhibition can be rescued by inhibiting DNA methylation or H3K27me3 deposition, and unexpectedly, this transcriptional reactivation is not linked to the re-establishment of H3K4me3. These data suggest that the function of MLL2 at these loci is to overcome or repel transcriptional

silencing exerted by H3K27me3 and/or DNA methylation, and that in the absence of these inhibitory marks, transcription can proceed without H3K4me3. Finally, we demonstrate that prevention of active repression by MLL2 and not H3K4me3 underlies transcriptional regulation on MLL2 targets and that this molecular mechanism is at least partially conserved during differentiation.

Discussion

Our study identifies a regulatory circuit among the COMPASS family, DNA methylation machinery and the Polycomb-deposited H3K27me3, which becomes apparent in the absence of MLL2 (a member of the COMPASS family). Our genome-scale screen indicated that depletion of SET1A/B in MLL2-KO mESCs was sufficient to rescue the loss of expression of ~1,200 MLL2-dependent genes. To our surprise, this rescue in gene expression as the result of SET1A/B loss in MLL2-KO cells did not require H3K4me3. These findings

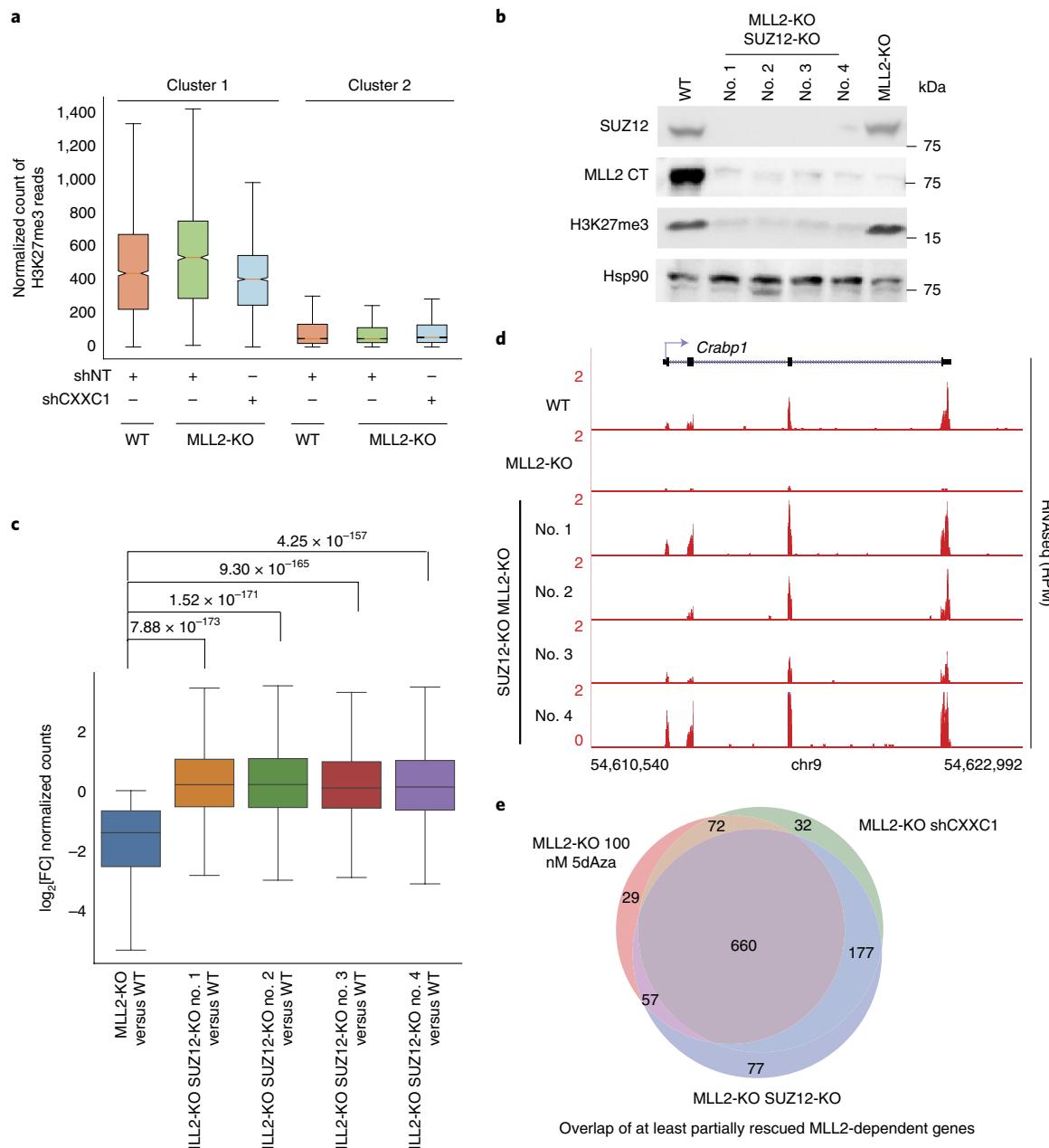


Fig. 6 | Interplay of H3K27me3, DNA methylation and MLL2 for gene transcription. **a**, Box plots quantifying the H3K27me3 ChIP-seq signal around the TSS (± 3 kb; $n=2$). **b**, A western blot of the SUZ12, MLL2, H3K27me3 and HSP90 (loading control) protein levels in WT, double-mutant MLL2- and SUZ12-KO (4 clones) and MLL2-KO cells. The experiment was repeated two times independently with similar results. **c**, Box-plot analysis of cluster-1 gene expression in MLL2-KO versus WT mESCs, and double-mutant MLL2- and SUZ12-KO versus WT mESCs (clones 1 to 4). The P values were computed using the Wilcoxon test (two-sided), $n=2,021$. The box plots indicate the median (middle line), the third and first quartiles (box) and the first and fourth quartiles (error bars). **d**, RNA-seq track examples for each clone. The experiment was repeated two times independently with similar results. **e**, A Venn diagram of RNA-seq comparing double-mutant MLL2- and SUZ12-KO, MLL2-KO shCXXC1 or MLL2-KO 5dAza-treated versus MLL2-KO cells showing the overlap of partially rescued genes (less than WT expression) among those conditions. Uncropped gels are available as source data.

are in accordance with the growing body of evidence interrogating the existence of an instructional role for H3K4 methylation in active transcription^{10,25,33,43,44}. Deletion of the only H3K4 methyltransferase in yeast, *SET1*, had no extensive transcriptional alteration effect⁴³. Accordingly, deletion of *CPS40* (*SPP1*) or *CPS35* (*SWD2*), resulting in severe loss of H3K4me3, had no effects on coding gene transcription⁴³. Similarly, in mammals, loss of H3K4me3 at active genes had only a very limited effect on gene expression⁴⁴ or transcriptional induction^{25,44}.

Our study has uncovered that affecting global DNA methylation or altering the DNA methylation of targeted regions was sufficient to restore MLL2-dependent transcriptional loss in MLL2-KO mESCs. Similarly, several recent studies using CRISPR-Cas9-based systems to modulate DNA methylation have demonstrated that targeted DNA demethylation of methylated promoters is associated with gene re-expression^{45–47}. The mechanism by which the removal of DNA methylation is sufficient to induce transcription remains an intriguing question and determining the factors that are recruited

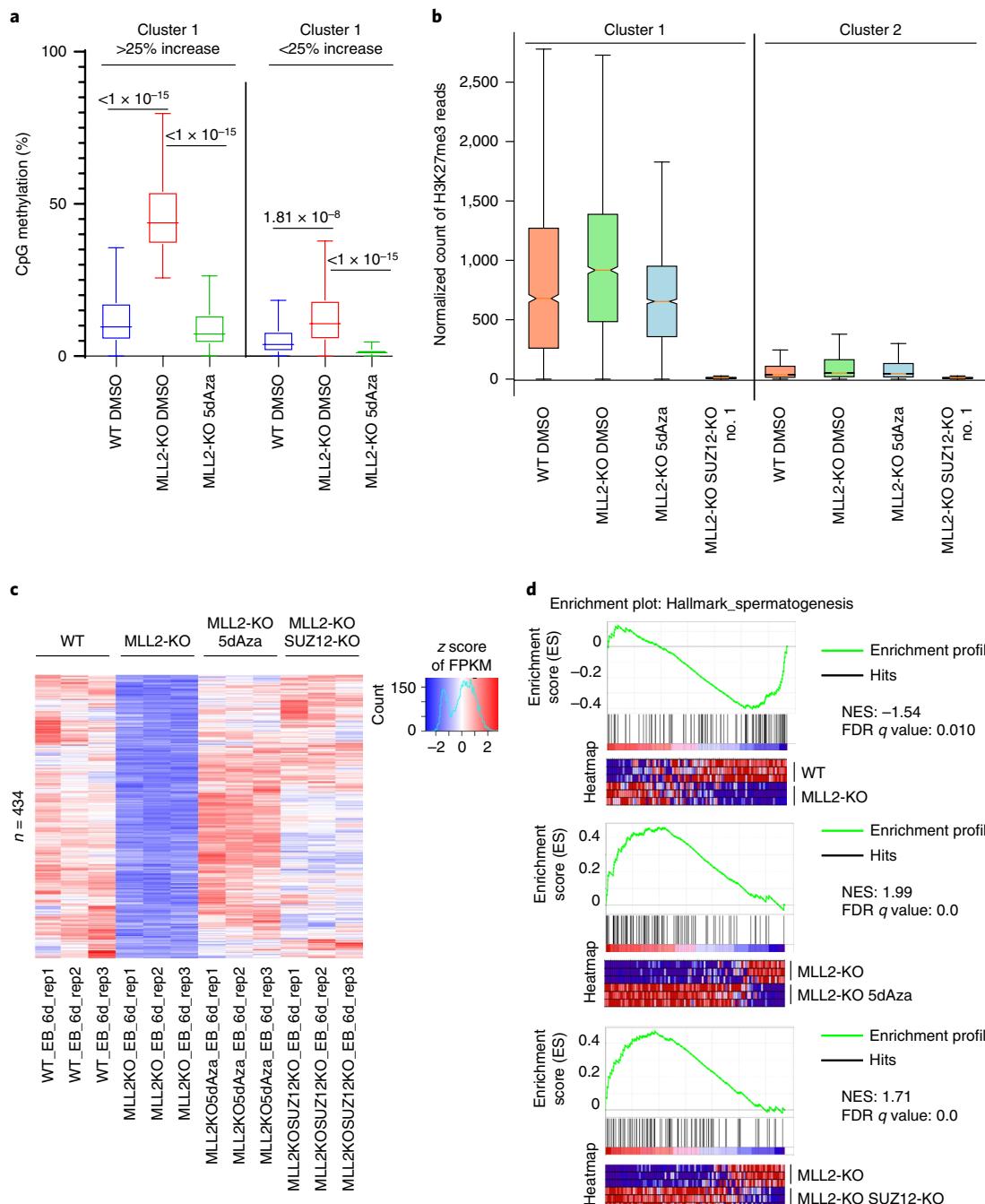


Fig. 7 | Interplay of H3K27me3, DNA methylation and MLL2 and its impact on MLL2-KO differentiation defects. **a**, A box-and-whisker plot quantifying the changes in CpG methylation around the TSS ($\pm 1\text{kb}$) for cluster 1, which was subclustered into genes with a more or a less than 25% increase in DNA methylation in MLL2-KO versus WT cells ($n=3$ biological replicates). The P values were calculated using a Brown-Forsythe and Welch ANOVA test assuming unequal variances with the Dunnett T3 correction for multiple comparisons (two-sided). The box plots indicate the median (middle line), the third and first quartiles (box) and the first and fourth quartiles (error bars). **b**, Box plots quantifying the H3K27me3 ChIP-seq signal around the TSS ($\pm 3\text{ kb}$; $n=2$). **c**, A heatmap of gene expression in WT, MLL2-KO, MLL2-KO 5dAza-treated and double-mutant MLL2- and SUZ12-KO EBs at day 6. Differentially expressed genes were selected for gene downregulation in MLL2-KO EBs compared to WT EBs and rescued following 5dAza treatment as well as SUZ12 KO. **d**, GSEA of genes downregulated between WT versus MLL2-KO EBs, MLL2-KO versus MLL2-KO 5dAza-treated EBs and MLL2-KO versus double-mutant MLL2- and SUZ12-KO EBs at day 6 (normalized enrichment score (NES); false discovery rate-adjusted P value (FDR q value)), $n=434$. The P values were computed using GSEA that utilizes an empirical phenotype-based permutation test procedure (two-sided).

following DNA demethylation for gene activation will be required to better understand this *cis*-regulatory mechanism.

To this end, we found that, to some extent, the loss of H3K27me3 mimics the phenotype of DNA methylation loss for MLL2-dependent gene expression in MLL2-KO mESCs, indicating

that DNA methylation or H3K27me3 can inhibit MLL2-dependent gene transcription in the absence of MLL2. This is particularly interesting considering the fact that an intact DNA methylome was shown to be required for proper H3K27me3 deposition⁴¹: at the CpG islands that contain a high level of H3K27me3, loss of

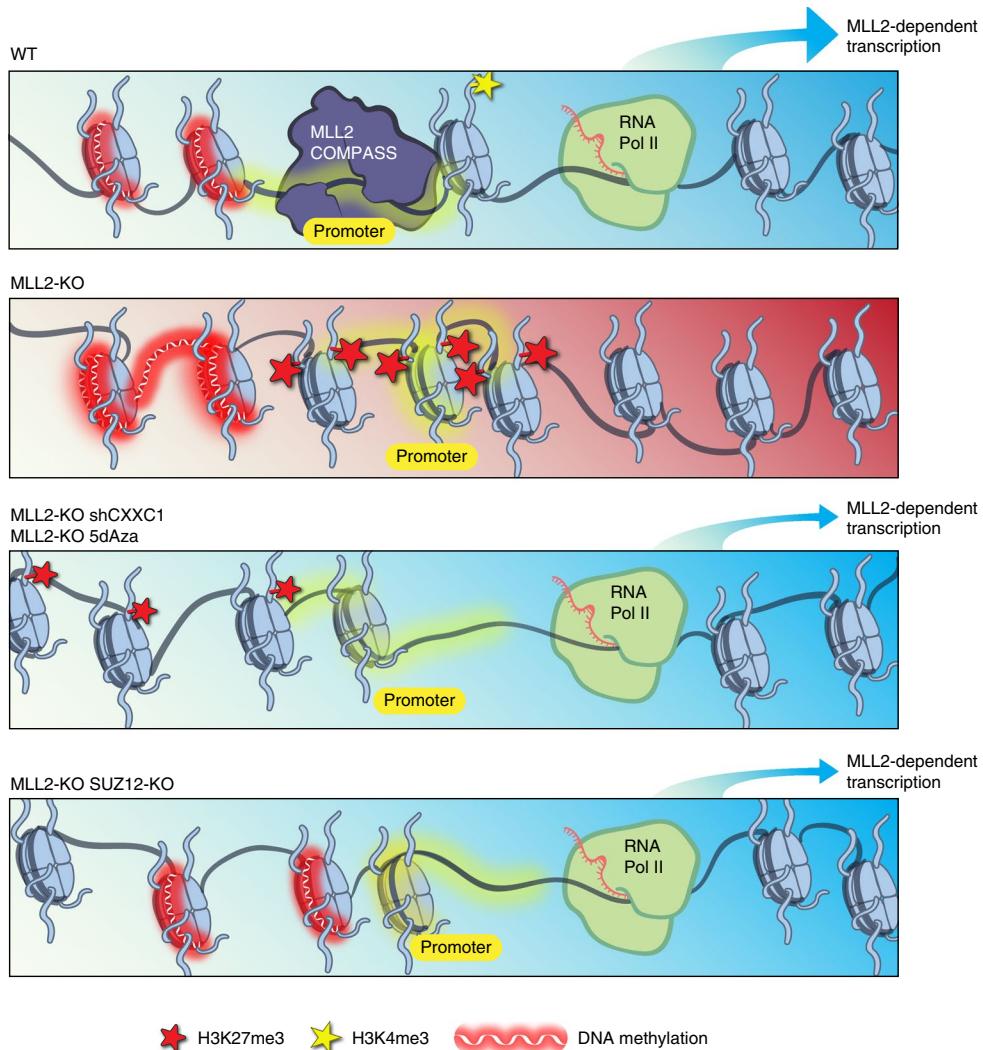


Fig. 8 | Model for an epigenetic equilibrium among Polycomb, COMPASS and DNA methylation machineries at most MLL2-dependent genes. In WT mESCs (top panel), MLL2/COMPASS binds to the promoters of MLL2-dependent genes and deposits H3K4me3 on these promoter regions. In the absence of MLL2 (second panel, MLL2 KO), H3K27me3 is deposited on the regions that are now free of MLL2, while DNA methylation remains for most of the genes at the shore of H3K27me3 marks. When DNA methylation is lost (third panel), either by targeting CXXC1, which affects *Dnmt1* expression, or by 5dAza treatment (MLL2-KO shCXXC1 or MLL2-KO 5dAza), H3K27me3 marks are spread to these newly CpG hypomethylated DNA regions, and consequently, the H3K27me3 signal previously observed on the promoter regions of MLL2-dependent genes is diluted and becomes insufficient to repress the expression of these genes. Similarly, after KO of SUZ12 in MLL2-KO mESCs (bottom panel, double-mutant MLL2 and SUZ12 KO), MLL2-dependent genes are transcribed in the absence of MLL2 owing to the absence of the repressive H3K27me3 mark.

DNA methylation (triple-KO (*Dnmt1kd,3a-/-,3b-/-*) mESCs) causes a concomitant decrease of H3K27me3 by dilution of this mark. According to this model, in the absence of DNA methylation, PRC2 activity can access newly unmethylated CpG sites, and therefore, H3K27me3 is redistributed in the genome, which results in de-repression of PRC2 targets. Our data show that the vast majority of genes that are rescued by inhibition of DNA methylation can also be rescued by PRC2 inhibition (Figs. 6e and 8). These results are in agreement with DNA methyltransferase inhibition inducing a reduction of H3K27me3 by dilution, as confirmed by our H3K27me3 ChIP-seq and bisulfite sequencing experiments (Fig. 7a,b). Therefore, inhibition of DNA methyltransferases can activate the transcription of MLL2-dependent genes by indirect reduction of H3K27me3 acquired following MLL2 loss, thus explaining how perturbations of mutually exclusive epigenetic marks can end up affecting the same set of genes. Interestingly, loss of H3K27me3 or reduction of DNA methylation was able to rescue the most

downregulated gene set ('hallmark_spermatogenesis') of MLL2-KO EBs at day 6 (Fig. 7d). This indicates that the molecular mechanism uncovered in this study for MLL2-dependent transcriptional regulation in normal growth conditions is at least partially conserved during differentiation. This is particularly interestingly in regard to the fact that MLL2 was recently shown to be mutated in early-onset dystonia^{20,21}. Our finding that the activity of MLL2-dependent genes can be modulated by DNA methylation and H3K27me3 raises the possibility that MLL2 haploinsufficiency could be potentially rescued by inhibiting DNA methylation and/or PRC2 activity.

Finally, *Magohb* encodes a core member of the exon-junction complex that is redundant with *Magoh* gene function²⁸. Interestingly, a recent paper determined that hemizygous loss of MAGOH is observed in a large number of tumors, establishing MAGOHB as a promising therapeutic target in this context²⁷. This unexpected dependency of cancer cells to paralog genes during tumor growth emphasizes the importance of performing genome-wide CRISPR

screens to gain molecular insights into context-dependent transcription regulation.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-020-0618-1>.

Received: 12 March 2019; Accepted: 26 March 2020;

Published online: 11 May 2020

References

- Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
- Meeks, J. J. & Shilatifard, A. Multiple roles for the MLL/COMPASS family in the epigenetic regulation of gene expression and in cancer. *Annu. Rev. Cancer Biol.* **1**, 425–446 (2017).
- Shilatifard, A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu. Rev. Biochem.* **81**, 65–95 (2012).
- Miller, T. et al. COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc. Natl Acad. Sci. USA* **98**, 12902–12907 (2001).
- Schneider, J. et al. Molecular regulation of histone H3 trimethylation by COMPASS and the regulation of gene expression. *Mol. Cell* **19**, 849–856 (2005).
- Rougue, A. et al. The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J.* **20**, 7137–7148 (2001).
- Krogan, N. J. et al. COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. *J. Biol. Chem.* **277**, 10753–10755 (2002).
- Mohan, M. et al. The COMPASS family of H3K4 methylases in *Drosophila*. *Mol. Cell. Biol.* **31**, 4310–4318 (2011).
- Wu, M. et al. Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. *Mol. Cell. Biol.* **28**, 7337–7344 (2008).
- Clouaire, T. et al. Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes Dev.* **26**, 1714–1728 (2012).
- Morgan, M. A. & Shilatifard, A. Chromatin signatures of cancer. *Genes Dev.* **29**, 238–249 (2015).
- Hu, D. et al. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Mol. Cell. Biol.* **33**, 4745–4754 (2013).
- Herz, H. M. et al. Enhancer-associated H3K4 monomethylation by Trithorax-related, the *Drosophila* homolog of mammalian Mll3/Mll4. *Genes Dev.* **26**, 2604–2620 (2012).
- Hu, D. et al. Not all H3K4 methylations are created equal: Mll2/COMPASS dependency in primordial germ cell specification. *Mol. Cell* **65**, 460–475.e6 (2017).
- Rickels, R. et al. An evolutionary conserved epigenetic mark of Polycomb response elements implemented by Trx/MLL/COMPASS. *Mol. Cell* **63**, 318–328 (2016).
- Tkachuk, D. C., Kohler, S. & Cleary, M. L. Involvement of a homolog of *Drosophila* trithorax by 11q23 chromosomal translocations in acute leukemias. *Cell* **71**, 691–700 (1992).
- Gu, Y. et al. The t(4;11) chromosome translocation of human acute leukemias fuses the *ALL-1* gene, related to *Drosophila* trithorax, to the *AF-4* gene. *Cell* **71**, 701–708 (1992).
- Djabali, M. et al. A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias. *Nat. Genet.* **4**, 431 (1993).
- Glaser, S. et al. The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis. *Epigenetics Chromatin* **2**, 5 (2009).
- Meyer, E. et al. Mutations in the histone methyltransferase gene *KMT2B* cause complex early-onset dystonia. *Nat. Genet.* **49**, 223–237 (2017).
- Zech, M. et al. Haplousinsufficiency of *KMT2B*, encoding the lysine-specific histone methyltransferase 2B, results in early-onset generalized dystonia. *Am. J. Hum. Genet.* **99**, 1377–1387 (2016).
- Sze, C. C. et al. Histone H3K4 methylation-dependent and -independent functions of Set1A/COMPASS in embryonic stem cell self-renewal and differentiation. *Genes Dev.* **31**, 1732–1737 (2017).
- 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).
- Rickels, R. et al. Histone H3K4 monomethylation catalyzed by Trx and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. *Nat. Genet.* **49**, 1647–1653 (2017).
- Hu, D. et al. The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1093–1097 (2013).
- Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
- Viswanathan, S. R. et al. Genome-scale analysis identifies paralog lethality as a vulnerability of chromosome 1p loss in cancer. *Nat. Genet.* **50**, 937–943 (2018).
- Singh, K. K., Wachsmuth, L., Kulozik, A. E. & Gehring, N. H. Two mammalian MAGOH genes contribute to exon junction complex composition and nonsense-mediated decay. *RNA Biol.* **10**, 1291–1298 (2013).
- Lee, J. H., Tate, C. M., You, J. S. & Skalnik, D. G. Identification and characterization of the human Set1B histone H3-Lys⁴ methyltransferase complex. *J. Biol. Chem.* **282**, 13419–13428 (2007).
- Austenaa, L. M. et al. Transcription of mammalian *cis*-regulatory elements is restrained by actively enforced early termination. *Mol. Cell* **60**, 460–474 (2015).
- Lee, J. H. & Skalnik, D. G. CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys⁴ methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J. Biol. Chem.* **280**, 41725–41731 (2005).
- Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950–953 (2013).
- Howe, F. S., Fischl, H., Murray, S. C. & Mellor, J. Is H3K4me3 instructive for transcription activation? *Bioessays* **39**, 1–12 (2017).
- Tate, C. M., Lee, J. H. & Skalnik, D. G. CXXC finger protein 1 restricts the Setd1A histone H3K4 methyltransferase complex to euchromatin. *FEBS J.* **277**, 210–223 (2010).
- Carlone, D. L. et al. Reduced genomic cytosine methylation and defective cellular differentiation in embryonic stem cells lacking CpG binding protein. *Mol. Cell. Biol.* **25**, 4881–4891 (2005).
- Li, Z. et al. Distinct roles of DNMT1-dependent and DNMT1-independent methylation patterns in the genome of mouse embryonic stem cells. *Genome Biol.* **16**, 115 (2015).
- Christman, J. K. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene* **21**, 5483–5495 (2002).
- McGrath-Morrow, S. A. et al. DNA methylation regulates the neonatal CD4⁺ T-cell response to pneumonia in mice. *J. Biol. Chem.* **293**, 11772–11783 (2018).
- Liu, X. S. et al. Rescue of fragile X syndrome neurons by DNA methylation editing of the FMR1 gene. *Cell* **172**, 979–992.e6 (2018).
- Piunti, A. & Shilatifard, A. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science* **352**, aad9780 (2016).
- Brinkman, A. B. et al. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res.* **22**, 1128–1138 (2012).
- Mas, G. et al. Promoter bivalency favors an open chromatin architecture in embryonic stem cells. *Nat. Genet.* **50**, 1452–1462 (2018).
- Margaritis, T. et al. Two distinct repressive mechanisms for histone 3 lysine 4 methylation through promoting 3'-end antisense transcription. *PLoS Genet.* **8**, e1002952 (2012).
- Clouaire, T., Webb, S. & Bird, A. Cfp1 is required for gene expression-dependent H3K4 trimethylation and H3K9 acetylation in embryonic stem cells. *Genome Biol.* **15**, 451 (2014).
- Xu, X. et al. A CRISPR-based approach for targeted DNA demethylation. *Cell Discov.* **2**, 16009 (2016).
- Morita, S. et al. Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nat. Biotechnol.* **34**, 1060–1065 (2016).
- Liu, X. S. et al. Editing DNA methylation in the mammalian genome. *Cell* **167**, 233–247.e17 (2016).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

mESC culture, knock-in, shRNA knockdown, complementation, KO and EB formation. V6.5 mESCs were grown in N2B27 medium supplemented with two inhibitors and leukemia inhibitory factor, as described previously⁴⁸. MLL2-KO cells were generated in the laboratory as previously described¹⁴. TMutant cells will be described in another manuscript (C.C.S. et al., manuscript in preparation). For 3×FLAG-mCherry knock-in, oligonucleotides encoding the desired sgRNA sequence (5'-CGGTATAGCGGGCTTACCG-3') were annealed and cloned into the pX459 plasmid according to a published protocol⁴⁹. A donor plasmid containing homology arms flanking the *Magohb* promoter and a 3×FLAG-mCherry sequence was cloned into the pUC57 vector. mESCs were transfected with pX459 containing the desired sgRNA-coding sequences and donor plasmid using Nucleofector 2b (Lonza) according to the manufacturer's instructions. Transfected mESCs were selected with 2 µg ml⁻¹ puromycin (Life Technologies) for 48 h and grown for 10 d until cell clones were pickable. Cell clones were screened by PCR and the genotypes were confirmed by sequencing. For double-mutant MLL2 and SUZ12 KO, mESCs were transfected with pX459 containing the desired sgRNA-coding sequences using Nucleofector 2b (Lonza) according to the manufacturer's instructions. Transfected mESCs were selected with 2 µg ml⁻¹ puromycin (Life Technologies) for 48 h and grown for 10 d until cell clones were pickable. Cell clones were screened by PCR and the genotypes were confirmed by sequencing. The guide RNA sequences used are listed in Supplementary Table 3.

shRNA knockdown was performed by transducing mESCs with lentiviruses expressing shRNA for 24 h before selection with 2 µg ml⁻¹ puromycin for 3 d. The lentiviral construct containing shRNA against CXXC1 (TRCN0000241395) was purchased from Sigma.

MLL2 re-expression in MLL2-KO cells was performed by nucleo-transfecting 2 × 10⁷ mESCs with 40 µg MLL2 plasmid (Promega, FHC12727). Flow cytometry experiments were performed 5 d after transfection.

Fuw-dCas9-dead Tet1CD-P2A-BFP and Fuw-dCas9-9-Tet1CD-P2A-BFP were a gift from R. Jaenisch (Addgene plasmid no. 108246 and Addgene plasmid no. 108245). mESCs were transduced with sgRNA-expressing vector (2 sgRNAs targeting a control region or 2 sgRNAs targeting the *Magohb* promoter; see Supplementary Table 1 for sequences) and selected for 2 d with puromycin. mESCs were then transfected with Fuw-dCas9-dead Tet1CD-P2A-BFP or Fuw-dCas9-9-Tet1CD-P2A-BFP and flow cytometric measurements were performed 2 d later.

EBs were generated as described previously²² using the hanging-drop method. Briefly, 1,500 ESCs were cultured in 25 µl of EB differentiation medium on the lid of 150-mm culture plates for 6 d. The EB differentiation medium was composed of DMEM supplemented with 15% fetal bovine serum (Gemini Bio-Products), 1× GlutaMAX (Gibco), 1× MEM non-essential amino acids (Gibco), 1× β-mercaptoethanol (Gibco) and 1× penicillin-streptomycin (Gibco).

RT-qPCR. RNA was extracted from approximately 1 × 10⁶ cells using the RNeasy Kit (Qiagen), followed by DNase digestion using RNase-Free DNase (Qiagen). A 1-µg quantity of total RNA was reverse transcribed using Superscript II RT according to the manufacturer's protocol (Invitrogen) and quantitative PCR was performed using SYBR Green dye (Roche Scientific) and a CFX96 instrument (BioRad). For data analysis, normalized gene expression values were calculated using the comparative CT method (ΔΔCT method). Ct values of the target gene (*Magohb*) were normalized to Ct values of 16S ribosomal RNA and to that of a control sample (Control) to obtain ΔΔCt values. FC differences in gene expression compared to controls (calculated as 2^{-ΔΔCt}) are represented in Fig. 2c.

ChIP and RNA extraction. ESCs were fixed with 1% formaldehyde for 10 min. After quenching and cell lysis, chromatin was fragmented to 200–600 base pairs (bp) using a Covaris E220 bath sonicator (Covaris). Chromatin was incubated with antibodies overnight at 4 °C, immunoprecipitated the following day by incubating with Protein A/G agarose beads (Santa Cruz Biotechnology), washed six times in RIPA buffer (25 mM Tris pH 7.5, 140 mM NaCl, 1% Triton X-100, 1 mM EDTA, 0.1% SDS, 0.1% sodium deoxycholate, 0.5 mM dithiothreitol) and eluted (0.1 M NaHCO₃, 1% SDS). After cross-link reversal and proteinase K digestion overnight at 65 °C, DNA was purified using Qiagen PCR purification spin columns. ChIP-seq libraries were generated with the KAPA HTP library preparation kit (KAPA Biosystems) following the manufacturer's instructions and loaded onto a NextSeq 500 sequencer or NovaSeq (Illumina) for sequencing. At least two biological replicates were performed for ChIP-seq of H3K4me3, H3K27me3, MLL2 and SET1A under each experimental condition. RNA was isolated using the RNeasy mini kit (Qiagen) following the manufacturer's instructions, and RNA-seq libraries were made using the TruSeq Stranded Total RNA with Ribo-Zero kit (Illumina).

Precision nuclear run-on sequencing. We performed PRO-seq following a previously described protocol⁵⁰ with some modifications. Nuclei were isolated by a Dounce homogenizer with a loose pestle. Approximately 10⁷ nuclei were subjected to nuclear run-on (30 °C, 3 min) in the presence of 25 µM biotin-11-ATP/GTP/CTP/UTP (PerkinElmer) and 6 × 10⁶ *Drosophila* S2 spike-in nuclei. Total RNAs were extracted and hydrolyzed in 0.2 M NaOH (on ice, 10 min). Biotinylated

nascent RNA was purified by Dynabeads M-280 streptavidin (Invitrogen). After 3' VRA3 adapter ligation and the second purification by Dynabeads, 5' cap and 5' hydroxyl RNA were converted to 5' monophosphorylated RNA by RppH (NEB) and PNK (NEB), respectively. After 5' VRA5 adapter ligation and the third purification by Dynabeads, complementary DNA was generated by reverse transcription with SuperScript III (Invitrogen) and RP1 primer. PCR using Phusion Hot Start (Thermo) and RP1/RPIx primer sets in 10 cycles amplified the indexed DNA libraries. DNA libraries of 140–350 bp were size-selected by Pippin HT with a 2% gel cassette and the marker 20B (Sage Science). Two independent cell cultures per condition were then sequenced by the NextSeq 500 system (Illumina) with single-read runs.

mRRBS. mRRBS was performed as previously reported^{38,51–53}. Briefly, genomic DNA was digested with MspI (New England BioLabs) before size selection of 100- to 250-bp fragments with solid-phase reversible immobilization beads (MagBio Genomics). DNA was bisulfite-converted with the EZ DNA Methylation-Lightning Kit (Zymo Research), resulting in a conversion efficiency of 99.6% ± 0.05% s.d. based on the observed CpG methylation frequency of an unmethylated λ-bacteriophage spike-in control genome (New England BioLabs, N3013S). Libraries were prepared with the Pico Methyl-Seq Library Prep Kit (Zymo Research) using Illumina TruSeq indices and sequenced using single-end reads (NextSeq 500, Illumina) with a 500/550 V2 High Output reagent kit (1 × 75 cycles). Bioinformatic processing and alignment of the sequenced libraries to the mm9 reference genome were performed as previously reported^{38,51–53}. The code used for mRRBS data processing is reported elsewhere⁵⁴. CpG methylation was quantified using the SeqMonk platform (v1.40.1) with the bisulfite feature methylation pipeline. Between-group comparisons of individual CpG values were performed using the Brown-Forsythe and Welch ANOVA method and the Games-Howell multiple comparisons test with individual variances computed for each comparison.

Antibodies. The following antibodies were generated in-house: H3K4me3 rabbit no. 519 (1:2,000), H3K27me3 rabbit no. 67 (1:2,000), MLL2Nterm rabbit no. 9170 (1:1,000), MLL2Cterm rabbit no. 9173 (1:1,000), WDR82 rabbit affinity-purified (1:1,000), SET1A rabbit no. 919 (1:1,000) and SET1B rabbit no. 386 (1:2,000). Antibody against CXXC1 was purchased from Abcam (ab56035), antibodies against DNMT1 and CAS9 were purchased from CST (5032S, 14697), antibody against DNMT3A was purchased from Abcam (ab2850), antibody against FLAG was purchased from Sigma (F3165) and antibodies against SUZ12, HSP90 and β-tubulin were purchased from Santa Cruz (catalog nos. sc-9104, sc-13119 and sc-46264).

Genome-wide CRISPR-Cas9 screen. Brie library amplification, lentiviral production, MOI determination and transduction were performed as described previously³⁵. Briefly, ~300 million MLL2-KO cells stably expressing Cas9 were infected with the genome-wide sgRNA Brie library²⁶ at a MOI < 0.3. Infected cells underwent puromycin selection (2 µg ml⁻¹) for 7 d. A total of 4–6 million cells were pelleted and frozen ('total population') and ~10 million cells were sorted for cells with an increased mCherry fluorescence compared to control MLL2-KO cells ('Sorted cells'). To achieve a better selection and reduce the number of false-positive cells, cells were sorted twice.

Genomic DNA was extracted from the 'Total population' and 'Sorted cells' and the sgRNA library was amplified by PCR as previously described³⁵. Deep sequencing on an Illumina Nextseq platform was used to monitor the library composition. Guide composition between the 'Total population' and 'Sorted cells' was compared: briefly, the enrichment of individual guides was calculated as the log₂ ratios of 'Sorted cells' and 'Total population', with each sgRNA with a log₂[FC] > 1 considered as enriched. Each gene was then classified into one of the five categories: zero sgRNAs enriched out of four sgRNAs; one sgRNA enriched out of four sgRNAs, two sgRNAs enriched out of four sgRNAs; three sgRNAs enriched out of four sgRNAs; or four sgRNAs enriched out of four sgRNAs.

KO of candidate genes in pooled ESCs. For the eight top candidate genes (four out of four sgRNAs enriched), individual targeted sgRNAs were cloned into the lentiCRISPRv2 plasmid (Addgene no. 52961). These constructs were used to produce lentivirus and transduce MLL2-KO cells, followed by puromycin selection for 7 d and western blot analyses. The guide RNA sequences used are listed in Supplementary Table 3.

Sequencing data processing. RNA-seq and ChIP-seq samples were sequenced with Illumina technology, and output data were processed with the bcl2fastq software tool. Sequence quality was assessed using FastQC v0.11.2 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), and quality trimming was performed using the FASTX toolkit. RNA-seq and ChIP-seq reads were aligned to the mm9 genome using TopHat v2.0.9 (ref. ⁵⁶) and Bowtie v0.12.9 (ref. ⁵⁷), respectively, and only uniquely mapped reads meeting a two-mismatch threshold were considered for downstream analysis. Gene annotations from Ensembl 67 were used. Output BAM files were converted into BigWig track files to display coverage throughout the genome (in reads per million) using the GenomicRanges package.

For Fig. 7d, the total number of ChIP-seq reads was normalized to the number of *Drosophila* spike-in reads.

ChIP-seq analysis and RNA-seq analysis. Peaks were called with MACS v1.4.2 (ref.⁵⁸) software using default parameters. Meta-plots and heat maps were generated using deeptools⁵⁹. k-means clustering was also performed using deeptools, and nearest-gene log[FC] in gene expression corresponding to the clustered peaks in the heat maps were determined using in-house scripts (gene count tables were used as input for edgeR 3.0.8 (ref.⁶⁰) to analyze RNA-seq data; the in-house script is available at <https://github.com/ebartom/NGSbartom>) and visualized with Java TreeView⁶¹.

PRO-seq analysis. PRO-seq raw data were trimmed by cutadapt 1.14 (ref.⁶²) and Trimmomatic 0.33 (ref.⁶³). Each sample was next mapped to the mouse and fruit fly genome assemblies (mm9 and dm3, respectively) using Bowtie. The number of fly reads in each experiment was used to normalize the corresponding genome-wide strand-specific profiles. Box plots quantifying the PRO-seq level were generated by quantifying the PRO-seq level around the TSS, considering the direction of transcription (± 3 kb) and normalized using the total number of reads per sample.

Statistical analysis. For statistical analyses, R, Python and GraphPad were used. Appropriate statistical tests were used for all data where a statistical analysis was reported. All experiments were conducted in unblinded conditions and the statistical tests used are reported in the figure legends.

GO terms. The P values for GO were determined with the hypergeometric test using the software Metascape⁶⁴. GSEA of the lists of genes pre-ranked by the edgeR stat value was performed with the GSEA software⁶⁵. Mouse genes were ranked by the ratio between WT and MLL2-KO RNA-seq expression on EBs at day 6, MLL2-KO and 5dAza-treated MLL2-KO RNA-seq expression on EBs at day 6 and MLL2-KO and double-mutant MLL2- and SUZ12-KO RNA-seq expression on EBs at day 6.

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

Data availability

RNA-seq, ChIP-seq and bisulfite sequencing raw data are available in the Gene Expression Omnibus database under accession [GSE129037](#). Additional data supporting the findings of this study are available from the corresponding author upon request. Source data for Figs. 2 and 6 and Extended Data Figs. 2, 3 and 5 (full scans of the immunoblots) and Figs. 5 and 7 (statistical source data) are provided with the paper.

Code availability

We made use of publicly available software and tools (as referenced in the Methods section, the in-house script used for ChIP-seq and RNA-seq analyses is available at <https://github.com/ebartom/NGSbartom>).

References

48. Cao, K. et al. SET1A/COMPASS and shadow enhancers in the regulation of homeotic gene expression. *Genes Dev.* **31**, 787–801 (2017).
49. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
50. Mahat, D. B. et al. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat. Protoc.* **11**, 1455–1476 (2016).
51. Weinberg, S. E. et al. Mitochondrial complex III is essential for suppressive function of regulatory T cells. *Nature* **565**, 495–499 (2019).
52. Wang, L. et al. TET2 coactivates gene expression through demethylation of enhancers. *Sci. Adv.* **4**, eaau6986 (2018).

53. Walter, J. M., Helmin, K. A., Abdala-Valencia, H., Wunderink, R. G. & Singer, B. D. Multidimensional assessment of alveolar T cells in critically ill patients. *JCI Insight* **3**, e12328 (2018).
54. Singer, B. D. A practical guide to the measurement and analysis of DNA methylation. *Am. J. Respir. Cell Mol. Biol.* **61**, 417–428 (2019).
55. Pusapati, G. V. et al. CRISPR screens uncover genes that regulate target cell sensitivity to the morphogen sonic hedgehog. *Dev. Cell* **44**, 271 (2018).
56. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
57. 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).
58. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
59. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
60. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
61. Saldaña, A. J. Java Treeview-extensible visualization of microarray data. *Bioinformatics* **20**, 3246–3248 (2004).
62. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
63. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
64. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).
65. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

Acknowledgements

We thank the Shilatifard laboratory members for helpful suggestions and discussions. K.A.H. and B.D.S. are supported by NIH K08HL128867. C.C.S. is supported, in part, by the NIH Predoctoral to Postdoctoral Transition Award F99CA234945. K.C. is supported, in part, by the NIH Pathway to Independence Award K99HD094906. A.P. is supported by the NIH Pathway to Independence Award K99CA234434. E.R.S. is supported by NIH R50CA211428. We thank N. J. Ethen for the graphical representation of the model (Fig. 8). Studies in the Shilatifard laboratory related to COMPASS are supported by the NCI Outstanding Investigator Award R35CA197569.

Author contributions

A.S. and D.D. conceived and initiated the project. D.D., C.C.S. and A.P.S. performed RNA-seq and ChIP-seq studies. D.D. wrote the manuscript. C.C.S., M.U., M.A.M. and K.C. generated mutant mESCs, and C.R., Z.Z. and C.C.S. assisted with mammalian studies. K.A.H. performed mRRBS, and mRRBS data were analyzed by B.D.S. RNA-seq and ChIP-seq data were analyzed by D.D. and E.T.B., while libraries were generated and sequenced by E.J.R., D.Z. and S.A.M. Critical feedback and advice were provided by E.R.S., A.P. and A.S. throughout the course of this project.

Competing interests

The authors declare no competing interests.

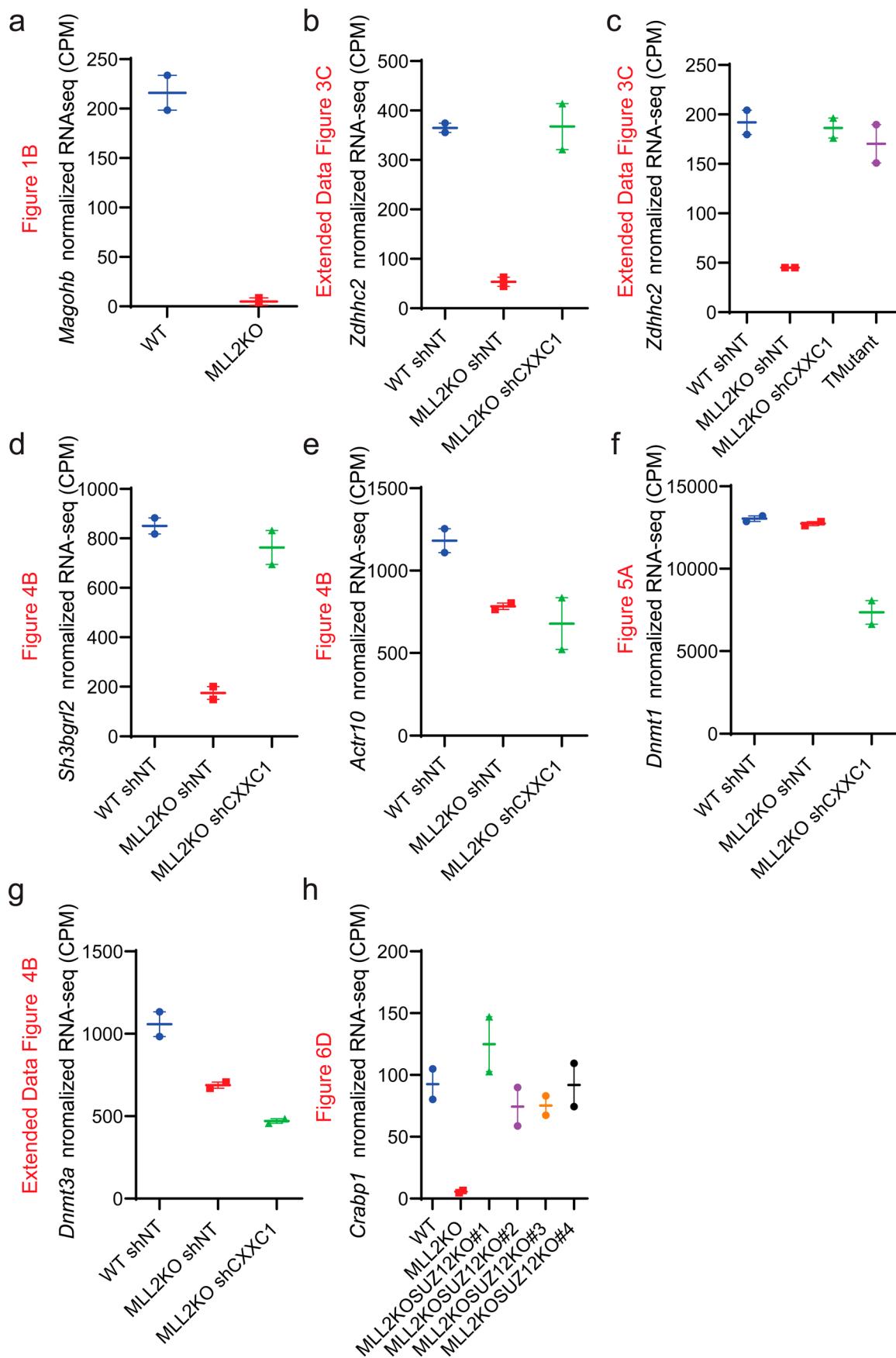
Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41588-020-0618-1>.

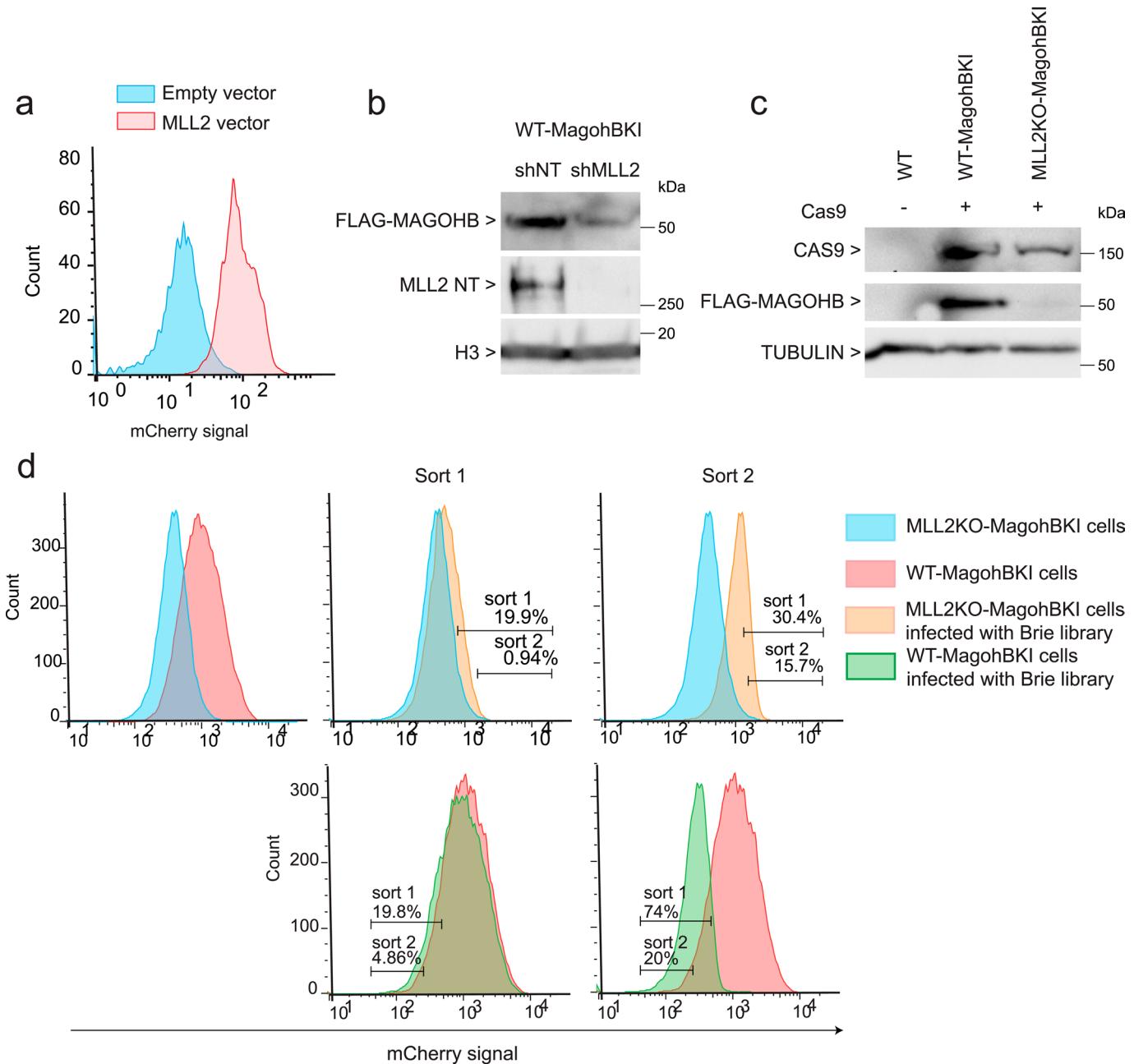
Supplementary information is available for this paper at <https://doi.org/10.1038/s41588-020-0618-1>.

Correspondence and requests for materials should be addressed to A.S.

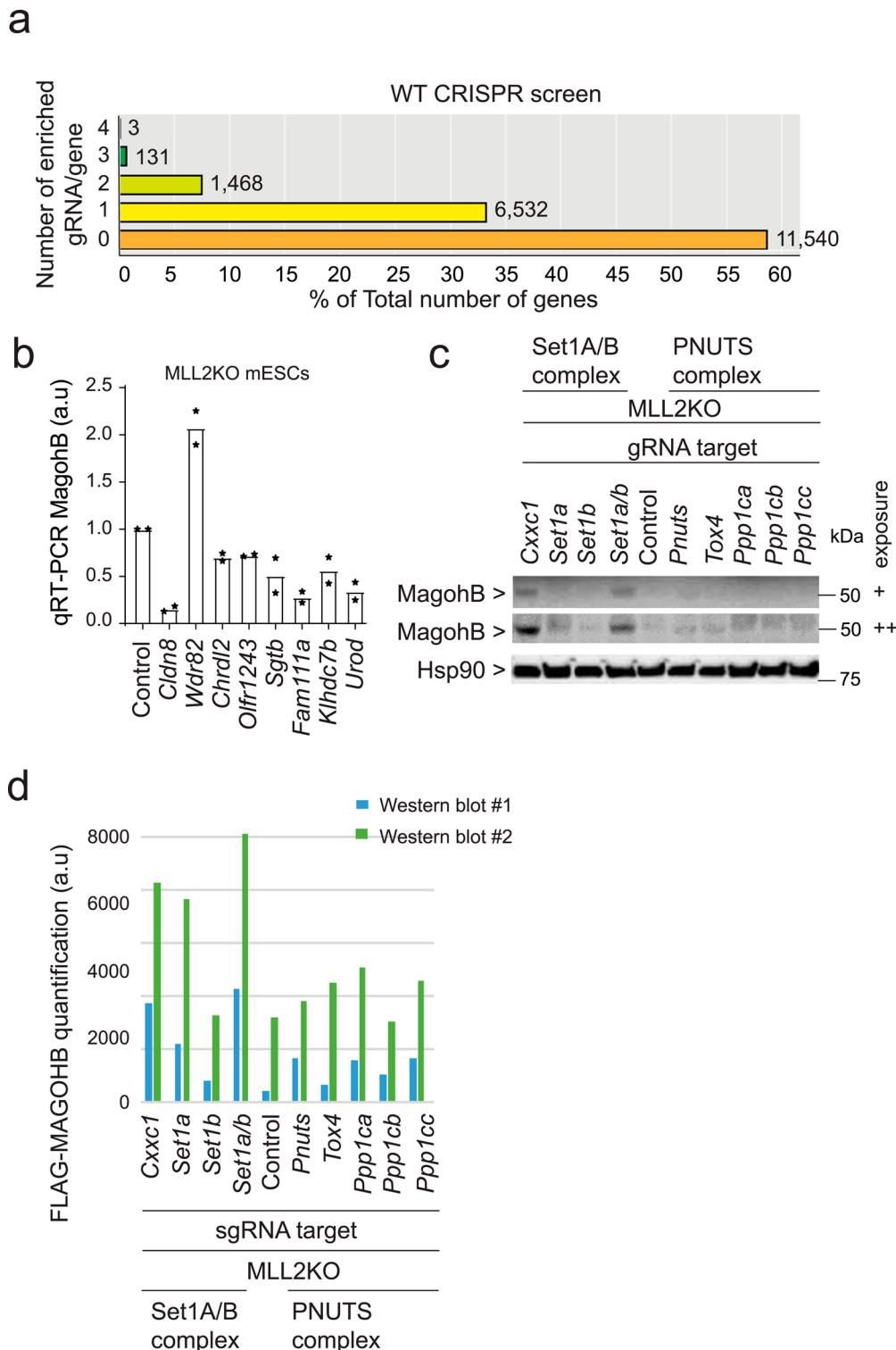
Reprints and permissions information is available at www.nature.com/reprints.



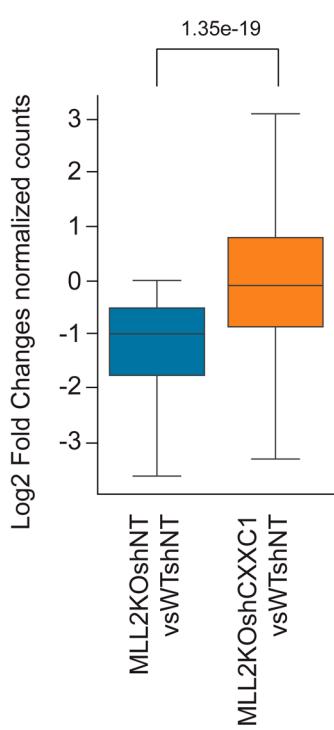
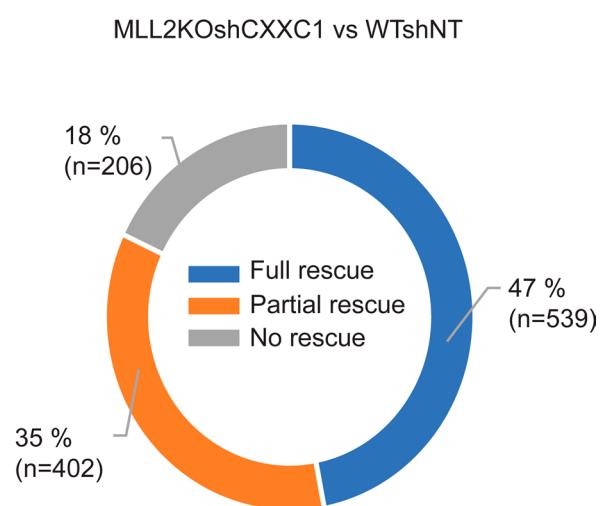
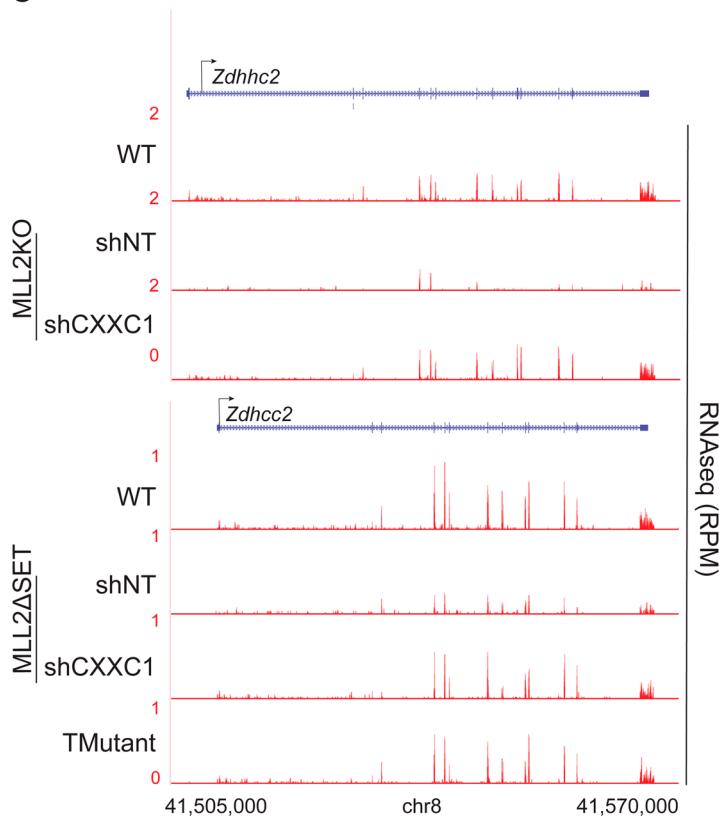
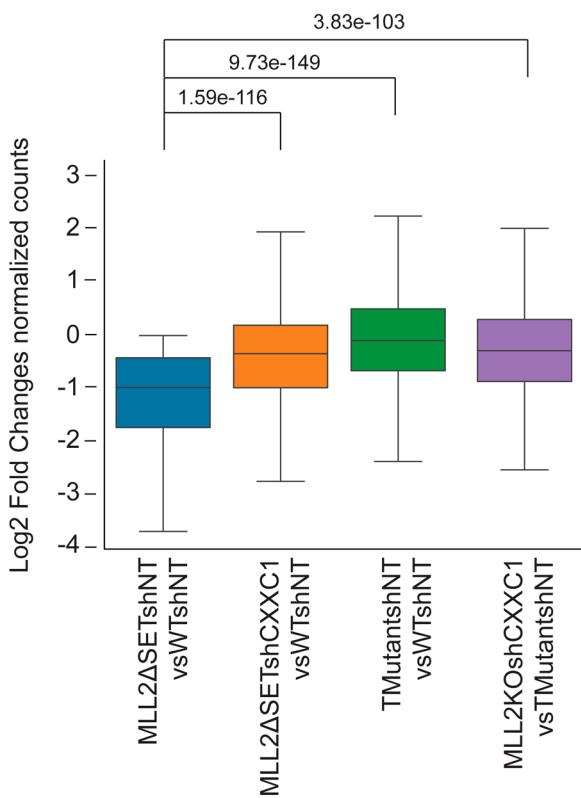
Extended Data Fig. 1 | Quantification of RNAseq changes. **a-h**, Interleaved scatter plot of indicated genes' normalized RNA-seq counts in the indicated conditions for the genes *Magohb* (a), *Zdhhc2* (b), *Zdhhc2*(c), *Sh3bgfl2* (d), *Actr10* (e), *Dnmt1* (f), *Dnmt3a* (g) and *Crabp1* (h) ($n=2$). CPM, counts per million.



Extended Data Fig. 2 | CRISPR-Cas9 screen using MAGOHB knock in in mESCs. **a**, FACS analysis of *mCherry-Magohb* MLL2KO mESCs transformed with an empty vector or a MLL2 expressing vector. The experiment was repeated three times independently with similar results. **b**, Western blot for FLAG-MAGOHB expression in WT cells expressing a shRNA control (shNT) or a shRNA targeting MLL2 (shMLL2). The experiment was repeated two times independently with similar results. **c**, Western blot for CAS9, FLAG-MAGOHB and TUBULIN (loading control) protein levels in WT, *Magohb* KI WT cells (WT-MagohbKI) and *Magohb* KI MLL2KO cells (MLL2KO-MagohbKI). The experiment was repeated two times independently with similar results. **d**, Top panel: *mCherry* high cells were isolated via two successive rounds of FACS in MLL2KO mESCs. Bottom panel: *mCherry* low cells were isolated via two successive rounds of FACS in WT mESCs. The experiment was repeated three times independently with similar results. Uncropped gels are available as source data.

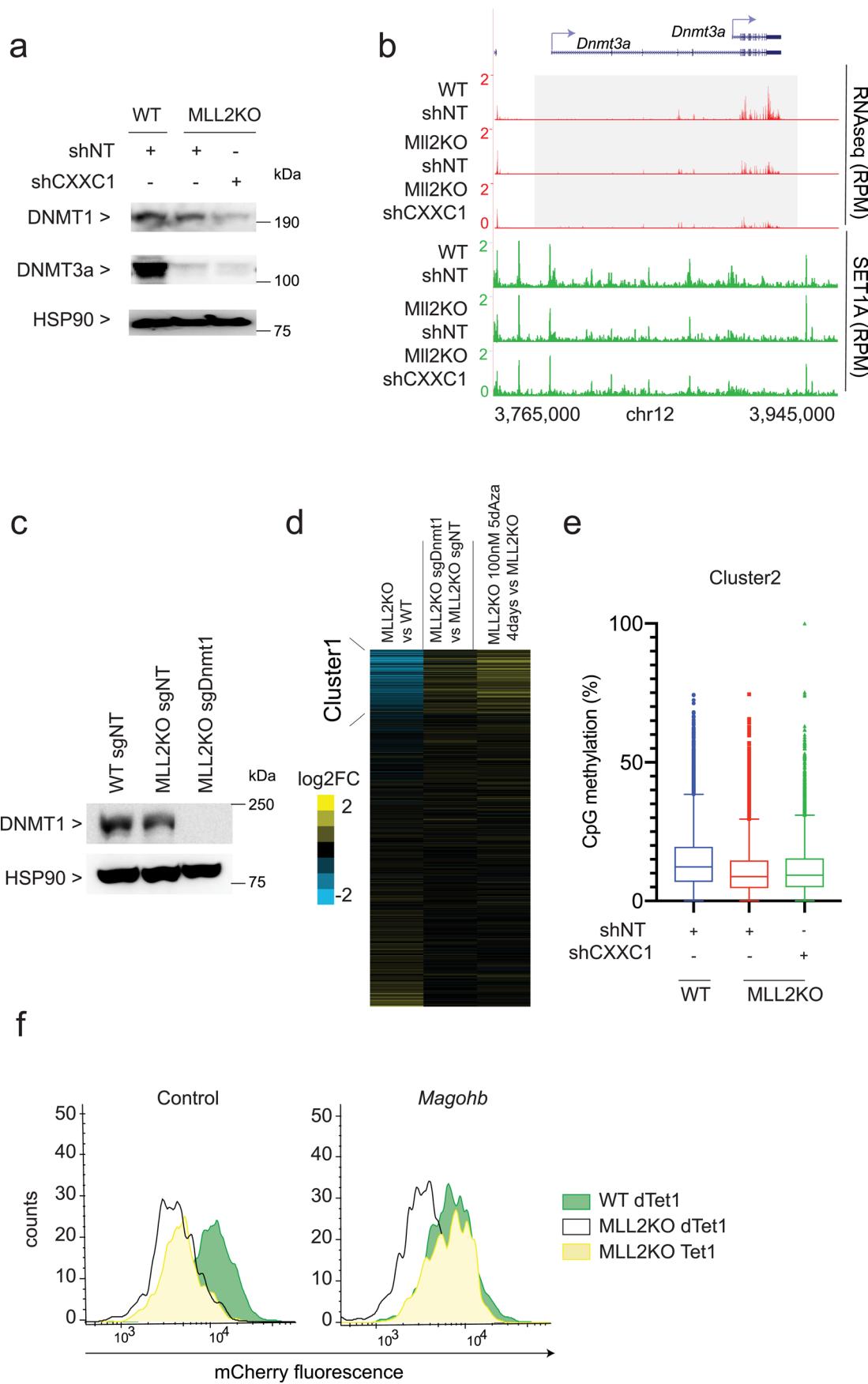


Extended Data Fig. 3 | CRISPR screen validation. **a**, CRISPR screen results representing the number of enriched sgRNA in the ‘Sorted cells’ compared to the ‘Total population’ per gene for WT screen. The majority of the genes fall into the category ‘0 sgRNA enriched out of 4’. **b**, Expression levels of *Magohb* as assessed by reverse transcription quantitative PCR in MLL2KO mESCs expressing a control sgRNA (Control) or sgRNAs targeting various genes ($n=2$). **c**, Western blot of MAGOHB and HSP90 (loading control) protein levels. Two different exposures are shown. The experiment was repeated two times independently with similar results. **d**, Quantification of two independent replicates of Extended Data Fig. 2c western blots. Uncropped gels are available as source data.

a**b****c****d**

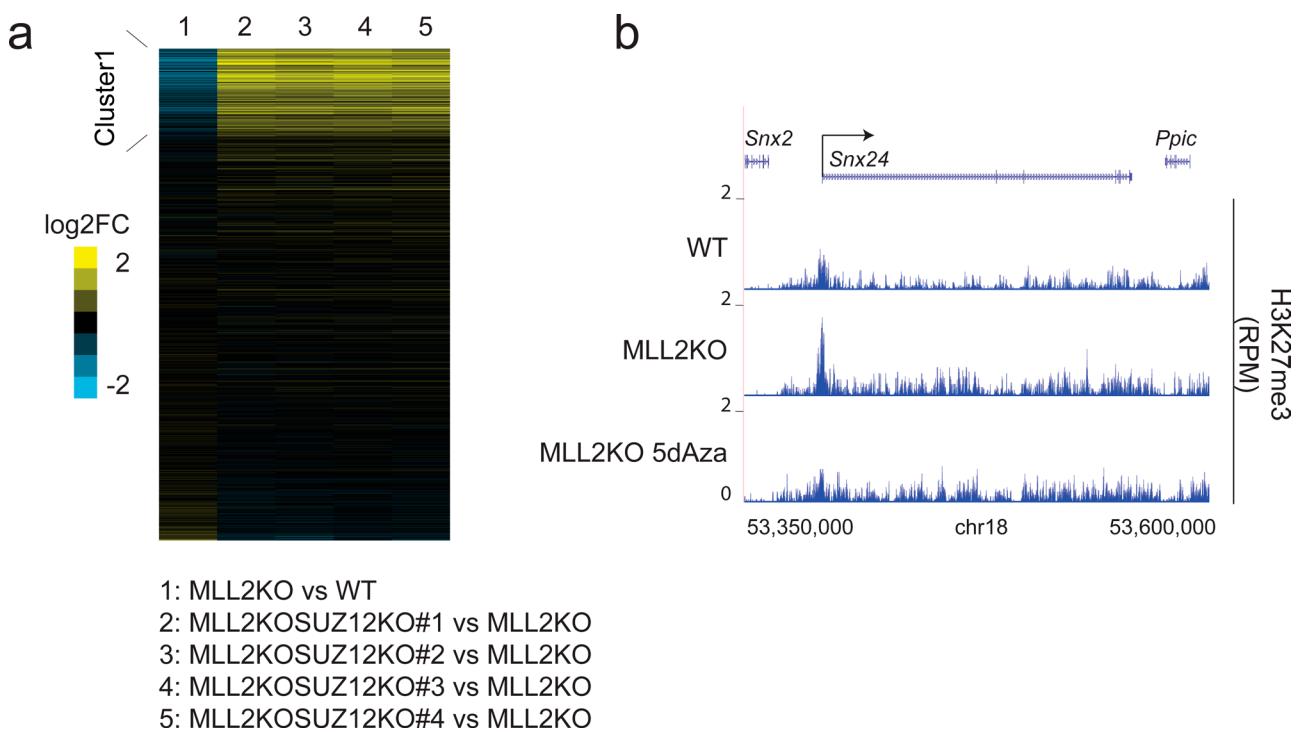
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Cluster-1 gene expression rescue by CXXC1 knockdown. **a**, Boxplot analysis of cluster1 gene expression in 1) MLL2KOshNT vs. WTshNT mESCs and 2) MLL2KOshCXXC1 vs. WTshNT mESCs. P values were computed using Wilcoxon test (two-sided), $n=2,021$. The boxplots indicate the median (middle line), the third and first quartiles (box) and the first and fourth quartiles (error bars). **b**, Donut chart of cluster 1 distribution of genes based on their level of rescue in MLL2KO shCXXC1 cells compared to WT cells. ‘Full rescue’ genes were characterized by a $\log_{2}FC \geq 0$ comparing MLL2KOshCXXC1 to WT mESCs, ‘Partial rescue’ genes were characterized by a $\log_{2}FC > 0$ comparing MLL2KOshCXXC1 and MLL2KO mESCs and ‘No rescue’ were characterized by a $\log_{2}FC \leq 0$ comparing MLL2KOshCXXC1 and MLL2KO mESCs. **c**, RNA-seq tracks for *Zdhhc2* in WT, MLL2KOshNT, MLL2KOshCXXC1, MLL2ΔSETshNT, MLL2ΔSETshCXXC1 and TMutant. The experiment was repeated two times independently with similar results. **d**, Boxplot analysis of cluster1 gene expression in 1) MLL2ΔSETshNT vs. WTshNT mESCs and 2) MLL2ΔSETshCXXC1 vs. WTshNT mESCs, 3) TMutantshNT vs. WTshNT mESCs, 4) TMutantshCXXC1 vs. WTshNT mESCs and 5) MLL2KOshCXXC1 vs TMutantshNT mESCs ($n=2$).



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | DNA methylation and MLL2 dependent transcription. **a**, Western blot of DNMT1, DNMT3A and HSP90 (loading control) protein levels in WTshNT, MLL2KOshNT and MLL2KOshCXXC1 mESCs. The experiment was repeated two times independently with similar results. **b**, RNA-seq and SET1A ChIP-seq tracks for *Dnmt3a*. The experiment was repeated two times independently with similar results. **c**, Western blot of DNMT1 and HSP90 (loading control) protein levels in WTsgNT, MLL2KOsgNT and MLL2KOsgDnmt1. The experiment was repeated two times independently with similar results. **d**, Heatmaps show the corresponding log₂ fold changes in gene expression in 1) MLL2KO vs. WT mESCs, 2) MLL2KO sgDnmt1 vs. MLL2KOsgNT mESCs and 3) MLL2KO treated for 4 days with 100 nM 5dAza vs. MLL2KO mESCs. **e**, Box-and-Whisker plot quantifying changes in CpG methylation around TSS (± 3 kb) of cluster2 clusters identified in Fig. 3a (N = 3 biological replicates, n = 13420). The boxplots indicate the median (middle line), the third and first quartiles (box) and the first and fourth quartiles (error bars). **f**, Flow cytometry on mCherry-Magohb KI WT or MLL2KO mESCs expressing a dead Tet1 catalytic domain (dTet1) or an active Tet1 catalytic domain (Tet1) targeted to a control region 5 kb upstream Magohb or Magohb promoter. The experiment was repeated two times independently with similar results. Uncropped gels are available as source data.



Extended Data Fig. 6 | H3K27me3 and MLL2 dependent transcription. **a**, Heatmaps show the corresponding log₂ fold changes in gene expression in 1) MLL2KO vs. WT mESCs, 2) MLL2KOSUZ12KO#1 vs. MLL2KO mESCs, 3) MLL2KOSUZ12KO#2 vs. MLL2KO mESCs, 4) MLL2KOSUZ12KO#3 vs. MLL2KO mESCs and 5) MLL2KOSUZ12KO#4 vs. MLL2KO mESCs. The experiment was repeated two times independently with similar results. **b**, ChIP-seq tracks of H3K27me3 occupancy at the *Snx24* locus in WT, MLL2KO and MLL2KO 100 nM 5dAza for 4 days treated mESCs. The experiment was repeated two times independently with similar results.

Reporting Summary

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

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

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

Our web collection on [statistics for biologists](#) may be useful.

Software and code

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

Data collection	Data collection was performed using BD FACSDiva Version 8.0.2 and FlowJo software v10.0.7.
Data analysis	mRRBS: The code used for mRRBS data processing is available: Singer, B.D. A Practical Guide to the Measurement and Analysis of DNA Methylation. Am J Respir Cell Mol Biol 61, 417-428 (2019). CpG methylation was quantified using the SeqMonk platform (v1.40.1) with the bisulfite feature methylation pipeline. CRISPR screen: The code used to create the count table named 'count_spacers.py' can be found here https://github.com/fengzhanglab/Screening_Proocols_manuscript/blob/master/count_spacers.py Sequencing data processing: bcl2fastq software tool, FastQC v0.11.2, FASTX toolkit, TopHat v2.0.9, Bowtie v0.12.9, GenomicRanges package ChIP-seq and RNA-seq: MACS v1.4.2, deeptools2.0, edgeR 3.0.8, the in house pipeline code can be found here https://github.com/ebartom/NGSbartom PRO-seq: cutadapt 1.14, Trimmomatic 0.33, Bowtie v0.12.9

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

Data

Policy information about [availability of data](#)

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

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

All processed and raw data are available from NCBI GEO GSE129037. Raw data was analyzed and described in the following figures and table: Figure1, Figure2, Figure3, Figure4, Figure5, Figure 6, Figure 7,Supplementary Figure2, Supplementary Figure3, Supplementary Figure4, Supplementary Figure5, Supplementary Table1, Supplementary Table2.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size	Sample size was determined to be adequate based on the magnitude and consistency of measurable differences between groups, and compared against sizes known from previous experience. Sample sizes of all experiments were chosen in agreement with guidelines for the analysis of next-generation sequencing data (n≥2).
Data exclusions	No samples were excluded.
Replication	All experimental findings were reproduced and results obtain from each. Reproducibility between biological replicates was assessed by PCA analysis, overall distribution and output.
Randomization	No randomization was required for this cell-culture based in vitro study.
Blinding	All experiments were conducted in unblinded conditions, as no experiment required blinding.

Reporting for specific materials, systems and methods

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used

The following antibodies were generated in-house: H3K4me3 rabbit #519 (1:2000), H3K27me3 rabbit #67 (1:2000), MLL2Nterm rabbit #9170 (1:1000), MLL2Cterm rabbit #9173 (1:1000), WDR82 rabbit affinity purified (1:1000), SET1A rabbit #919 (1:1000) and SET1B rabbit #386 (1:2000). Antibody to CXXC1 (1:1000) was purchased from Abcam (ab56035), antibodies to DNMT1 (1:1000) and CAS9 (1:2000) were purchased from CST (5032S, 14697), antibody to DNMT3A (1:1000) was purchased from Abcam (ab2850), antibody to FLAG (1:500) was purchased from Sigma (F3165) and antibodies to SUZ12 (1:2000), HSP90 (1:4000) and β-TUBULIN (1:1000) were purchased from Santa Cruz (#sc-9104, #sc-13119 and #sc-46264).

Validation

All the antibodies used were tested for specificity by the distributors or in previous studies (using a depletion approach or KO cells): Thornton, J.L. et al. Context dependency of Set1/COMPASS-mediated histone H3 Lys4 trimethylation, Hu, D. et al. Not Not

All H3K4 Methylation Are Created Equal: MLL2/COMPASS Dependency in Primordial Germ Cell Specification., Wu, M. et al. Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS, Sze, C.C. et al. Histone H3K4 methylation-dependent and -independent functions of Set1A/COMPASS in embryonic stem cell self-renewal and differentiation, Wang, L. A cytoplasmic COMPASS is necessary for cell survival and triple-negative breast cancer pathogenesis by regulating metabolism.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	The v6.5 mouse embryonic stem cell line was established from cells derived from the inner cell mass (ICM) of a 3.5 day old male mouse embryo from a C57BL/6 X 129/sv cross.
Authentication	This cell line was not authenticated.
Mycoplasma contamination	v6.5 mouse embryonic stem cells were tested by total RNA-seq for presence of mycoplasma
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study

ChIP-seq

Data deposition

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

Data access links

May remain private before publication.

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE129037>

Files in database submission

```
GSM3692066 mES_WT_Mll2ChIP_rep1
GSM3692067 mES_Mll2KO_Mll2ChIP_rep1
GSM3692068 mES_input_Mll2ChIP_rep1
GSM3692069 mES_WT_Mll2ChIP_rep2
GSM3692070 mES_Mll2KO_Mll2ChIP_rep2
GSM3692071 mES_input_Mll2ChIP_H3K4me3ChIP_rep2
GSM3692072 mES_WT_shNT_H3K4me3ChIP_rep1
GSM3692073 mES_Mll2KO_shNT_H3K4me3ChIP_rep1
GSM3692074 mES_Mll2KO_shCXXC1_H3K4me3ChIP_rep1
GSM3692075 mES_input_H3K4me3ChIP_rep1
GSM3692076 mES_WT_shNT_H3K4me3ChIP_rep2
GSM3692077 mES_Mll2KO_shNT_H3K4me3ChIP_rep2
GSM3692078 mES_Mll2KO_shCXXC1_H3K4me3ChIP_rep2
GSM3692109 mES_WT_shNT_Set1aChIP
GSM3692110 mES_Mll2KO_shNT_Set1aChIP
GSM3692111 mES_Mll2KO_shCXXC1_Set1aChIP
GSM4002621 mES_WT_shNT_Set1aChIP_rep2
GSM4002622 mES_Mll2KO_shNT_Set1aChIP_rep2
GSM4002623 mES_Mll2KO_shCXXC1_Set1aChIP_rep2
GSM4002624 mES_input_Set1aChIP
GSM4002625 mES_WT_shNT_H3K27me3_rep1
GSM4002626 mES_Mll2KO_shNT_H3K27me3_rep1
GSM4002627 mES_Mll2KO_shCXXC1_H3K27me3_rep1
GSM4002628 mES_input_H3K27me3_rep1
GSM4002629 mES_WT_shNT_H3K27me3_rep2
GSM4002630 mES_Mll2KO_shNT_H3K27me3_rep2
GSM4002631 mES_Mll2KO_shCXXC1_H3K27me3_rep2
GSM4002632 mES_input_H3K27me3_rep2
GSM4002633 mES_WTRNA_double_rep1
GSM4002634 mES_Mll2KORNA_double_rep1
GSM4002635 mES_Mll2KOSuz12KO#1RNA_double_rep1
GSM4002636 mES_Mll2KOSuz12KO#2RNA_double_rep1
GSM4002637 mES_Mll2KOSuz12KO#3RNA_double_rep1
GSM4002638 mES_Mll2KOSuz12KO#4RNA_double_rep1
GSM4002639 mES_WTRNA_double_rep2
GSM4002640 mES_Mll2KORNA_double_rep2
GSM4002641 mES_Mll2KOSuz12KO#1RNA_double_rep2
GSM4002642 mES_Mll2KOSuz12KO#2RNA_double_rep2
GSM4002643 mES_Mll2KOSuz12KO#3RNA_double_rep2
GSM4002644 mES_Mll2KOSuz12KO#4RNA_double_rep2
GSM4339122 mES_WT_DMSO_mRRBS_rep1
GSM4339123 mES_Mll2KO_DMSO_mRRBS_rep1
GSM4339124 mES_Mll2KO_100nM5dAza_mRRBS_rep1
```

GSM4339125 mES_WT_DMSO_mRRBS_rep2
 GSM4339126 mES_MII2KO_DMSO_mRRBS_rep2
 GSM4339127 mES_MII2KO_100nM5dAza_mRRBS_rep2
 GSM4339128 mES_WT_DMSO_mRRBS_rep3
 GSM4339129 mES_MII2KO_DMSO_mRRBS_rep3
 GSM4339130 mES_MII2KO_100nM5dAza_mRRBS_rep3
 GSM4339131 mES_WT_DMSO_H3K27me3_rep1
 GSM4339132 mES_MII2KO_DMSO_H3K27me3_rep1
 GSM4339133 mES_MII2KO_100nM5dAza_H3K27me3_rep1
 GSM4339134 mES_input_WT_DMSO_H3K27me3_rep1
 GSM4339135 mES_input_WT_DMSO_H3K27me3_rep1
 GSM4339136 mES_input_MII2KO_DMSO_H3K27me3_rep1
 GSM4339137 mES_input_MII2KO_100nM5dAza_H3K27me3_rep1
 GSM4339138 mES_input_MII2KOSuz12KO_DMSO_H3K27me3_rep1
 GSM4339139 mES_WT_DMSO_H3K27me3_rep2
 GSM4339140 mES_MII2KO_DMSO_H3K27me3_rep2
 GSM4339141 mES_MII2KO_100nM5dAza_H3K27me3_rep2
 GSM4339142 mES_MII2KOSuz12KO_DMSO_H3K27me3_rep2
 GSM4339143 mES_input_WT_DMSO_H3K27me3_rep2
 GSM4339144 mES_input_MII2KO_DMSO_H3K27me3_rep2
 GSM4339145 mES_input_MII2KO_100nM5dAza_H3K27me3_rep2
 GSM4339146 mES_input_MII2KOSuz12KO_DMSO_H3K27me3_rep2
 GSM4339147 mES_WT_EB_6d_DMSORNA_rep1
 GSM4339148 mES_MII2KO_EB_6days_DMSORNA_rep1
 GSM4339149 mES_MII2KO_EB_6days_100nM5dAzaRNA_rep1
 GSM4339150 mES_MII2KOSuz12KO_EB_6days_RNA_rep1
 GSM4339151 mES_WT_EB_6d_DMSORNA_rep2
 GSM4339152 mES_MII2KO_EB_6days_DMSORNA_rep2
 GSM4339153 mES_MII2KO_EB_6days_100nM5dAzaRNA_rep2
 GSM4339154 mES_MII2KOSuz12KO_EB_6days_RNA_rep2
 GSM4339155 mES_WT_EB_6d_DMSORNA_rep3
 GSM4339156 mES_MII2KO_EB_6days_DMSORNA_rep3
 GSM4339157 mES_MII2KO_EB_6days_100nM5dAzaRNA_rep3
 GSM4339158 mES_MII2KOSuz12KO_EB_6days_RNA_rep3

Genome browser session
(e.g. [UCSC](#))

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE129037>

Methodology

Replicates

At least two replicate for each ChIPseq, RNAseq, PROseq or bisulfite sequencing experiments.

Sequencing depth

Samples were sequenced with 20-50 million reads for each sample

Antibodies

The following antibodies generated in-house were used for ChIPseq: H3K4me3 rabbit #519, H3K27me3 rabbit #67, MLL2Cterm rabbit #9173, and SET1A rabbit #919.

Peak calling parameters

Peaks were called with MACS v1.4.2 software using default parameters. Meta-plots and heat maps were generated using deeptools. k-means clustering was also performed using deeptools, and nearest-gene log fold changes in gene expression corresponding to the clustered peaks in the heat maps were determined using in-house scripts (Gene count tables were used as input for edgeR 3.0.8 51 to analyze RNA-seq data) and visualized with Java TreeView.

Data quality

Mapping quality filtering to exclude secondary alignments and reads aligning to multiple locations, a MACS pvalue cutoff of 1.00e-05 and visual inspection of ChIP, input and peak tracks on UCSC browser.

Software

Information regarding the software used to analyze our data is provided in the methods section under the subheadings: NGS Data Processing, ChIP-seq Analysis, and RNA-seq Analysis. All custom code used for this study will be available as a supplementary file.

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- 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).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	WT or MII2KO mESCs with mCherry-MagohB knockin were used
Instrument	BDFACS Aria
Software	FlowJo
Cell population abundance	Approximatively 10 millions cells were sorted for each screening.
Gating strategy	WT and MII2KO mESCs with mCherry-MagohB knockin were used to establish cell gating for Brie library sorting. Supplementary figure 1

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