

RNA LEVER Mediates Long-Range Regulation of ϵ -globin by Keeping PRC2 in Check

Wei Wen Teo^{1,§}, Xinang Cao^{1,2,§}, Chan-Shuo Wu¹, Hong Kee Tan^{1,3}, Qiling Zhou¹, Henry Yang¹,
Li Chai,⁴ and Daniel G. Tenen^{1,5,6}

¹Cancer Science Institute of Singapore, National University of Singapore, Singapore 117599

² Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, 119228

³National University of Singapore Graduate School for Integrative Sciences and Engineering, 21 Lower Kent Ridge, Singapore 119077

⁴Department of Pathology, Brigham and Women's Hospital, Boston, MA 02115, USA

⁵Harvard Stem Cell Institute, Harvard Medical School, Boston, MA 02115 USA

⁶Harvard Medical School Initiative for RNA Medicine, Harvard Medical School, Boston, MA 02115, USA

[§]These authors contributed equally to this work

Correspondence: daniel.tenen@nus.edu.sg (D.G.T)

1 **Summary**

2 Polycomb Repressive Complex 2 (PRC2) is an epigenetic regulator required for
3 gene silencing during embryonic development. Previous studies have reported that
4 PRC2 interacts with RNA in a promiscuous manner, but the biological functions of
5 such interaction are unknown. Here we present a seesaw mechanism for the
6 regulation of ϵ -globin through inactivating **EZH2** by an upstream non-coding **RNA**
7 (LEVER). We show that LEVER, a non-coding RNA identified by RNA
8 immunoprecipitation sequencing (RIP-seq) of the PRC2 core subunit EZH2 and
9 Nanopore sequencing, binds PRC2 and thereby prevents the accumulation of
10 H3K27 methylation along the genomic region where LEVER RNA is transcribed. The
11 open chromatin within the LEVER locus in turn competes for the chromatin
12 interaction between the ϵ -globin promoter and the Locus Control Region (LCR),
13 working as a negative regulatory element of ϵ -globin expression. Hence, LEVER
14 RNA negatively regulates ϵ -globin by sequestering PRC2 from repressing the
15 LEVER locus, which is a competitor of the ϵ -globin-LCR interaction.

1 Introduction

2 Polycomb Repressive Complex 2 (PRC2) is an epigenetic regulator that plays
3 essential roles in cellular differentiation and early embryonic development. These
4 biological functions are achieved mainly through establishing and maintaining the
5 global methylation landscape of lysine 27 on histone H3. In particular, the final
6 H3K27 tri-methylated product (H3K27me₃) generally marks transcriptionally inactive
7 regions and is considered as the major epigenetic output through which PRC2
8 supervises the repressive genome ¹.

9 However, in mammals, how PRC2 selectively represses targets globally remains
10 elusive. Many long non-coding RNAs (lncRNAs), such as XIST ² and HOTAIR ^{3,4},
11 have been proposed to serve as co-factors to direct PRC2 to its specific targets in
12 the mammalian genome. Importantly, this model highlights a specific lncRNA-PRC2-
13 genome loci axis.

14 This lncRNA-directed PRC2-mediated gene repression model has been
15 questioned by several follow-up studies on PRC2-RNA interaction. First, the
16 promiscuous RNA binding pattern of PRC2 *in vitro* and in cells challenges the
17 selectivity of PRC2 for its specific lncRNA partners ⁵. Second, RNA can inhibit PRC2
18 enzymatic activity ⁶ through competing with histone for PRC2 binding *in vitro* ⁷.
19 Furthermore, extensive interaction between PRC2 and nascent RNAs was observed
20 in mouse embryonic stem cells (ESCs) at transcriptionally active gene loci ^{8,9}, which
21 brings into question the transcription repressive effect by local PRC2-RNA
22 interaction at these loci. In line with these findings, blocking RNA transcription in
23 mouse ESCs increases global PRC2 occupancy and H3K27me₃ levels at previously
24 active gene loci ¹⁰.

1 To investigate the biological functions of PRC2-RNA interactions, we identified
2 robust PRC2-bound RNAs through RIP-seq in K562 cells. This led to the
3 identification of a novel PRC2-interacting non-coding RNA (termed LEVER, short for
4 the regulation of ϵ -globin through inactivating EZH2 by an upstream non-coding
5 RNA), and we further characterized LEVER by direct-RNA Nanopore sequencing.
6 Depletion of LEVER RNA increased PRC2 occupancy and H3K27me2/me3 levels at
7 the LEVER locus. Importantly, down-regulation of LEVER enhanced the expression
8 of the downstream neighboring ϵ -globin gene. Mechanistically, the epigenetic
9 change at the LEVER locus caused by increased PRC2 binding upon LEVER RNA
10 depletion remodeled the local chromatin interaction involving the β -globin Locus
11 Control Region (LCR) and thereby activated ϵ -globin transcription. In summary, our
12 study reveals a model that LEVER RNA sequesters PRC2 from its genome locus
13 and blocks PRC2 activity *in cis*, which remodels local chromatin interaction and
14 regulates neighboring gene expression.

15

16 **Results**

17 **EZH2 binds G/A-rich nascent transcripts**

18 To identify PRC2 associated RNAs, we performed RIP-seq in K562 cells for
19 endogenous EZH2, which is the enzymatic subunit and the subunit with the highest
20 RNA-binding affinity in PRC2⁶. The two well-established PRC2-interacting lncRNAs,
21 XIST² and KCNQ1OT1¹¹, were both consistently enriched in our biological
22 replicates (Supplementary Figure 1A and 1B), validating our EZH2 RIP-seq. 1528
23 high-confidence EZH2-interacting RNA fragments were reproducibly identified from
24 both replicates using a stringent peak-calling algorithm. Annotation of these

1 fragments indicates no preference for transcripts generated from coding genes, non-
2 coding genes, or intergenic regions (Figure 1A). Detailed feature annotation of the
3 fragments from the annotated gene loci further reveals that EZH2 preferentially binds
4 intronic rather than exonic RNAs, disregarding the coding potential of the
5 corresponding gene (Figure 1A). To further study the identity of these EZH2-
6 interacting RNAs, we examined their cellular localization by quantifying their reads-
7 per-million-reads (RPM) values in RNA-seq of nuclear or cytoplasmic fractions of
8 K562 cells. 89% of these EZH2-bound RNA fragments have higher read counts in
9 nuclear than in the cytoplasmic fraction, suggesting their nuclear localization (Figure
10 1B). Further comparison of the RNA-seq read counts between nuclear polyA+ and
11 nuclear polyA- fractions suggests more than 75% of these EZH2-interacting
12 fragments are non-poly-adenylated (Figure 1B). Collectively, these results support
13 previous findings⁹ that EZH2 predominantly binds nascent transcripts in the nuclei.

14 To further investigate whether EZH2 possesses any binding preference for
15 specific RNA sequences, we performed motif discovery analysis for EZH2-enriched
16 RNA fragments with MeMe Suite¹². Two sets of flanking sequences were included in
17 the analysis to eliminate background noise. One motif rich in guanidine (G) and
18 adenosine (A) was significantly enriched from our motif analysis (Figure 1C). This
19 G/A-rich motif revealed by RIP-seq in cells highly agrees with previous results from
20 *in vitro* RNA binding experiments¹³. Collectively, our EZH2 RIP-seq in K562 cells
21 reveals that EZH2 prefers to bind G/A-rich nascent transcripts in cells.

22

23 **EZH2-bound nascent transcripts antagonize PRC2 activity**

1 As nascent transcripts are geometrically close to their corresponding gene loci,
2 nascent transcript-EZH2 interaction may locally influence the histone methylation
3 function of PRC2. To investigate this, we profiled genome-wide H3K27 methylation
4 status as the readout of PRC2 activity at the gene loci where EZH2-interacting
5 transcripts are generated (Figure 1D). We refer to these gene loci as "EZH2-
6 interacting gene loci", while "non-EZH2-interacting gene loci" refer to those in which
7 no EZH2-bound transcript was identified by our RIP-seq. This definition of
8 'interacting' is based on RNA-binding (RIP-seq) instead of the chromatin binding
9 (ChIP-seq) of EZH2. Consistent with the nature of a repressive histone mark, the
10 H3K27me3 level is remarkably higher in the gene body region (from the transcription
11 start site (TSS) to the transcription end site (TES)) of non-expressed genes
12 compared with their flanking regions. In contrast, the H3K27me3 of the EZH2-
13 interacting gene loci showed the opposite pattern, reflecting active transcription in
14 these loci. We further compared the H3K27me3 level of EZH2-interacting gene loci
15 with non-EZH2-interacting gene loci. To minimize the epigenetic bias caused by
16 different gene expression levels, a set of non-EZH2-interacting genes with gene-by-
17 gene matched expression levels as those EZH2-interacting genes were selected for
18 comparison (this set of genes is termed non-EZH2-interacting matched-FPKM,
19 Supplementary Figure 1C). Interestingly, we noticed the EZH2-interacting genes
20 have higher H3K27me3 basal levels in the flanking regions than their FPKM
21 matched non-EZH2-interacting genes and the genomic background (flanking regions
22 of non-EZH2-interacting and non-expressed genes), suggesting their residence in
23 genomic regions that are generally more repressed and with more stable contact
24 with PRC2. Moreover, the decrease of H3K27me3 levels from the flanking regions to
25 the gene body is steeper in the EZH2-interacting gene loci than in the non-EZH2-

1 interacting gene loci, suggesting that RNA-binding may have played a PRC2-
2 inhibitory function. In sum, these observations suggest that EZH2 interacts with RNA
3 within densely H3K27 methylated genomic regions in which the emerging RNA can
4 be crucial in curtailing the repressive activity of PRC2 at the corresponding gene
5 locus.

6 To further study how RNA affects PRC2 chromatin occupancy biochemically, we
7 inhibited global transcription by treating K562 cells with the RNA Polymerase II
8 (Pol II) inhibitor 5, 6-dichloro-1- β -D-ribofuranosylbenzimidazole (DRB) in culture and
9 performed co-immunoprecipitation against PRC2 subunits EZH2 or SUZ12
10 (Supplementary Figure 1D). While lack of RNA did not affect the binding of EZH2 or
11 SUZ12 to other PRC2 subunits (SUZ12/EZH2 and RBBP4), both EZH2 and SUZ12
12 showed increased interaction with lysine 27 tri-methylated histone H3 (H3K27me3)
13 and, to a lesser extent, with total H3 upon transcription blockage (Figure 1E). In line
14 with this observation, the depletion of RNA by RNase A treatment also markedly
15 enhanced PRC2 interaction with H3 and H3K27me3 (Figure 1F). Of note, both DRB
16 and RNase A treated K562 cells showed a global increase in H3K27me3 but not
17 total H3 in the input samples, revealing that RNA depletion enhanced PRC2 binding
18 to the chromatin and led to increased histone methylation levels, supporting previous
19 reports in mESC¹⁴. Collectively, these results strongly support the antagonistic effect
20 of RNA on PRC2 chromatin binding and its methyltransferase activity.

21

22 **PRC2 binds a non-coding RNA upstream of β -globin cluster**

23 Our EZH2 RIP-seq enriched many uncharacterized RNAs transcribed from
24 unannotated intergenic regions (Figure 1A). Among these uncharacterized RNAs, we

1 identified one long transcript (which we termed LEVER) whose transcription start site
2 locates 235kb upstream of the β -globin gene cluster on chromosome 11 (Figure 2A).
3 Transcripts from this locus were highly enriched by RIP-seq with multiple significant
4 peaks (labeled as LVR1-4), and this was further validated by RIP-qPCR of three
5 independent biological replicates (Figure 2B). Notably, LEVER RNA is enriched by
6 EZH2 at a level comparable to that of XIST, which strongly suggests the validity of
7 LEVER as an EZH2-interacting RNA.

8 To characterize the nascent and spliced transcripts from the LEVER locus, we
9 assembled the RNA sequences produced by long-read Nanopore cDNA sequencing
10 of K562 nuclear RNA, as well as Nanopore direct sequencing of K562 total polyA+
11 RNA, which revealed spliced transcripts from this locus (Supplementary Figure 2).
12 By combining RIP-seq enrichment with Nanopore results, we confirmed that EZH2
13 binds multiple intronic rather than exonic RNAs from this region (Figure 2A). Digital
14 droplet PCR of K562 nuclear and cytoplasmic RNA further revealed that the EZH2-
15 bound fragments of LEVER are nuclear-localized (Figure 2C).

16

17 **LEVER RNA sequesters PRC2 and inhibits H3K27 methylation at the LEVER** 18 **locus**

19 Given that loss of nascent transcripts globally increased PRC2-chromatin
20 interactions and PRC2-related histone methylation levels (Figure 1E, 1F), we
21 wondered if loss of specific transcripts would result in a similar observation. We
22 thereby knocked out LEVER using CRISPR-Cas9 in K562 cells. To completely
23 disrupt LEVER RNA generation while preserving corresponding gene loci for
24 following epigenetic studies, we designed single guide RNAs targeting the LEVER

1 promoter region in order to generate small deletions that disrupted RNA Pol II
2 initiation at the LEVER locus (Supplementary Figure 3A). We obtained a LEVER-
3 silenced clone with a homozygous deletion (270bp) spanning the TSS site
4 (Figure 3A, Supplementary Figure 3B). Loss of LEVER transcripts increased EZH2
5 occupancy at the locus (Figure 3B), which was accompanied by markedly elevated
6 H3K27me3 levels at the same loci (Figure 3C). In contrast, no noticeable increase in
7 either EZH2 or H3K27me3 was detected at MYT1 (serving as a positive control for
8 EZH2 and H3K27me3 ChIP) or CCDC26 (negative control) regions, suggesting the
9 PRC2 related chromatin changes at the LEVER locus is site-specific upon LEVER
10 RNA knock-out. To validate the above observations, we further constructed a
11 doxycycline-inducible LEVER knock-down K562 line using the CRISPRi-dCas9
12 system. Of note, the dCas9 we used in our CRISPRi¹⁵ experiments is not KRAB-
13 fused¹⁶, and thus the RNA knock-down effect is achieved by blockage of Pol II at the
14 LEVER promoter without introducing any direct epigenetic modifications
15 (Supplementary Figure 3A). Consistently, inducible knock-down of LEVER (Figure
16 3D) remodeled the PRC2-related epigenetic features along the LEVER locus but not
17 nearby regions. Histone ChIP-seq showed increased H3K27me2/me3 and
18 decreased H3K27ac at the LEVER locus (Figure 3E), suggesting locally enhanced
19 PRC2 activity and reduced chromatin accessibility. These ChIP-seq results were
20 further validated at selected regions by qPCR (Supplementary Figure 3C, 3D, 3E).
21 Collectively, the depletion of LEVER transcripts specifically increased PRC2
22 occupancy and its progressive H3K27 methylation products on the local chromatin.
23 This gene-specific observation reveals that LEVER RNA regulates the homeostasis
24 of PRC2 related histone modifications *in cis* and is consistent with the above
25 observations from global RNA depletion (Figure 1D, 1E, 1F).

1

2 **LEVER RNA negatively regulates ϵ -globin**

3 LEVER-PRC2 interaction and its epigenetic effects *in cis* may regulate local gene
4 transcription. Due to the genomic proximity of LEVER to the β -globin cluster, we
5 examined the regulation of expression of ϵ -globin, the β -globin member adjacent to
6 LEVER (Figure 4A), by LEVER RNA. Knock-out and knock-down of LEVER RNA
7 upregulated ϵ -globin mRNA levels (Figure 4B), and this increase was further
8 observed at the protein level (Figure 4C). In contrast, γ -globin mRNA level was not
9 affected by the loss of LEVER (Supplementary Figure 4A). Of note, although
10 lncRNAs have been reported to affect mRNA stability¹⁷, the increase of ϵ -globin
11 mRNA abundance here was achieved by transcriptional activation, as reflected by
12 increased RNA Pol II occupancy at ϵ -globin promoter upon LEVER knock-down
13 (Figure 4D).

14 To further validate the regulatory effect of LEVER RNA on ϵ -globin expression,
15 we over-expressed two RNA fragments encompassing the consensus EZH2-bound
16 guanidine-rich motif (Figure 2A) within LVR-1 and LVR-2 regions (termed LVR-1
17 MeMe and LVR-2 MeMe respectively) independently in LEVER KO cells. Restored
18 expression of EZH2-interacting LEVER fragments (Supplementary Figure 4B)
19 mitigated ϵ -globin activation at mRNA (Figure 4E) and protein levels (Figure 4F),
20 showing an inverse correlation and supports the negative regulatory effect of LEVER
21 RNA on ϵ -globin.

22 Moreover, we observed that the de-repression of LEVER RNA upon doxycycline
23 withdrawal was associated with a reversible decrease in ϵ -globin RNA (Figure 5A;

1 Supplementary Figure 4C). In sum, these dynamic studies clearly support a negative
2 regulatory effect of LEVER RNA on ϵ -globin expression.

3

4 **Hypoxia regulates LEVER and ϵ -globin**

5 The embryonic form of β -globin, ϵ -globin, is expressed almost exclusively in the early
6 human embryo in which the environment is hypoxic¹⁸. To study whether LEVER can
7 be a *bona fide* regulator of ϵ -globin, we examined the expression homeostasis
8 between these two under environmental perturbations, e.g., oxygen levels. Culturing
9 K562 cells in hypoxic conditions (1% oxygen) repressed LEVER while enhanced ϵ -
10 globin expression over time (Figure 5B, 5C), which reproduced the inverse-
11 correlation between LEVER and ϵ -globin expression. Of note, LEVER RNA level was
12 more drastically changed in terms of fold difference than ϵ -globin on the first day of
13 hypoxic culture, suggesting LEVER as an upstream regulator of ϵ -globin.

14

15 **The LEVER locus regulates ϵ -globin through long-range chromatin** 16 **interactions**

17 Although we have demonstrated that LEVER RNA inhibits EZH2 function at its
18 own locus *in cis*, LEVER RNA may also act directly through PRC2 to repress the ϵ -
19 globin locus *in trans*. We thereby examined the H3K27me3 status at the ϵ -globin
20 locus. We did not detect noticeable H3K27me3 coverage at the ϵ -globin locus at
21 baseline. This observation suggests that ϵ -globin is not a direct repressive target of
22 PRC2 mediated by H3K27 methylation in K562 cells.

1 To understand the role of EZH2 and LEVER in ϵ -globin transcriptional regulation,
2 we knocked out EZH2 in K562 cells. Interestingly, ϵ -globin expression was reduced
3 in EZH2 null cells in which global H3K27me3 was not detectable by western blot
4 (Figure 6A, 6B). Re-expression of wild type EZH2 and its activating mutant
5 (EZH2^{Y641F})^{19, 20} in the knock-out cells restored ϵ -globin expression, while the
6 truncated EZH2 mutant without methyltransferase motif (EZH2 ^{Δ SET}) failed to do so
7 (Figure 6A, 6B). In addition, the EZH2 enzymatic inhibitor GSK343 curtailed ϵ -globin
8 up-regulation in LEVER knock-out cells (Supplementary Figure 5). In summary, this
9 data suggests that this ϵ -globin activating effect requires EZH2 methyltransferase
10 activity. Moreover, it appears that EZH2 function affected H3K27 methylation within
11 the LEVER locus but not at the ϵ -globin gene. These results collectively suggest that
12 EZH2 and LEVER regulate ϵ -globin in K562 cells.

13 The β -globin cluster is well known to be regulated by long-range chromatin
14 interactions. Therefore, we asked whether the LEVER locus itself may work as a
15 negative regulatory element of the ϵ -globin gene through long-range chromatin
16 interactions. To test this hypothesis, we performed 4C-seq analysis before and after
17 LEVER promoter knock-out to compare local chromatin interaction changes upon
18 LEVER RNA loss. We positioned two viewpoints at the ϵ -globin promoter, and
19 another two viewpoints at HS3 and HS4 regions within the β -globin Locus Control
20 Region (LCR), which are well-established activating elements of the β -globin cluster
21 ²¹ (Figure 6C). As expected, our 4C-seq analysis suggested that the ϵ -globin
22 regulatory domains, including both the ϵ -globin promoter and the two LCR regions,
23 interact weakly but broadly to LEVER chromatin in K562 cells. Loss of LEVER RNA
24 resulted in the reduction of such interaction (Figure 6C, Supplementary Figure 6),

1 presumably due to the compromised accessibility of LEVER locus caused by PRC2
2 methylation. In addition, the interaction between ϵ -globin promoter and LCR regions
3 was significantly enhanced upon LEVER KO, as supported by the results from both
4 viewpoints (Figure 6C, Supplementary Figure 6). Together with the increased Pol II
5 occupancy at ϵ -globin promoter (Figure 4D), this observation suggests that the
6 strengthened ϵ -globin promoter-LCR enhancer interaction could be the cause of ϵ -
7 globin transcriptional activation upon LEVER RNA loss. Since the LEVER locus
8 interacts with both the ϵ -globin promoter and the LCR, we propose that it may work
9 as a competitor of the ϵ -globin promoter for LCR binding, thereby exerting its
10 negative regulatory effect on ϵ -globin expression. Collectively, our 4C-seq results
11 from different viewpoints support LEVER RNA as a negative regulator of ϵ -globin.

12

13 **Discussion**

14 Following the initial discovery of the specific PRC2-lncRNA interaction and its
15 supportive effect on PRC2 functions at PRC2 target genes^{2-4, 11}, more pieces of
16 evidence revealed a promiscuous PRC2-nascent RNA interaction that inhibits local
17 PRC2 function^{5-8, 14}. While previous studies of the latter viewpoint were performed
18 either *in vitro* or exclusively in mESC, our study in the human chronic myelogenous
19 leukemia cell line K562 provides further evidence to support this model. In summary,
20 we demonstrated that PRC2 manifests genome-wide interaction with nuclear-
21 localized, non-polyA tailed, guanidine-rich intronic RNAs that are transcribed from
22 active gene loci. These genes, as suggested by local H3K27me3 ChIP-seq profiling,
23 tend to localize within densely H3K27 methylated genomic regions, but they are
24 rather sparsely methylated within their gene body. These observations suggest that

1 the PRC2-sequestering function of their nascent transcripts may play a role in
2 antagonizing PRC2 repression *in cis* within PRC2-dense environment. Following the
3 global analysis, we further identified and characterized the LEVER locus that
4 generates PRC2-binding nascent transcripts near the β -globin cluster. Using both
5 CRISPR knock-out and inducible CRISPRi knock-down systems, we showed that
6 LEVER RNA prevents PRC2 from occupying the chromatin and thereby curtails
7 EZH2 methyltransferase activity at the LEVER locus (Figure 7). These observations
8 made at single gene locus further support the model generalized from global
9 analysis that nascent transcripts deprive PRC2 from the chromatin and block PRC2
10 function locally.

11 Our study further reveals a novel model of epigenetic regulation on ϵ -globin
12 expression (Figure 7). Our results showed that in the presence of the LEVER RNA,
13 the LEVER locus competes with the ϵ -globin promoter to bind with the LCR, a well-
14 established enhancer region for the β -globin cluster ^{21, 22}, and thereby suppresses ϵ -
15 globin transcription. In the absence of LEVER RNA, the released PRC2 occupies the
16 LEVER chromatin, facilitates H3K27 methylation, and reduces the accessibility of the
17 LEVER locus. This in turn permits the interaction between the LCR and the ϵ -globin
18 promoter, which promotes the transcription of ϵ -globin. In sum, we demonstrated that
19 the LEVER RNA can activate the LEVER locus that works as a negative regulatory
20 element of ϵ -globin gene expression through chromatin looping.

21 Interestingly, we found that transcripts generated from similar genomic regions
22 upstream of the β -globin cluster in the mouse genome were previously identified and
23 annotated as polycomb-associated ncRNAs, similar to LEVER ²³. But the two loci
24 from the two species share poor sequence homology (data not shown). We wonder

1 whether the regulatory effect of the corresponding LEVER loci and RNAs on β -globin
2 cluster are functionally conserved in mouse although the loci are poorly conserved.
3 Considering the promiscuous RNA binding pattern of PRC2, some PRC2-binding
4 non-coding RNAs may have evolved to become opportunistic regulatory elements
5 that are epigenetically functional to neighboring genes, and thereby conserved their
6 relative position to neighboring protein-coding genes without the need to preserve
7 their sequences in different species. The syntenic transcription of lncRNA has been
8 proposed together with conserved sequence, structure, and function as the four
9 dimensions of lncRNA conservation²⁴. Our model in PRC2-nascent RNA interaction
10 and its regulatory function on nearby genes may explain such synteny conservation
11 observed in some lncRNAs.

12 In addition, our identification of the LEVER locus as a negative regulator of ϵ -
13 globin shows potential clinical significance. β -hemoglobinopathies, including sickle
14 cell disease (SCD) and β -thalassemia, are caused by genetic mutations in the adult-
15 form β -globin gene. They are among the most common genetic diseases and
16 influence tens of millions of patients worldwide²⁵. Reactivation of the fetal γ -globin
17 gene has been demonstrated to ameliorate the severity of these diseases²⁶.
18 Therefore, much effort has been made to understand the molecular basis of γ -globin
19 regulation and to develop therapeutic strategies to reactivate γ -globin expression in
20 β -hemoglobinopathy patients^{27, 28}. However, the embryonic-form globin, ϵ -globin, is
21 much less studied than its fetal counterpart²⁹, even though limited studies indeed
22 suggested the therapeutic potential of ϵ -globin reactivation in the SCD mouse model,
23 similar to γ -globin^{30, 31}. In this work, we provided an in-depth understanding of the
24 regulation of ϵ -globin by its upstream LEVER loci through PRC2. More importantly,

1 we have achieved multi-fold ϵ -globin upregulation through knocking-out and
2 knocking-down the non-coding transcript from LEVER. This moderate enhancement
3 of ϵ -globin expression, however, may ameliorate red blood cell sickling and thereby
4 mitigate SCD symptoms as suggested by previous studies^{30, 31}. Taken together, our
5 findings suggest a novel therapeutic strategy of using epigenetic modulators, RNAi,
6 or even gene editing to reactivate ϵ -globin for the treatment of β -hemoglobinopathy
7 patients.

8
9

- 10 1. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature*
11 **469**, 343-349 (2011).
- 12 2. Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J. & Lee, J.T. Polycomb proteins targeted by a
13 short repeat RNA to the mouse X chromosome. *Science* **322**, 750-756 (2008).
- 14 3. Rinn, J.L. *et al.* Functional demarcation of active and silent chromatin domains in
15 human HOX loci by noncoding RNAs. *Cell* **129**, 1311-1323 (2007).
- 16 4. Gupta, R.A. *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to
17 promote cancer metastasis. *Nature* **464**, 1071-1076 (2010).
- 18 5. Davidovich, C., Zheng, L., Goodrich, K.J. & Cech, T.R. Promiscuous RNA binding by
19 Polycomb repressive complex 2. *Nat Struct Mol Biol* **20**, 1250-1257 (2013).
- 20 6. Cifuentes-Rojas, C., Hernandez, A.J., Sarma, K. & Lee, J.T. Regulatory interactions
21 between RNA and polycomb repressive complex 2. *Mol Cell* **55**, 171-185 (2014).
- 22 7. Wang, X. *et al.* Molecular analysis of PRC2 recruitment to DNA in chromatin and its
23 inhibition by RNA. *Nat Struct Mol Biol* **24**, 1028-1038 (2017).
- 24 8. Kaneko, S., Son, J., Bonasio, R., Shen, S.S. & Reinberg, D. Nascent RNA interaction
25 keeps PRC2 activity poised and in check. *Genes Dev* **28**, 1983-1988 (2014).
- 26 9. Kaneko, S., Son, J., Shen, S.S., Reinberg, D. & Bonasio, R. PRC2 binds active promoters
27 and contacts nascent RNAs in embryonic stem cells. *Nat Struct Mol Biol* **20**, 1258-1264
28 (2013).
- 29 10. Riising, E.M. *et al.* Gene silencing triggers polycomb repressive complex 2 recruitment
30 to CpG islands genome wide. *Mol Cell* **55**, 347-360 (2014).
- 31 11. Pandey, R.R. *et al.* Kcnq1ot1 antisense noncoding RNA mediates lineage-specific
32 transcriptional silencing through chromatin-level regulation. *Mol Cell* **32**, 232-246
33 (2008).
- 34 12. Bailey, T.L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids*
35 *Res* **37**, W202-208 (2009).
- 36 13. Wang, X. *et al.* Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats
37 of Consecutive Guanines. *Mol Cell* **65**, 1056-1067 e1055 (2017).
- 38 14. Beltran, M. *et al.* The interaction of PRC2 with RNA or chromatin is mutually
39 antagonistic. *Genome research* **26**, 896-907 (2016).

- 1 15. Qi, L.S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific
2 control of gene expression. *Cell* **152**, 1173-1183 (2013).
- 3 16. Thakore, P.I. *et al.* Highly specific epigenome editing by CRISPR-Cas9 repressors for
4 silencing of distal regulatory elements. *Nat Methods* **12**, 1143-1149 (2015).
- 5 17. He, R.Z., Luo, D.X. & Mo, Y.Y. Emerging roles of lncRNAs in the post-transcriptional
6 regulation in cancer. *Genes Dis* **6**, 6-15 (2019).
- 7 18. Simon, M.C. & Keith, B. The role of oxygen availability in embryonic development and
8 stem cell function. *Nat Rev Mol Cell Biol* **9**, 285-296 (2008).
- 9 19. Sneeringer, C.J. *et al.* Coordinated activities of wild-type plus mutant EZH2 drive
10 tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-
11 cell lymphomas. *Proc Natl Acad Sci U S A* **107**, 20980-20985 (2010).
- 12 20. Yap, D.B. *et al.* Somatic mutations at EZH2 Y641 act dominantly through a mechanism
13 of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. *Blood*
14 **117**, 2451-2459 (2011).
- 15 21. Levings, P.P. & Bungert, J. The human beta-globin locus control region. *Eur J Biochem*
16 **269**, 1589-1599 (2002).
- 17 22. Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction
18 between hypersensitive sites in the active beta-globin locus. *Mol Cell* **10**, 1453-1465
19 (2002).
- 20 23. Lee, J.T., Zhao, J., Sarma, K., Borowsky, M. & Ohsumi, T.K. (Google Patents, 2016).
- 21 24. Diederichs, S. The four dimensions of noncoding RNA conservation. *Trends Genet* **30**,
22 121-123 (2014).
- 23 25. Weatherall, D.J. The inherited diseases of hemoglobin are an emerging global health
24 burden. *Blood* **115**, 4331-4336 (2010).
- 25 26. Olivieri, N.F. & Weatherall, D.J. The therapeutic reactivation of fetal haemoglobin.
26 *Hum Mol Genet* **7**, 1655-1658 (1998).
- 27 27. Wilber, A., Nienhuis, A.W. & Persons, D.A. Transcriptional regulation of fetal to adult
28 hemoglobin switching: new therapeutic opportunities. *Blood* **117**, 3945-3953 (2011).
- 29 28. Ghiaccio, V., Chappell, M., Rivella, S. & Breda, L. Gene Therapy for Beta-
30 Hemoglobinopathies: Milestones, New Therapies and Challenges. *Mol Diagn Ther* **23**,
31 173-186 (2019).
- 32 29. Kingsley, P.D. *et al.* "Maturation" globin switching in primary primitive erythroid cells.
33 *Blood* **107**, 1665-1672 (2006).
- 34 30. He, Z. & Russell, J.E. Expression, purification, and characterization of human
35 hemoglobins Gower-1 ($\zeta 2\epsilon 2$), Gower-2 ($\alpha 2\epsilon 2$), and Portland-2 ($\zeta 2\beta 2$) assembled in
36 complex transgenic-knockout mice. *Blood* **97**, 1099-1105 (2001).
- 37 31. He, Z. & Russell, J.E. A human embryonic hemoglobin inhibits Hb S polymerization in
38 vitro and restores a normal phenotype to mouse models of sickle cell disease. *Proc*
39 *Natl Acad Sci U S A* **99**, 10635-10640 (2002).

Figure 1. Nascent RNA binds and sequesters EZH2 from histone H3 in K562 cells.

(A) EZH2 preferentially binds intronic rather than exonic RNAs.

Upper panel: Percentage of EZH2 RIP-seq enriched RNA fragments annotated to coding genes, non-coding genes, or intergenic regions. The number of identified peaks annotated to each category are labelled in the brackets in the pie-chart legends.

Lower panel: Percentage of EZH2 RIP-seq enriched RNA fragments originated from coding or non-coding genes annotated to different gene features. UTR, untranslated regions; CDS, coding sequences; ncRNA, non-coding RNA.

(B) EZH2 predominantly binds nuclear polyA- nascent transcripts.

Left panel: Scatter plot showing normalized read counts in nuclear (x-axis) and cytoplasmic (y-axis) fractioned RNA-seq for each EZH2 RIP-seq enriched RNA fragment. The read counts are normalized to total mapped reads as RPM (reads per million mapped reads). Enrichment of dots below the labeled diagonal reveals higher RNA-seq reads in the nuclear than in the cytoplasmic RNA fraction for most of the EZH2 RIP-seq fragments.

Right panel: similar as in left, except that the read counts of the identified RNA fragments are quantified in the nuclear polyA- (x-axis) and nuclear polyA+ (y-axis) fractions.

(C) RNA motif identified from a set of 900 EZH2-bound RNA fragments by MeMe motif discovery analysis.

(D) Read coverage profiles for H3K27me3 ChIP-seq spanning the upstream, gene body, and downstream regions of the following gene sets in K562 cells:

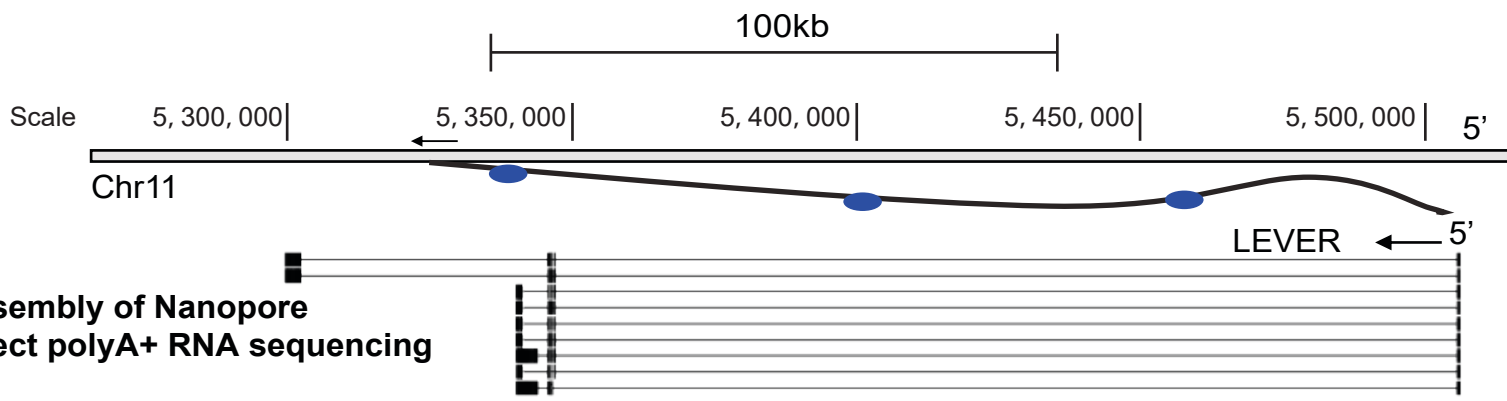
Figure 1. Nascent RNA binds and sequesters EZH2 from histone H3 in K562 cells. (continued)

Red line: EZH2-interacting genes; blue line: non-EZH2-interacting genes with matched expression level as EZH2-interacting genes; green line: total non-EZH2 interacting expressed genes; and yellow line: non-expressed genes. Notably, the "EZH2-interacting" stated here is based on the RNA-binding as identified by EZH2 RIP-seq rather than ChIP-seq (chromatin binding) of EZH2 at these gene loci. Shown in the X-axis is distance from the transcription start site (TSS) to the transcription end site (TES).

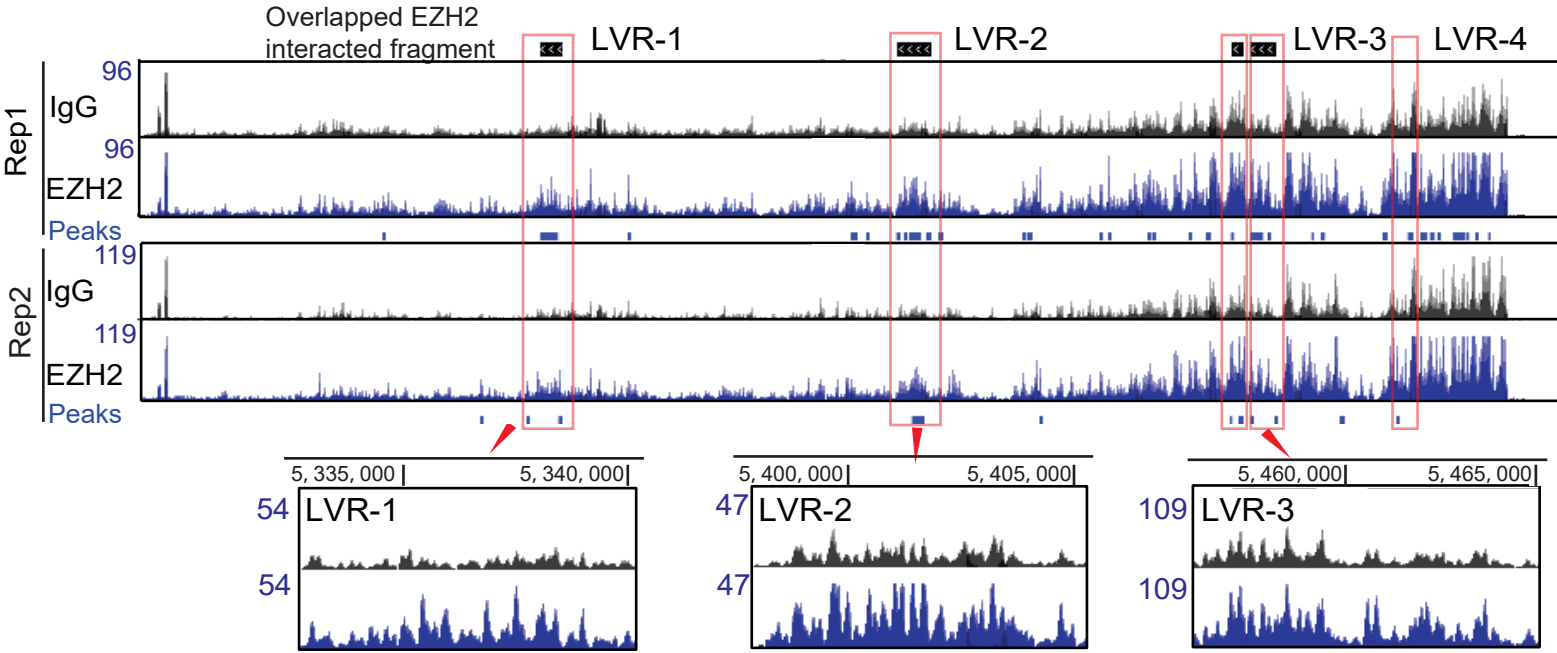
(E) (F) EZH2 or SUZ12 co-immunoprecipitation (co-IP) in K562 cells treated with (E) 5,6-dichloro-1- β -D-ribofuranosylbenzimidazole (DRB) or (F) RNase A. Co-IP of PRC2 subunits (EZH2, SUZ12, RBBP4), PRC2-interacting protein (DNMT1), and histones (total histone H3 and histone H3 with H3K27me3 methylation) were measured by western blot. IgG isotype control antibody was used as co-IP negative control.

Figure 2

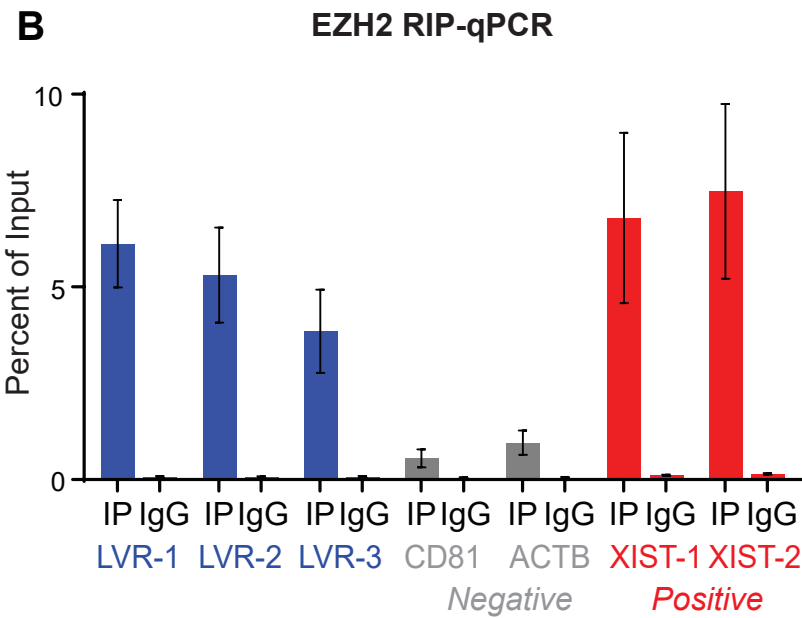
A



EZH2 RIP-seq



B



C

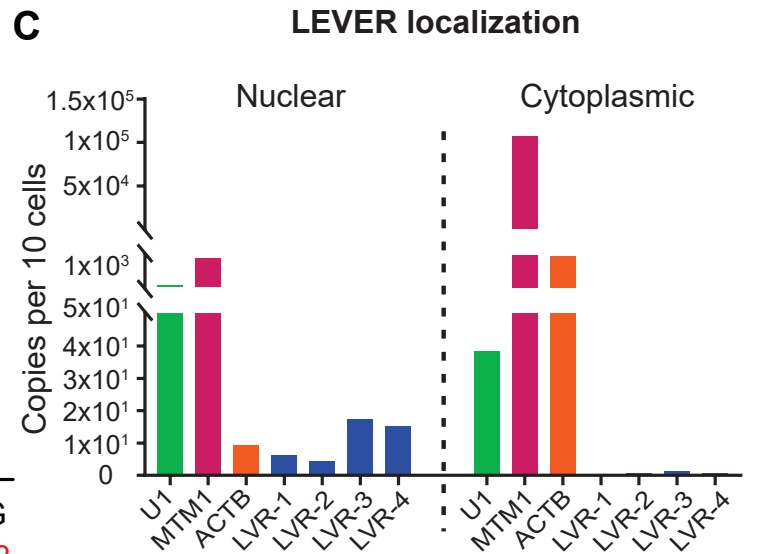


Figure 2. EZH2 binds a nascent transcript at LEVER locus.

- (A) EZH2 RIP-seq and Nanopore direct sequencing of total polyA+ RNA in K562 cells demonstrating EZH2 binding intronic RNAs from the LEVER locus (pictured in supplementary Figure 2). In the Nanopore track, the bars represent genomic regions covered by actual Nanopore reads, and they are linked by lines to represent splice junctions. In the EZH2 RIP-seq tracks, the EZH2 enriched RNA peaks (blue bars below signal tracks of each replicate) are called by ≥ 2 -fold read coverage of EZH2 over IgG with Poisson distribution p -value $\leq 10^{-5}$. Peaks generated from two biological replicates are further overlapped and merged as "Overlapped EZH2 interacted fragments" (black bars above the RIP-seq tracks). Selected overlapped fragments are highlighted in red boxes with their read-coverage tracks shown in a zoomed-in view. The blue ovals in the top diagram represent the relative positions of LVR-1, LVR-2, and LVR-3.
- (B) EZH2 RIP-qPCR on three LEVER fragments (LVR-1, LVR-2, and LVR-3, highlighted in Figure 2A), CD81 and ACTB (RIP-negative controls) and two XIST fragments (RIP-positive controls), quantified as percent of total input. Data from three biological replicates are shown.
- (C) Digital droplet RT-PCR of LEVER fragments (LVR-1 through LVR-4), U1 and MTM1 (nuclear enriched), and ACTB (cytoplasmic enriched) RNAs in K562 nuclear or cytoplasmic RNA fractions.

Figure 3

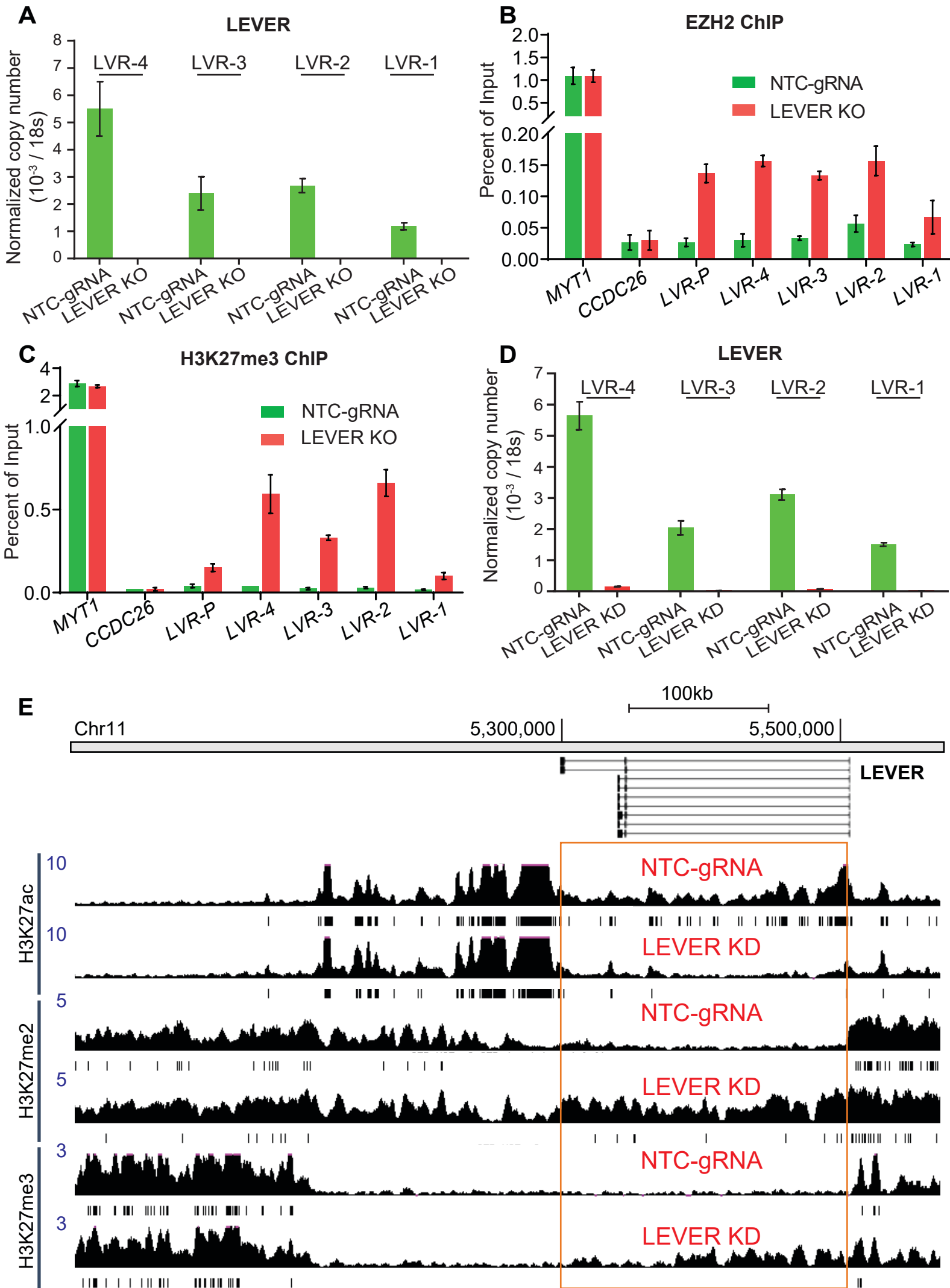


Figure 3. LEVER RNA sequesters EZH2 and blocks histone H3K27 methylation at the LEVER locus *in cis*.

- (A) LEVER RNA expression measured by primers detecting four selected regions (marked in Figure 2A) in non-targeting control (NTC-gRNA) or LEVER knock-out (LEVER KO) K562 cells. LEVER RNA levels were below the level of detection following LEVER knock-out.
- (B) (C) EZH2 (B) or H3K27me3 (C) chromatin occupation at selected regions within the LEVER locus. MYT1 and CCDC26 serve as ChIP positive and negative controls, respectively. LVR-P primer pair is designed at the LEVER promoter region (Supplementary Figure 3A).
- (D) LEVER RNA expression in doxycycline (dox) inducible non-targeting control (NTC-gRNA) or LEVER knock-down (LEVER KD) K562 cells treated with dox for 28 days quantified by RT-qPCR.
- (E) H3K27ac, H3K27me2, and H3K27me3 ChIP-seq at the LEVER locus in dox-inducible non-targeting control (NTC-gRNA) or LEVER knock-down (LEVER KD) K562 cells treated with dox for 28 days.

Figure 4

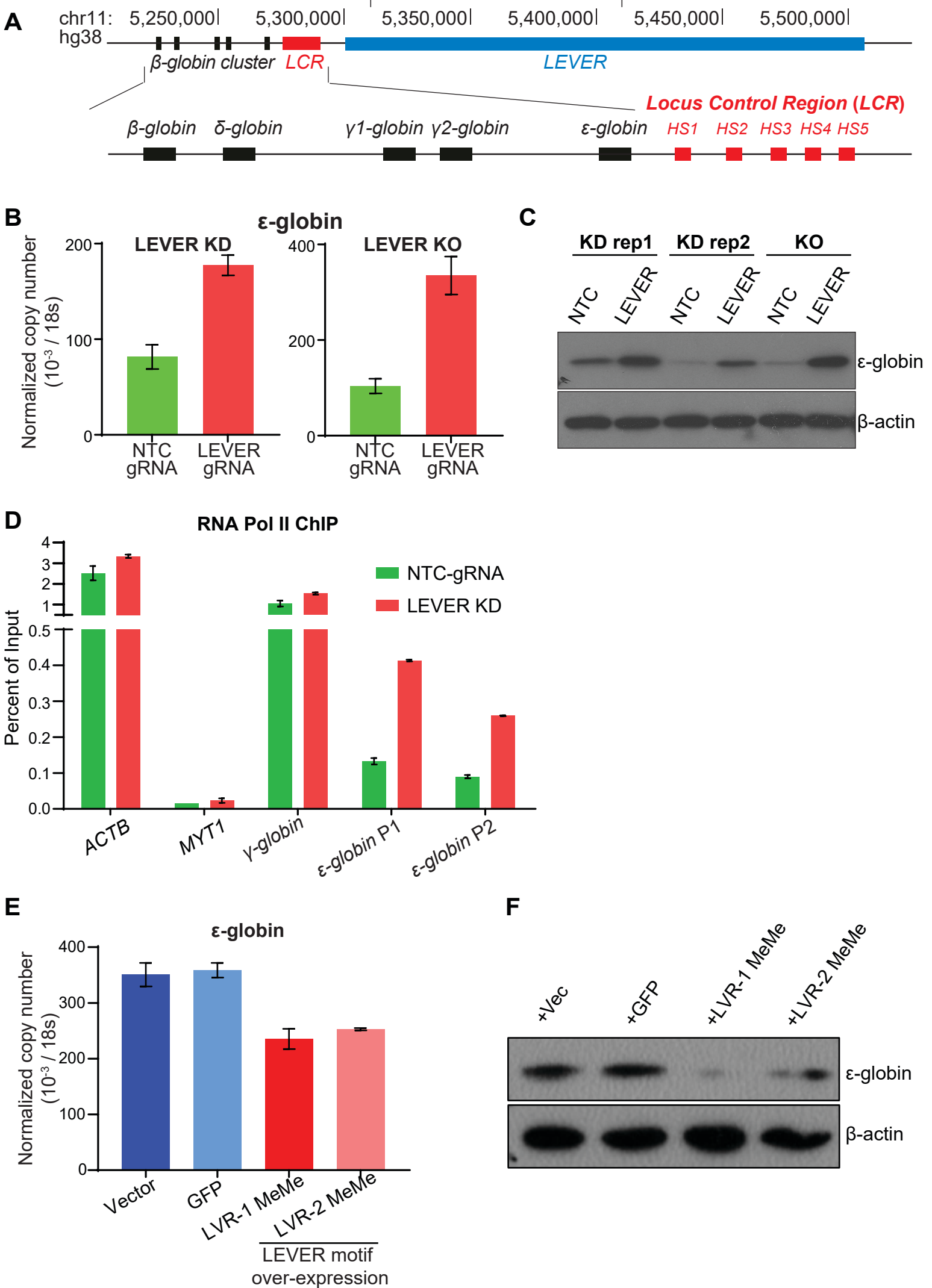


Figure 4. LEVER RNA negatively regulates ϵ -globin.

- (A) Chromosome map of the β -globin cluster and LEVER locus. Transcription of LEVER (in blue) and the globin genes (in black) proceeds in a right to left direction. HS1 through HS5 (in red) correspond to the hypersensitive regions in the LCR.
- (B) ϵ -globin expression after inducible LEVER knock-down (KD) or LEVER knock-out (KO) in K562 cells measured by RT-qPCR.
- (C) ϵ -globin protein abundance in LEVER KD or KO K562 cells measured by western blot. β -actin serves as the loading control.
- (D) ChIP-qPCR of RNA Pol II at ϵ -globin and γ -globin promoter in non-targeting control (NTC-gRNA) and LEVER knock-down (LEVER KD) cells. ACTB and MYT1 serve as Pol II ChIP positive and negative controls, respectively. Pol II occupation at ϵ -globin promoter is measured by two primer sets (ϵ -globin P1 and ϵ -globin P2).
- (E, F) ϵ -globin expression in LEVER KO K562 cells rescued by LVR-1 MeMe and LVR-2 MeMe. LEVER KO K562 cells were transduced with empty vector, GFP, LVR-1 MeMe, or LVR-2 MeMe expressing lentivirus and analysed for ϵ -globin expression using RT-qPCR (E) and western blot (F). LVR-1 and LVR-2 MeMe represent RNA fragments encompassing MeMe enriched EZH2-interacting RNA motif (Figure 1C) within the LVR-1 and LVR-2 regions.

Figure 5

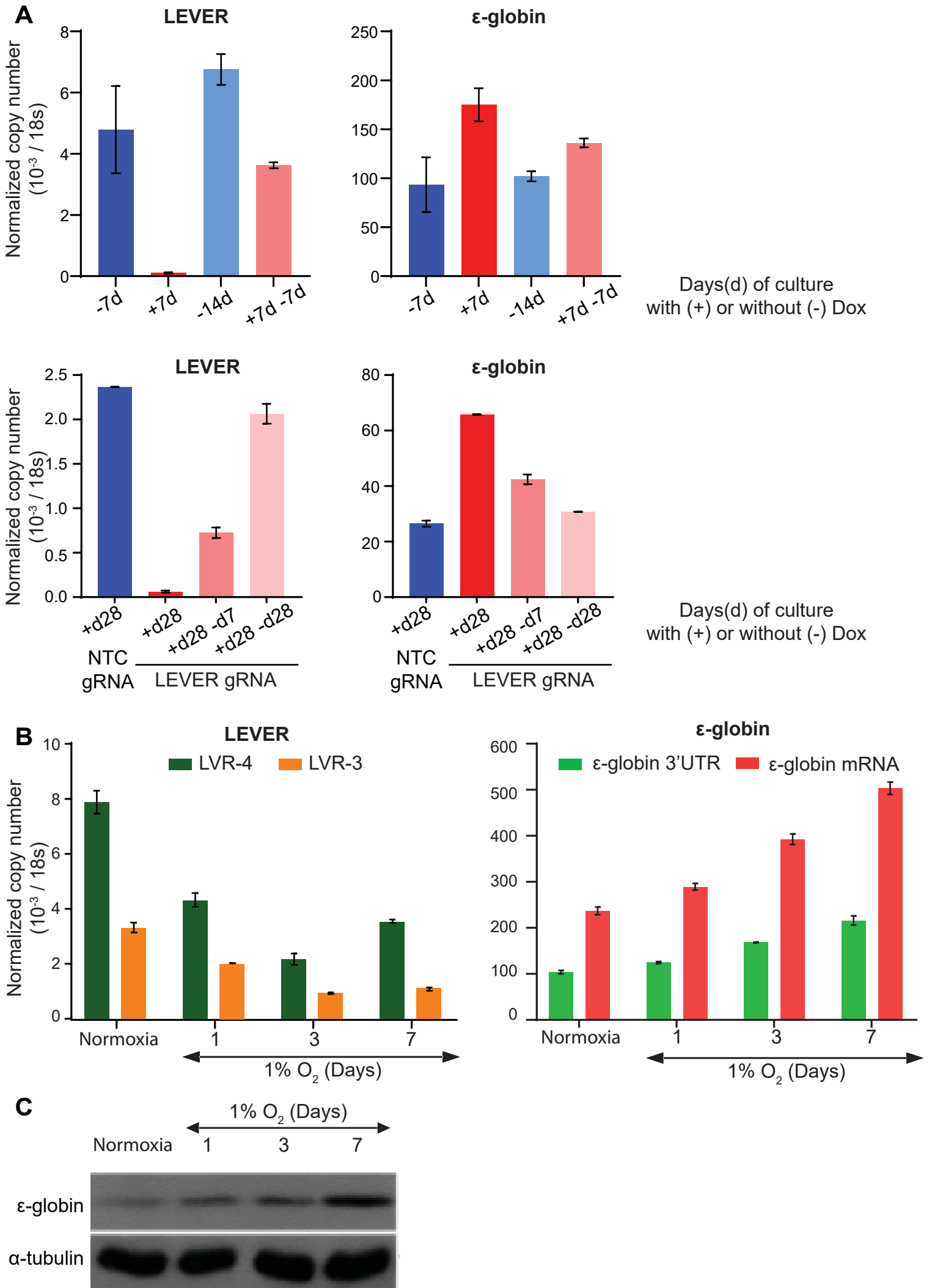
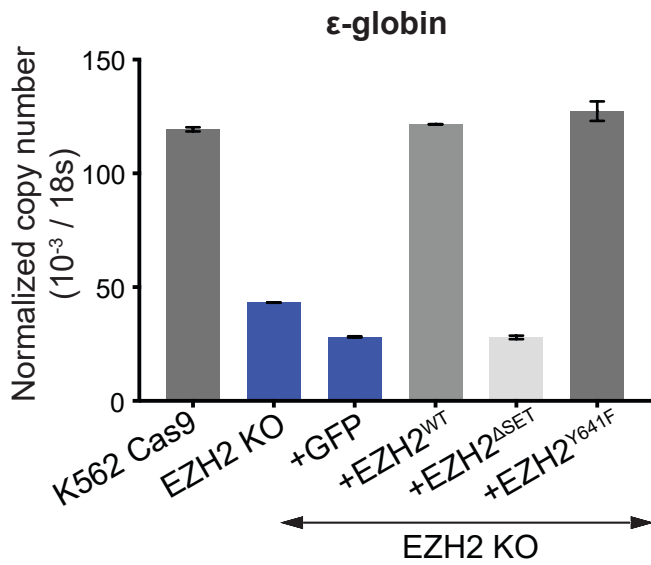


Figure 5. Dynamic LEVER regulation validates inverse correlation between LEVER and ϵ -globin expression.

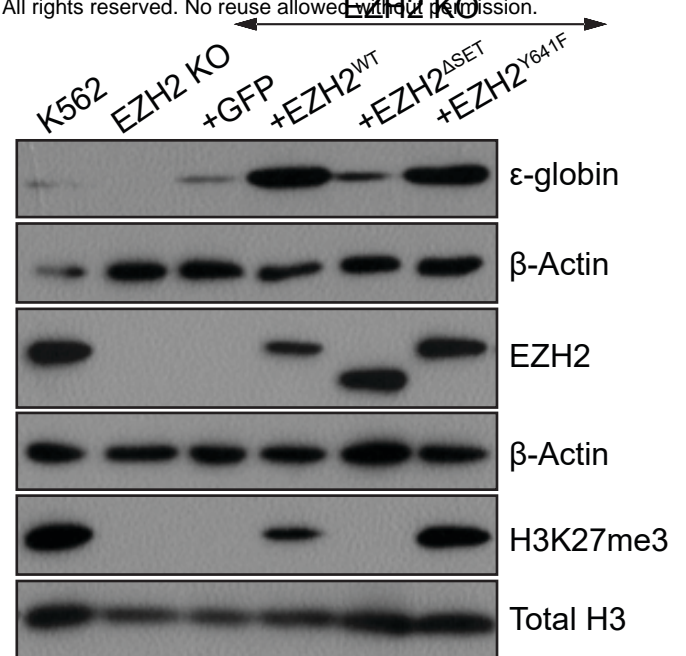
- (A) LEVER and ϵ -globin expression in inducible LEVER KD K562 cells with (+) or without (-) doxycycline treatment for the indicated duration. LEVER expression is quantified at the LVR-4 region.
- (B) Time course RT-qPCR analysis of LEVER RNA and ϵ -globin in K562 cells during 7-day hypoxic culture under 1% oxygen.
- (C) ϵ -globin protein abundance in K562 cells cultured under 1% oxygen. α -tubulin serves as the loading control.

Figure 6

A



B



C

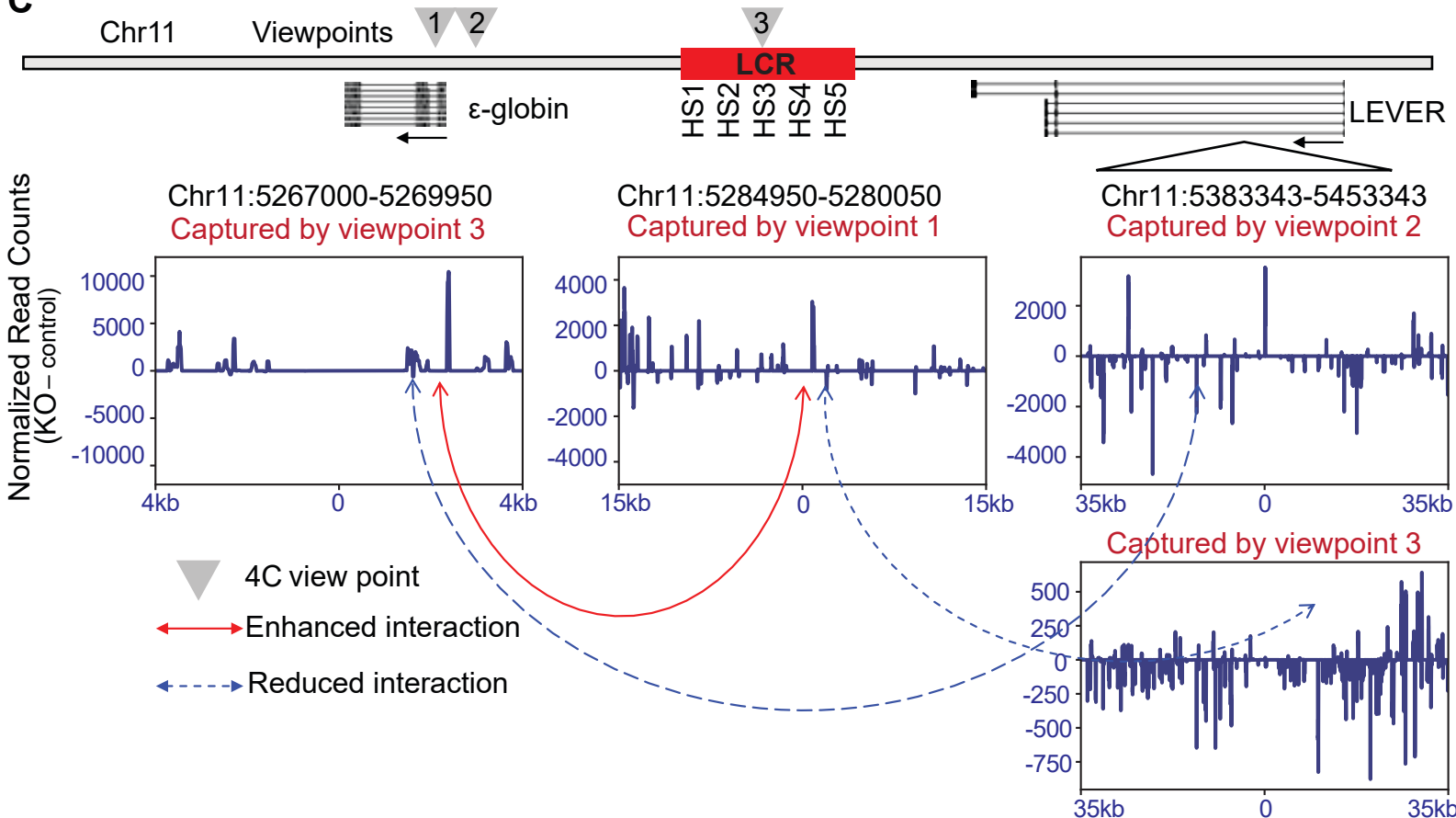


Figure 6. The LEVER locus regulates ϵ -globin through long-range chromatin interactions.

- (A) RT-qPCR analysis of ϵ -globin in K562 Cas9-expressing parental cells, EZH2 knock-out (EZH2-KO) cells, and EZH2-KO cells rescued with GFP, EZH2, EZH2 Δ SET and EZH2^{Y641F}. EZH2 Δ SET, truncated EZH2 without methyltransferase enzymatic domain; EZH2^{Y641F}, EZH2 gain of methyltransferase enzymatic function mutation. The EZH2-KO cells were transduced with lentivirus to stably express these constructs.
- (B) Western blot analysis of ϵ -globin, EZH2, and global H3K27me3 in K562 wild type, EZH2 KO, and rescue cells as in (A). β -actin and total histone H3 were used as loading controls.
- (C) 4C analysis comparing local chromatin interactions near the β -globin cluster in LEVER KO and NTC cells shows increased ϵ -globin-LCR interaction upon LEVER knock-out. We positioned two viewpoints at the ϵ -globin promoter and two viewpoints at HS3 and HS4 regions within the β -globin locus control region (LCR). The viewpoints at ϵ -globin promoters and HS3 are shown here and indicated as viewpoints 1-3, respectively. Captured genomic fragments displayed in the windows correspond to the genomic coordinates annotated above each window. Differentially interacted regions are shown in subtracted reads by comparing LEVER KO and NTC cells, thereby positive or negative values represent increased (solid red line) or decreased (dotted blue lines) interactions upon LEVER KO, respectively.

Figure 7

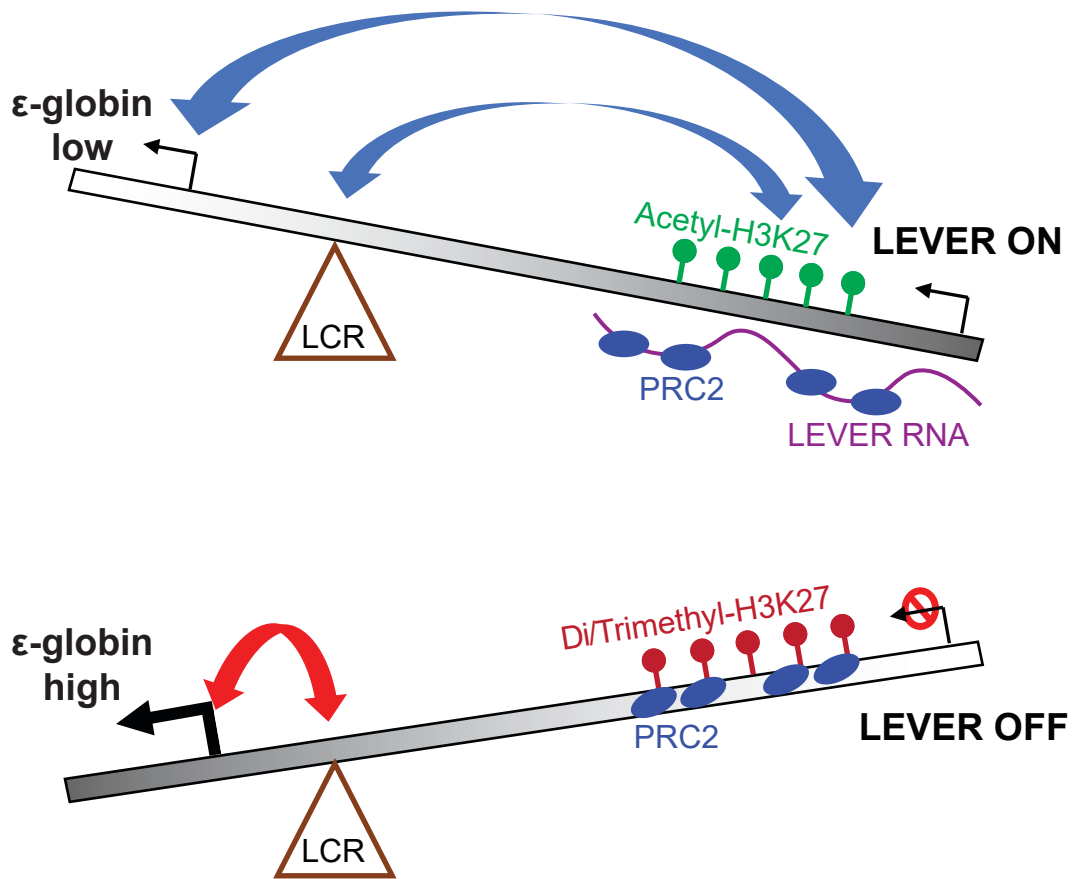
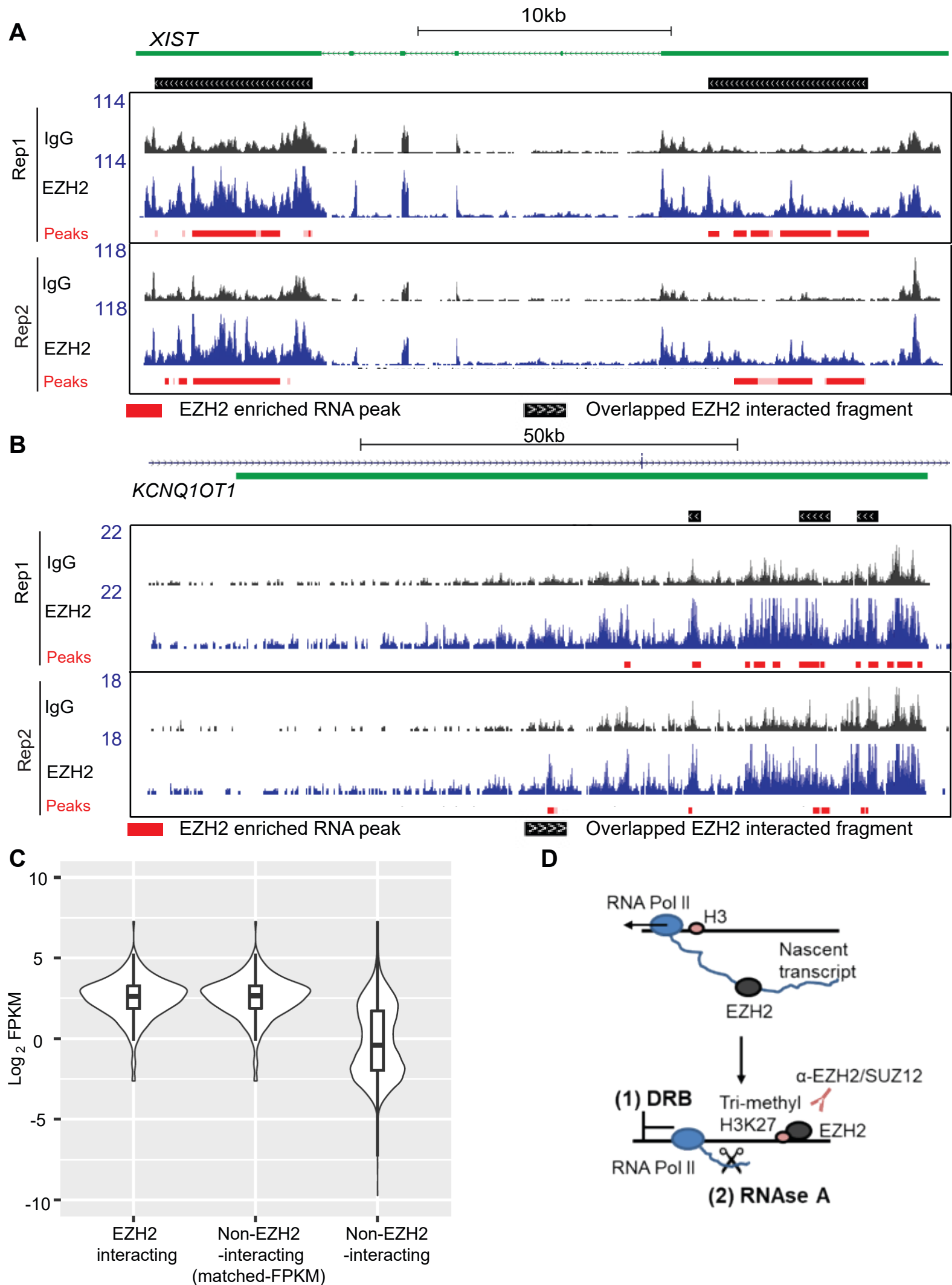


Figure 7. The proposed model of LEVER RNA regulating ϵ -globin transcription.

Top diagram: LEVER RNA sequesters and antagonises PRC2 *in cis*. This maintains the accessibility of the LEVER locus and facilitates its chromatin interaction with the ϵ -globin promoter and LCR region, which negatively regulates ϵ -globin transcription.

Bottom: When LEVER is silenced, the LCR can now interact with the ϵ -globin promoter to induce ϵ -globin transcription.

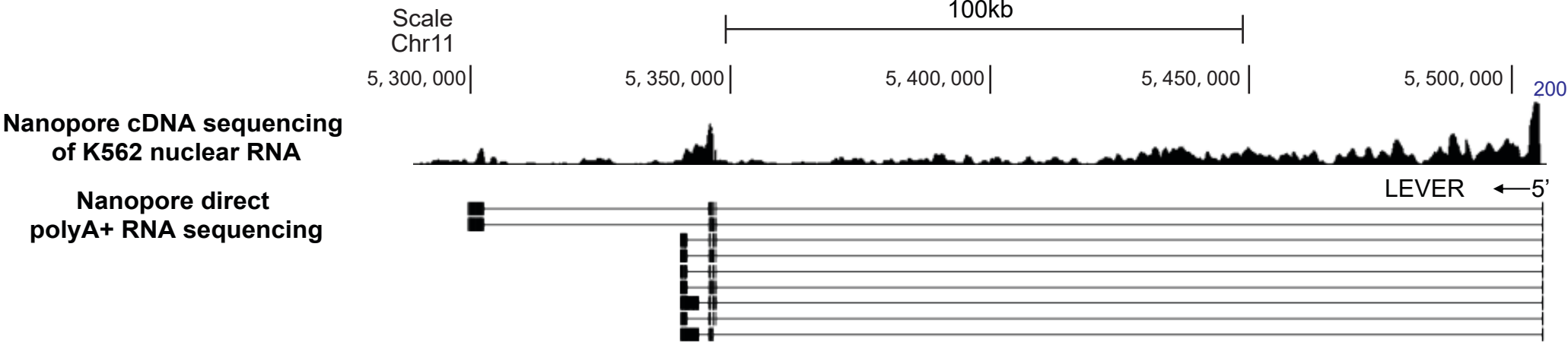
Supplementary Figure 1



Supplementary Figure 1

- (A) Genome tracks of enriched EZH2 RIP peaks of XIST RNA with reference to Genecode v29 in K562 cells.
- (B) Genome tracks of enriched EZH2 RIP peaks of KCNQ1OT1 RNA with reference to Genecode v29 in K562 cells.
- (C) Log₂ converted FPKM values of EZH2-interacting RNAs, a FPKM-matched subset of non-EZH2-interacting RNAs, and the total non-EZH2 interacting RNAs.
- (D) Design of experiments to demonstrate dependence on transcription and/or RNA (Figure 1E and 1F). K562 cells were treated with (1) 100 μ M DRB or vehicle (DMSO); or (2) 1mg/ml RNase A or vehicle. Immunoprecipitation of EZH2 or SUZ12 was performed after treatment to analyse the binding of PRC2 to histone H3.

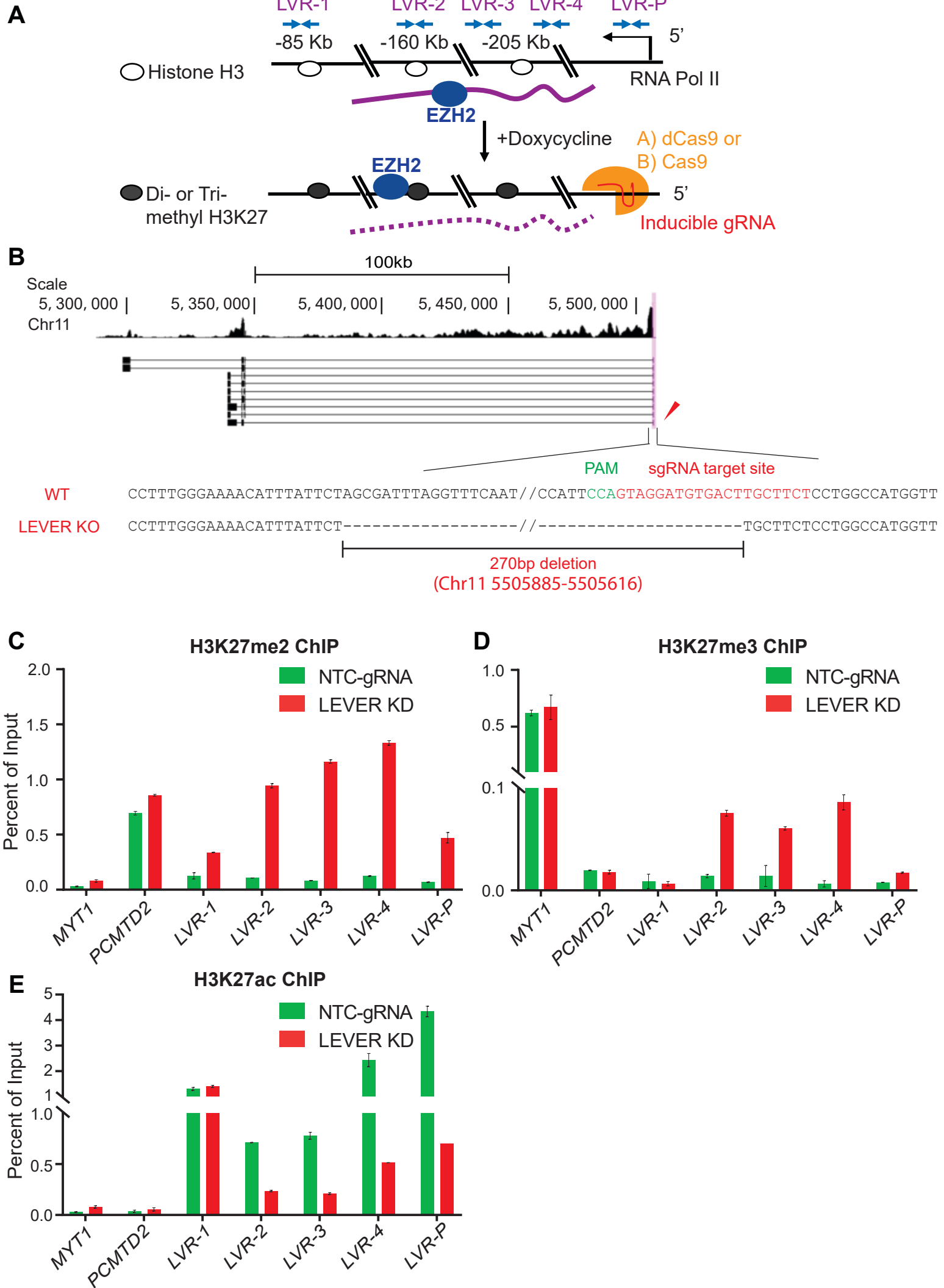
Supplementary Figure 2



Supplementary Figure 2

Mapping of LEVER RNA to the human genome with Nanopore cDNA sequencing of K562 nuclear RNA or direct-RNA sequencing of K562 total polyA+ RNA.

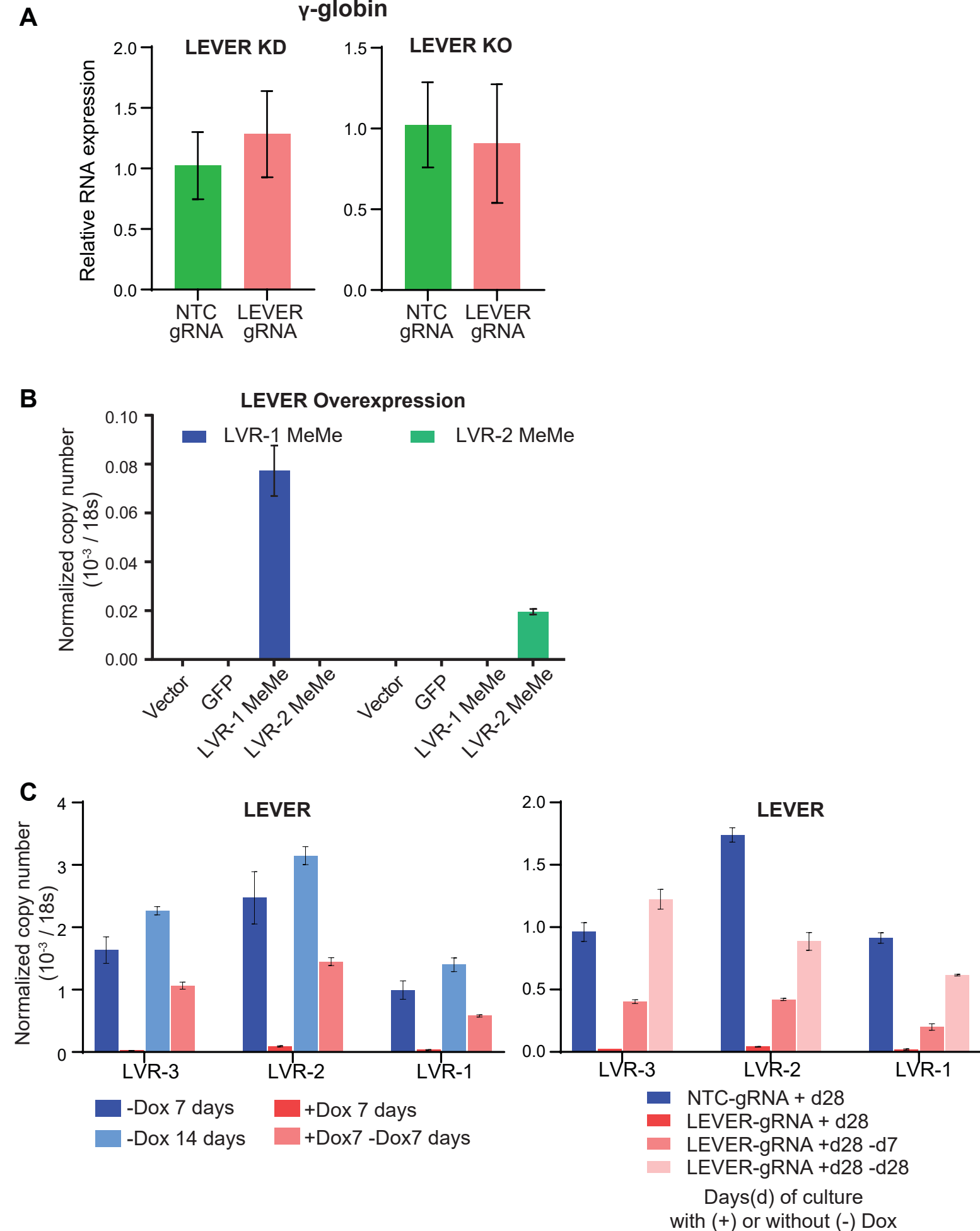
Supplementary Figure 3



Supplementary Figure 3.

- (A) Illustration of LEVER knockout (KO) or doxycycline-inducible knock down (KD) using the CRISPR-Cas9 or CRISPRi-dCas9 system. The guide RNA was designed to target at the LEVER TSS region.
- (B) CRISPR-Cas9 generated LEVER KO cells with homozygous deletion spanning the LEVER promoter.
- (C,D,E) CHIP-qPCR analysis of (C) H3K27me2, (C) H3K27me3, and (D) H3K27ac along the LEVER locus in doxycycline treated LEVER KD cells or NTC cells after 28 days of doxycycline. MYT1 serves as a CHIP positive control and PCMTD2 serves as a CHIP negative control.

Supplementary Figure 4



Supplementary Figure 4.

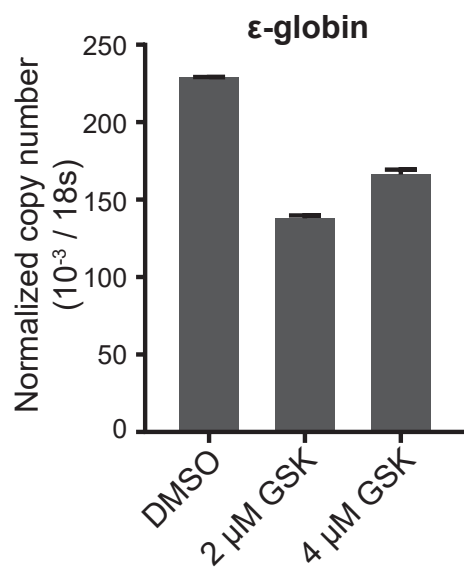
(A) RT-qPCR analysis of γ -globin in LEVER knock-down and knock-out cells.

Notably, the β - and δ -globin genes are not stably expressed in K562 wild-type or LEVER depleted cells.

(B) RT-qPCR analysis of LEVER-MeMe fragments in LEVER-KO+Vector control, KO+GFP control, KO+LVR-1 MeMe, and KO+LVR-2 MeMe.

(C) RT-qPCR of LEVER expression at LVR-1, LVR-2, and LVR-3 regions in corresponding cells and culturing conditions for Figure 5A and 5B, respectively.

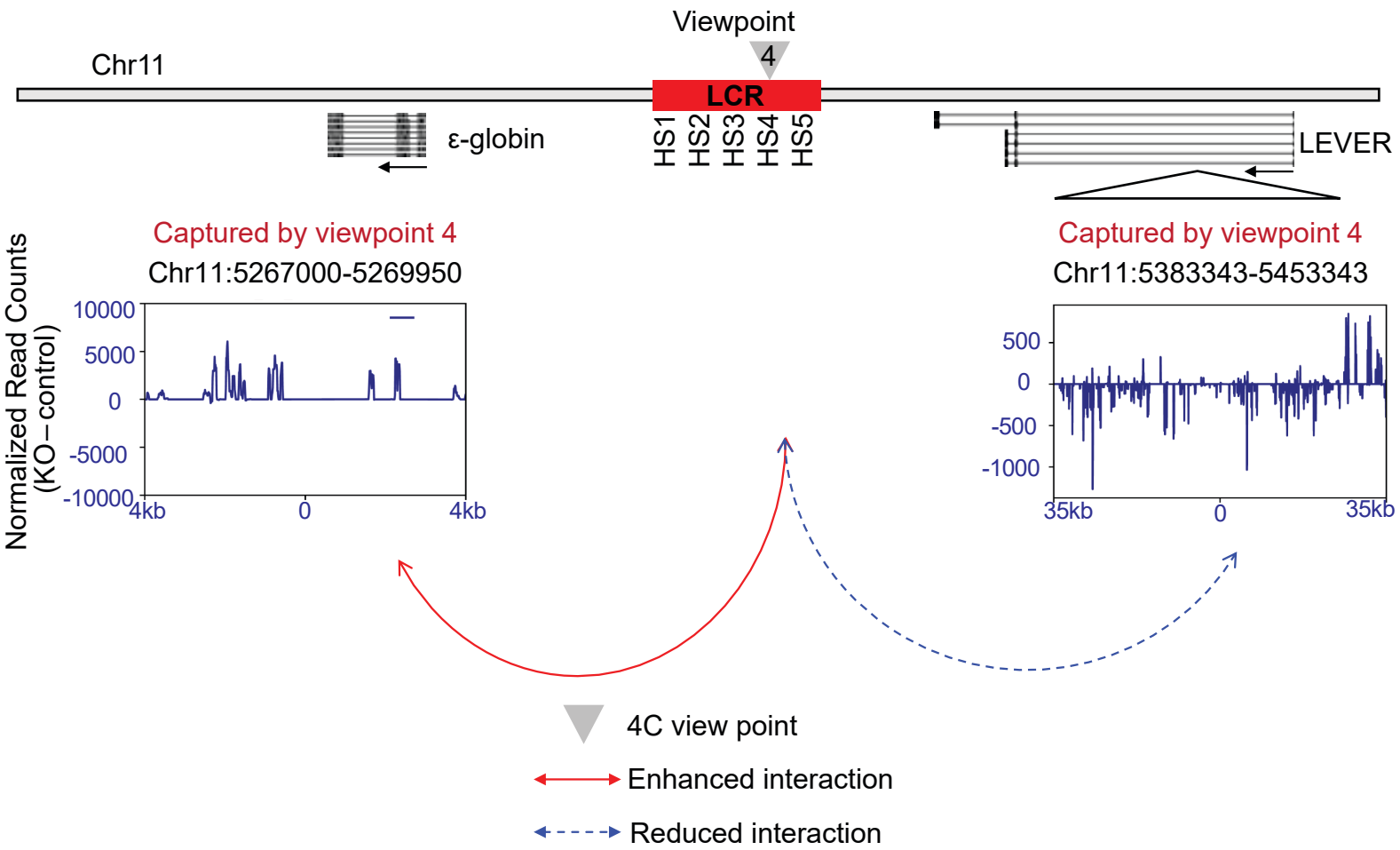
Supplementary Figure 5



Supplementary Figure 5

Quantitative RT-PCR of ϵ -globin in LEVER knock-out cells treated with DMSO, 2 μ M GSK343, or 4 μ M GSK343.

Supplementary Figure 6



Supplementary Figure 6

4C analysis of NTC and LEVER KO K562 cells at viewpoint 4 as described in

Figure 6C.

1 **Materials and Methods**

2 **Cell culture**

3 K562 and HEK293T cell lines were obtained from the American Type Culture
4 Collection (ATCC) and were used in experiments at their early passages. K562 cells
5 were cultured in RPMI (Biowest) supplemented with 10% fetal bovine serum (FBS)
6 (Biowest). HEK293T cells were cultured in DMEM supplemented with 4500 mg/L
7 glucose (Biowest), 2mM L-glutamine (Thermo Scientific), 1X MEM non-essential
8 amino acids (Thermo Scientific), 1mM sodium pyruvate, and 10% fetal bovine serum
9 (FBS) (Biowest). All cell lines were tested for mycoplasma every six months.

10

11 **Lentivirus production**

12 pCMV-dR8.91, pCMV-VSV-G, and lentivectors were transfected into 10 million
13 293T cells using Lipofectamine 2000 following the manufacturer's instructions.
14 Supernatants were harvested at 48 and 72 h after transfection. Viruses were
15 concentrated 100–300 times through centrifugation after filtering through 0.45 µM
16 syringe filters and stored at –80°C. The titers of viruses were determined on HeLa
17 cells using flow cytometry. The plasmids used in this study are listed in
18 Supplementary Table S2.

19

20 **CRISPR/Cas9 and CRISPRi/dCas9**

21 K562 cells were transduced with the FUCas9Cherry or FUdCas9Cherry
22 lentiviruses at 0.5 MOI with 5µg/mL polybrene. The transduced cells were sorted for
23 mCherry positive as K562-Cas9/dCas9 stable lines. FgH1tUTG-sgRNA viruses were
24 further transduced into these K562-Cas9/dCas9 stable lines and sorted for GFP
25 positives. To generate EZH2 KO, in vitro synthesized sgRNA (Synthego) was

1 transiently transfected into K562-Cas9 cells with DharmaFECT 1 (DHARMACON)
2 following the manufacturer's instructions. Of note, the DNA editing efficiency of the
3 sgRNAs at their specific targeting regions was determined in K562-Cas9-sgRNA
4 cells by the T7 endonuclease I (NEB) assay (as described in the manufacturer's
5 protocol) after sgRNA induction by 1 µg/mL doxycycline for 3 days. EZH2-sgRNA
6 transfected (48 hours) K562-Cas9 cells or Induced K562-Cas9-sgRNA cells were
7 seeded into single-cell clones in 96-well plates. Successful KO clones screened by
8 RT-qPCR (for LEVER KO) or western blot (for EZH2 KO) were expanded and
9 analyzed by genotyping. The list of sgRNAs used can be found in Supplementary
10 Table S2.

11

12 **Ectopic EZH2 or LEVER expression**

13 For the ectopic expression of EZH2 and its mutants, N-terminal Flag-tagged
14 EZH2 (WT) was cloned from K562 cDNA into a modified pLX304 plasmid in which
15 the CMV promoter was replaced with the CAG promoter. The EZH2^{ΔSET} and
16 EZH2^{Y641F} mutants were derived from EZH2 WT and cloned into the same vector.
17 For the ectopic expression of LEVER fragments, LVR-1 or LVR-2 were amplified
18 from K562 genomic DNA and cloned into a modified pLX304CAG plasmid, in which
19 the CAG promoter and the hPGK promoter were inverted, and a BGH stop sequence
20 was added to ensure that only the designated LEVER RNA fragments are being
21 expressed. Lentivirus was produced as stated above and transduced into
22 corresponding knock-out K562 cells and selected with blasticidin.

23

24 **Western blot**

1 Protein lysates were prepared in RIPA buffer (50mM Tris-Cl pH 7.5, 150 mM
2 NaCl, 1 mM EDTA, 1% NP-40, 0.5% sodium deoxycholate (NaDOC), 0.1% SDS),
3 and the lysates were mildly sonicated into clear supernatants before subjected to
4 protein quantification by the Pierce BCA assay (Thermo Scientific). 10-20 μ g of
5 protein lysate was loaded into each well of SDS-PAGE mini gel. The electrophoresis
6 was performed in Tris-glycine buffer supplemented with 0.1% SDS. Proteins were
7 then transferred onto nitrocellulose membranes (Bio-rad) in Tris-glycine-methanol
8 transfer buffer. The membrane was blocked by 5% skim milk/TBST and incubated
9 with primary antibodies followed by HRP-conjugated secondary antibodies diluted in
10 5% skim milk/TBST. Chemiluminescence was detected by exposing X-ray film to the
11 blot pre-incubated with the Luminata Crescendo Western HRP substrate (Millipore).
12 The list of antibodies used can be found in Supplementary Table S1.

13

14 **Quantitative PCR (qPCR)**

15 1 μ g of RNA was reverse transcribed using Qscript cDNA Supermix (QuantaBio).
16 qPCR on cDNA or ChIP DNA was performed with GoTaq qPCR Master Mix
17 (Promega) under the manufacturer's instruction on a QS5 system (Thermo
18 Scientific). The copy numbers were measured by predetermined standard curves.
19 The relative expression or enrichment to control samples were calculated based on
20 the $\Delta\Delta$ Ct method. Quantitative digital droplet PCR was performed on a QX200
21 Droplet Digital PCR (ddPCR) System (Bio-Rad) with EvaGreen Digital PCR
22 Supermix (Bio-Rad) under the manufacturer's instructions. Primers used in this study
23 are listed in Supplementary Table S2.

24

25 **RNA fractionation and sequencing**

1 Nuclear RNA purification was performed as described previously ¹. Briefly, 12
2 million cells were harvested and washed first with 20mL of cold PBS supplemented
3 with 5mM ribonucleoside Vanadyl complex (NEB) and 1mM phenylmethylsulphonyl
4 fluoride (PMSF), and then with 20mL of cold PBS supplemented with 1mM PMSF.
5 The cell pellet was lysed in 1.7 mL of cold hypotonic buffer (10mM HEPES pH7.6,
6 25mM KCl, 0.15mM spermine, 0.5mM spermidine, 1mM EDTA, 2mM sodium
7 butyrate, 1.25M sucrose, 10% glycerol, 5mg/mL BSA, 0.5% NP-40) supplemented
8 with 1mM PMSF, EDTA-free protease inhibitor cocktail (Roche) and 20U/mL
9 SUPERase-In (Invitrogen). The cells were homogenized with a 15 mL Dounce
10 homogenizer (by 10 strokes with Pestle A and 45 strokes with Pestle B). The
11 homogenized cells were diluted in 7.8 mL of NP-40 free hypotonic buffer containing
12 2M sucrose. 3.2 mL of cushion buffer (BSA and NP-40 free hypotonic buffer
13 containing 2M sucrose) was added to the bottom of a polypropylene tube (Beckman
14 Coulter, 331372), and the cell lysate was overlaid on the cushion buffer without
15 disrupting the bottom phase. Nuclei were harvested after ultra-centrifugation at
16 100,000 x g for one hour at 4°C with a SW41Ti rotor and Beckman Coulter TH641
17 ultra-centrifuge. The supernatant was removed, and the nuclear RNA was extracted
18 by Trizol according to the manufacturer's protocol. Nuclear polyA⁺ RNA was
19 extracted with a NEBNext Poly(A) mRNA Magnetic Isolation Module according to the
20 manufacturer's protocol, and the nuclear polyA⁻ fraction was extracted from the
21 unbound supernatant by ethanol precipitation. Cytoplasmic RNA was isolated with a
22 Qiagen RNeasy kit according to the manufacturer's protocol. The fractionated RNA
23 was subjected to DNase I digestion and then re-purified using Trizol. The RNA was
24 quantified with the Qubit RNA HS assay kit (Thermo Scientific), and rRNA was
25 removed by the Ribo-Zero Gold rRNA Removal Kit (Illumina). The stranded RNA-seq

1 libraries were constructed with the Script-seq V2 RNA-seq library preparation kit
2 (Illumina). The 250–500 bp size-selected libraries were sequenced on the Illumina
3 MiniSeq platform with MiniSeq High Output 2X150 paired-end kit.

4

5 **DRB / RNase A treatment and immunoprecipitation**

6 For DRB treatment, 12 million K562 cells were treated with 100 μ M DRB for 24
7 hours before subjected to nuclei extraction. For RNase A treatment, 10 million K562
8 cells were first harvested and washed with 1X PBS, and then treated with 1 mg/mL
9 RNase A in 0.05% Tween-20 for 10 mins at RT before subjected to nuclei extraction.
10 The DRB/RNase A treated cells were lysed on ice for 5 min with cell lysis buffer (20
11 mM Tris pH 7.5, 10 mM KCl, 1.5 mM MgCl₂, 0.1% Triton X-100) supplemented with
12 EDTA-free protease inhibitor cocktail (Roche), phosphatase inhibitor (Roche), 1 mM
13 DTT, and 1 mM PMSF and then centrifuged at 800 RPM for 10min at 4°C. The nuclei
14 pellets were washed once with the cell lysis buffer without Triton X-100 and
15 harvested by centrifugation at 1300 x g for 4 min at 4°C. Nuclei were resuspended in
16 resuspension buffer (50 mM Tris pH 7.5, 150 mM NaCl, 2 mM EDTA) supplemented
17 with EDTA-free protease inhibitor cocktail (Roche), phosphatase inhibitor (Roche),
18 80 U/mL RNaseOUT (ThermoScientific), and 1 mM PMSF. The nuclei were sonicated
19 using EpiShear Probe Sonicator (Active Motif) for 15 seconds at 30% amplitude. 500
20 μ l of the lysate was incubated with EZH2 antibody (1:250), SUZ12 antibody (2 μ g), or
21 rabbit IgG isotype (2 μ g) in the cold room for 12 hours. The lysate was incubated with
22 50 μ l pre-blocked dynabeads protein A (Invitrogen) for 2 hours at 4 °C. The beads
23 were washed with washing buffer (150mM NaCl, 10mM Tris-HCl, pH7.4, 1mM
24 EDTA, 1mM EGTA, pH8.0, 1% Triton X-100, 0.5% NP-40, 1U/mL SUPERase-In) for

1 three times before eluted with 30 μ l of RIPA buffer and 15 μ l of 6X Laemmli buffer at
2 95°C for 10 mins. The eluted complex was analyzed with western blot.

3

4 **RNA immunoprecipitation (RIP)**

5 RIP was performed as described previously ¹ with the following modifications. (A)
6 SUPERase-In was used in substitution of ribonucleoside Vanadyl complex from the
7 cell lysis step. (B) Fixed K562 cells were lysed with Dounce homogenizer by 10
8 strokes using Pestle A and 60 strokes using Pestle B. (C)The nuclei were sonicated
9 with Bioruptor (Diagenode) for 8 cycles (30s on, 30s off, high power). (D) Nuclear
10 lysates were pre-cleared with Dynabeads protein G twice at 4°C. (E) 5 μ g of EZH2 or
11 IgG isotype control antibodies were used in each RIP experiment. RIP experiments
12 were performed in three biological replicates, and two of these replicates were
13 independently sequenced. For sequencing, the immunoprecipitated RNA was
14 quantified with the Qubit RNA HS assay kit (Thermo Scientific) before rRNA
15 Removal with Ribo-Zero Gold rRNA Removal Kit (Illumina). 15ng of the rRNA-
16 depleted RNA was used for sequencing library construction. The stranded RIP-seq
17 libraries were constructed with a Script-seq V2 RNA-seq library preparation kit
18 (Illumina). The libraries were subjected to size selection (250–500 bp) on a 4–20%
19 TBE PAGE gel (Thermo Scientific). The recovered libraries were sequenced on the
20 Illumina MiniSeq platform using the MiniSeq High Output kit with 2 \times 150 paired-end.

21

22 **Chromatin immunoprecipitation (ChIP)**

23 ChIP was performed as described previously ² with minimal modifications. Briefly,
24 K562 chromatin was sonicated with Bioruptor (Diagenode) for 15-25 cycles (30s on,
25 30s off, high power), and immunoprecipitation was performed with Dynabeads

1 protein G. For ChIP-sequencing, the libraries were constructed using a Thru-PLEX
2 DNA-seq 12S kit (TakaraBio). The libraries were subjected to size selection (250–
3 500 bp) on a 4–20% TBE PAGE gel (Thermo Scientific). The recovered libraries
4 were sequenced on the Illumina Nextseq platform using Nextseq 500 high-output kit
5 with 2 × 76 paired-end. For H3K27me3, H3K27me2, or H3K27Ac ChIP-seq, one
6 replicate was sequenced, and the results were validated by ChIP-qPCR with
7 independent biological replicates.

8

9 **Oxford Nanopore sequencing**

10 Fractionated nuclear RNA was used to construct a cDNA library using a 1D
11 sequencing kit (SQK-LSK108) according to the manufacturer's instructions. In brief,
12 first-strand cDNA was generated from 20 ng of rRNA depleted nuclear RNA using
13 SuperScript IV reverse transcriptase with 0.35 pmol/μl random hexamer at 52°C for
14 60 mins. Second strand cDNA was synthesized with the NEBNext second strand
15 synthesis kit for 1 hour at 16°C. cDNA was purified with AMPure XP bead. Library
16 was constructed with LongAmp Taq PCR for 18 cycles. The PCR product was
17 purified with phenol: chloroform: isoamyl alcohol (25:24:1) and ethanol precipitated in
18 the presence of glycogen and was finally dissolved in 10 mM Tris buffer (pH8). The
19 cDNA library was sequenced on a Nanopore flow cell (FLO-MIN106). Enriched
20 polyA+ RNA was used to construct the library for direct RNA sequencing (SQK-
21 RNA001) according to the manufacturer's instructions. In brief, 400 ng of total polyA+
22 RNA was used for reverse transcription adaptor (RTA) ligation before subjected to
23 reverse transcription. RNA/DNA hybrids were purified with RNAClean XP beads and
24 ligated with RNA adaptors as direct RNA sequencing library. The constructed library
25 was sequenced on the Nanopore flow cell (FLO-MIN106).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

Circularized Chromosome Conformation Capture (4C)

4C-seq was performed as described previously³ with modifications. K562 cells were fixed as ChIP protocol except that the cells were fixed in 2% formaldehyde at the concentration of 1 million cells per mL for 10min. Cells were lysed in complete lysis buffer (10 mM Tris-HCl, pH 8, 10 mM NaCl, 0.2% NP-40, and complete protease inhibitor cocktail with EDTA (Roche)). The nuclei were pelleted, washed once with complete lysis buffer, and then lysed with 100ul 0.5% SDS solution at 62°C with 400RPM agitation for <10 mins. 299 µl nuclease-free water and 20 µl of 20% Triton X-100 were added before 15 min incubation at 37 °C to quench SDS. 1X DpnII buffer (NEB), 0.8U/µl DpnII restriction enzyme (NEB), and 0.08 mg/µl BSA were supplemented to the chromatin solution for DpnII digestion at 37°C for 4 hours with 1000 RPM agitation. Additional 400U of DpnII was added to the sample for overnight digestion. DpnII was heat-inactivated at 62°C for <20 mins. An aliquot of the digested chromatin was sampled for DNA purification followed by DNA quantification, agarose gel electrophoresis, and PCR to monitor digestion efficiency. Digested chromatin equivalent to total 10 µg of DNA was ligated with 660U HC DNA ligase (Invitrogen, 30U/µl) in 1X ligation buffer (Invitrogen, supplemented with 1% Triton X-100, 0.1 mg/mL BSA) for 8-10 hours at 16°C without shaking. The reaction was incubated at room temperature for 30 mins. The chromatin was reverse-crosslinked with 0.5% SDS and 50 µg/mL of Proteinase K (Invitrogen) at 64°C overnight. RNA in the chromatin was digested with 2mg/mL RNase A at 37°C for one hour. The ligated DNA was extracted with phenol: chloroform: isoamyl alcohol (25:24:1) followed by chloroform and then precipitated with ethanol (68% to avoid SDS precipitation) in the presence of glycogen. The ligated DNA product was

1 analyzed by agarose gel electrophoresis, and the DNA concentration was
2 determined by Qubit DNA HS assay kit.

3 3 µg of the DNA product was further digested with 3U of specific second cutter
4 (NlaIII, MseI, or MluCI). Restriction enzyme was heat-inactivated by incubating the
5 chromatin at 65°C for 20 mins (MluCI was heat-inactivated in the presence of 1.17%
6 SDS). The double-digested DNA was analyzed by agarose gel electrophoresis prior
7 to phenol: chloroform: isoamyl alcohol (25:24:1) extraction and ethanol precipitation
8 in the presence of glycogen. 1 µg of purified DNA was ligated with HC DNA ligase
9 (Invitrogen, 30U/µl) at 20U/ug DNA in 1X ligation buffer (Invitrogen). For NlaIII or
10 MluCI digested samples, the ligation was performed at 16°C overnight. For MseI
11 digested samples, the ligation was performed at 4°C for 24 hours. The ligated DNA
12 was recovered by phenol: chloroform: isoamyl alcohol (25:24:1) extraction and
13 ethanol precipitation. 800 ng DNA from each of the biological triplicates were pooled
14 to prepare 4C library (2.4 µg of DNA in total). The library for each viewpoint was
15 constructed by PCR with 0.05 µM 4C adaptor primers (see primers section) and 2
16 ng/µl DNA templates using KAPA HiFi HotStart ReadyMix (KK2602). The libraries
17 were subjected to size selection (250–500 bp) on a 4–20% TBE PAGE gel (Thermo
18 Scientific). The recovered libraries were sequenced on the Illumina Nextseq platform
19 using Nextseq 500 mid-output kit with 2x150 paired-end.

20

21 **Bioinformatics**

22 The NGS data generated in this study have been deposited to the NCBI Gene
23 Expression Omnibus (GEO) under accession number GSE141083.

24 For ChIP-seq analysis, reads were trimmed of adaptor sequences by TrimGalore
25 ⁴ and were mapped by bowtie2 ⁵ against human reference genome GRCh38. The

1 PCR duplicates were removed by SAMtools rmdup ⁶. The mapping results were
2 converted into signals by BEDTools ⁷ in bedGraph format. Then they were converted
3 to bigWig format by bedGraphToBigWig ^{8,9} for visualization on UCSC genome
4 browser ¹⁰. Peak calling was performed by MACS2 ¹¹.

5 RNA-seq reads were mapped by STAR ¹² against human reference genome
6 GRCh38 with GENCODE transcriptome annotation (v26) ¹³. RNA-seq reads were
7 trimmed of adaptor sequences by TrimGalore ⁴ and then mapped by STAR to the
8 human reference genome GRCh38 with reference gene annotation GENCODE ²⁶ ^{12,}
9 ¹³. PCR duplicates were removed in the paired-end alignments by SAMtools rmdup.
10 Alignments with mapping quality <20 were removed. Gene expression levels in
11 FPKM were determined by cuffdiff ¹⁴. RNA-seq read counts are generated by
12 featureCounts ¹⁵.

13 For RIP-seq, reads from the RIP data set and its control data set were trimmed of
14 adaptor sequences by TrimGalore ⁴ and were mapped respectively by STAR ¹²
15 against the human reference genome GRCh38 with GENCODE ¹³ transcriptome
16 annotation (v26). The resulting alignments were separated into two parts. (1) The
17 exonic part consisted of alignments belonging to GENCODE annotated transcripts.
18 (2) The non-exonic part consisted of the other alignments. Based on the number of
19 reads that can be mapped into the transcriptome, the larger data set was shrunk
20 down to fit the size of the smaller one. Then the normalized read count of each
21 genomic position was generated for the EZH2 RIP data set and IgG control data set,
22 respectively. The read coverage of each position was defined as the average of
23 normalized read counts within ± 150 bp region. Based on the comparison of the read
24 coverage between the EZH2 RIP and the IgG control, sites with ≥ 2 -fold enrichment
25 and Poisson distribution p -value $\leq 10^{-5}$ were defined as peaks. Each peak was

1 extended to surrounding areas until the fold enrichment dropped below 2. Peaks that
2 can be found in both replicates within 1000bp range were merged and considered
3 reproducible peaks. Sequences covered by reproducible peaks were extracted for
4 further motif discovery analysis. The upstream and downstream sequences of
5 reproducible peaks were also extracted and served as the background sequences in
6 the motif finding analysis. Motif discovery is performed with the MEME suite using
7 differential enrichment mode ^{16, 17}.

8 For Nanopore sequencing, raw read collected from MinKNOW were converted to
9 FASTQ file by Albacore 2.3.3 with the following arguments: -k SQK-RNA001 -f FLO-
10 MIN106 -o fastq for direct RNA sequencing; or the following arguments: Albacore
11 2.3.1 with --flowcell FLO-MIN106 --kit SQK-PCS108 --output_format fastq for cDNA
12 sequencing. For direct RNA sequencing, the reads were mapped to hg38 by
13 Minimap2 ¹⁸ with -ax splice -uf -k14. The bam files were converted to bed12 by
14 bamToBed -bed 12 and were converted to bigBed file by bedToBigBed -type=bed12.
15 Only uniquely mapped reads were retained. For cDNA sequencing, the reads were
16 mapped to hg38 by LAST with lastal -d90 -m50 -D10 and last-split -d2 ^{19, 20}. The maf
17 files were converted to psl file with maf-convert -j1e6 psl. The psl files were
18 subsequently converted to bam file with psl2sam.pl and cigar_tweaker scripts within
19 the SAMtools package ⁶. For visualization, the read alignments were converted into
20 bedGraph by BEDTools ¹², and finally, into bigWig format by bedGraphToBigWig ^{8, 9}.

21 For 4C-seq analysis, the long-range genomic interaction regions generated by
22 the 4C-Seq experiment were analyzed by the R package r3CSeq v1.30.0 ²¹. Briefly,
23 for each replicate, the raw reads were aligned to the masked version of the reference
24 human genome (masked for the gap, repetitive and ambiguous sequences)
25 downloaded from the R Bioconductor repository

1 (BSgenome.Hsapiens.UCSC.hg19.masked). The chromosome 11 was selected as
2 the viewpoints, and DpnII was used as the restriction enzyme to digest the genome.
3 A non-overlapping fragment was selected to identify the interaction regions. BAM
4 files were converted to read coverage by bedtools genomecov. The read coverage
5 was normalized according to the sequencing depth. The read coverage difference of
6 the treatment and the control was calculated based on the normalized read
7 coverage. The resulting bedGraph files were converted to bigWig format by
8 bedGraphToBigWig. The genome version of bigWig files was then converted to hg38
9 by CrossMap²². The profile of read coverage difference on genomic intervals was
10 plotted by deepTools²³.

- 1 1. Di Ruscio, A. *et al.* DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature*
2 **503**, 371-376 (2013).
- 3 2. Tan, H.K. *et al.* DNMT3B shapes the mCA landscape and regulates mCG for promoter
4 bivalency in human embryonic stem cells. *Nucleic Acids Res* **47**, 7460-7475 (2019).
- 5 3. Matelot, M. & Noordermeer, D. Determination of High-Resolution 3D Chromatin
6 Organization Using Circular Chromosome Conformation Capture (4C-seq). *Methods Mol Biol*
7 **1480**, 223-241 (2016).
- 8 4. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
9 *2011* **17**, 3 %J EMBnet.journal (2011).
- 10 5. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**,
11 357-359 (2012).
- 12 6. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-
13 2079 (2009).
- 14 7. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic
15 features. *Bioinformatics* **26**, 841-842 (2010).
- 16 8. Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S. & Karolchik, D. BigWig and BigBed: enabling
17 browsing of large distributed datasets. *Bioinformatics* **26**, 2204-2207 (2010).
- 18 9. Raney, B.J. *et al.* Track data hubs enable visualization of user-defined genome-wide
19 annotations on the UCSC Genome Browser. *Bioinformatics* **30**, 1003-1005 (2014).
- 20 10. Kent, W.J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).
- 21 11. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
- 22 12. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
- 23 13. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE
24 Project. *Genome Res* **22**, 1760-1774 (2012).
- 25 14. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated
26 transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-515
27 (2010).
- 28 15. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for
29 assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
- 30 16. Bailey, T.L. & Elkan, C. Fitting a mixture model by expectation maximization to discover
31 motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**, 28-36 (1994).
- 32 17. Bailey, T.L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**,
33 W202-208 (2009).
- 34 18. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100
35 (2018).
- 36 19. Frith, M.C. & Kawaguchi, R. Split-alignment of genomes finds orthologies more accurately.
37 *Genome Biol* **16**, 106 (2015).
- 38 20. Hamada, M., Ono, Y., Asai, K. & Frith, M.C. Training alignment parameters for arbitrary
39 sequencers with LAST-TRAIN. *Bioinformatics* **33**, 926-928 (2017).
- 40 21. Thongjuea, S., Stadhouders, R., Grosveld, F.G., Soler, E. & Lenhard, B. r3Cseq: an
41 R/Bioconductor package for the discovery of long-range genomic interactions from
42 chromosome conformation capture and next-generation sequencing data. *Nucleic Acids Res*
43 **41**, e132 (2013).
- 44 22. Zhao, H. *et al.* CrossMap: a versatile tool for coordinate conversion between genome
45 assemblies. *Bioinformatics* **30**, 1006-1007 (2014).
- 46 23. Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data
47 analysis. *Nucleic Acids Res* **44**, W160-165 (2016).

48