Hijacking of transcriptional condensates by endogenous retroviruses

Vahid Asimi
Max Planck Institute for Molecular Genetics
Sara Hetzel
Max Planck
Abhishek Sampath Kumar
MPIMG
Henri Niskanen
MPIMG
Christina Riemenschneider
MPIMG
Julian Naderi
MPIMG
Nina Fasching
Institute of Molecular Biotechnology (IMBA)
Niko Popitsch
Institute of Molecular Biotechnology (IMBA)
Helene Kretzmer
Max Planck Institute for Molecular Genetics
Zachary Smith
Harvard University
Maria Walther
Max Planck Institute for Molecular Genetics
David Meierhofer
Max Planck Institute for Molecular Genetics
Lars Wittler
Max Planck Institute for Molecular Genetics
René Buschow
MPIMG https://orcid.org/0000-0002-9800-2578
Bernd Timmermann
Sequencing Core Facility, Max Planck Institute for Molecular Genetics
Stefan Ameres
Institute of Molecular Biotechnology (IMBA)
Alexander Meissner

Max Planck Institute for Molecular Genetics https://orcid.org/0000-0001-8646-7469
Denes Hnisz (hnisz@molgen.mpg.de)
Max Planck https://orcid.org/0000-0002-6256-1693

Article

Keywords: Endogenous retroviruses, retrotransposons, transcriptional condensates

Posted Date: September 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-802409/v1

Version of Record: A version of this preprint was published at Nature Genetics on July 21st, 2022. See the published version at https://doi.org/10.1038/s41588-022-01132-w.

Hijacking of transcriptional condensates by endogenous retroviruses

Vahid Asimi[1]†, Sara Hetzel[1]†, Abhishek Sampath Kumar[1]†, Henri Niskanen[1]†, Christina Riemenschneider[1,2]†, Julian Naderi[1], Nina Fasching[3], Niko Popitsch[3], Helene Kretzmer[1], Zachary D. Smith[4,5], Maria Walther[1], David Meierhofer[6], Lars Wittler[7], Rene Buschow[8], Bernd Timmermann[9], Stefan L. Ameres[3,10], Alexander Meissner[1,4,5], Denes Hnisz[1*]

[1] Department of Genome Regulation, Max Planck Institute for Molecular Genetics, Berlin, Germany
[2] Institute of Biotechnology, Technische Universität Berlin, Berlin, Germany
[3] Institute of Molecular Biotechnology (IMBA), Vienna BioCenter (VBC), Vienna, Austria.
[4] Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
[5] Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts, USA
[6] Max Planck Institute for Molecular Genetics, Mass Spectrometry Facility, Berlin, Germany
[7] Department of Developmental Genetics, Max Planck Institute for Molecular Genetics, Berlin, Germany
[8] Microscopy Core Facility, Max Planck Institute for Molecular Genetics, Berlin, Germany
[9] Sequencing Core Facility, Max Planck Institute for Molecular Genetics, Berlin, Germany
[10] Department of Biochemistry and Cell Biology, Max Perutz Labs, University of Vienna, Vienna BioCenter (VBC), Vienna, Austria
† these authors contributed equally

[*] Correspondence: hnisz@molgen.mpg.de

Endogenous retroviruses (ERVs) comprise ~10% of mammalian genomes, and ERVs are repressed by multiple cellular mechanisms including heterochromatin, DNA methylation, and methylation of their RNA transcripts [1-13]. Although the vast majority of ERVs in mammals are incapable of retrotransposition, ERV de-repression during early development is associated with embryonic lethality [14-16], suggesting that ERV transcription or RNA transcripts may underlie essentiality of ERV repression. Here we report that rapid and selective degradation of the TRIM28 heterochromatin adapter protein in murine embryonic stem cells triggers dissociation of transcriptional condensates from loci encoding super-enhancer -driven pluripotency genes, and association of transcriptional condensates with transcribed ERV loci. Knockdown of ERV RNA or forced expression of super-enhancer -enriched transcription factors rescued condensate localization at super-enhancers in TRIM28-degraded cells. In a biochemical reconstitution system, ERV RNA facilitated phase separation of RNA Polymerase II, and partitioning of NFY and Mediator into heterotypic droplets, suggesting a mechanistic basis for the association of transcriptional condensates with ERVs. Using a zygotic perturbation platform, we found that timing and amount of ERV transcription correlate with the onset of lethality, and that loss of TRIM28 leads to specific depletion of pluripotent lineages in mouse embryos. We propose that retrotransposons contribute to the genomic distribution of nuclear condensates, and that coding and non-coding nascent RNAs may facilitate "hijacking" of transcriptional condensates in various developmental and disease contexts.

Heterochromatin plays essential roles in repressing retrotransposons such as endogenous retroviruses (ERVs) during embryonic development [1-6]. One of the best studied repressive pathways involves the TRIM28 co-repressor recruited by KRAB-ZFP transcription factors to ERVs in pluripotent embryonic stem cells (ESCs), where it recruits the histone H3K9 methyltransferase SETDB1 and HP1α that establish a repressive chromatin environment [7,9,17-20] (Fig. 1a-b, Extended Data Fig. 1a-j). In contrast, heterochromatin components typically do not occupy transcription-factor bound enhancers that drive cell-specific transcriptional programs (Fig. 1a-b, Extended Data Fig. 1a-j) [7,21,22]. ERV de-repression is associated with lethality at various embryonic stages and in pluripotent stem cells deficient for the TRIM28-HP1α pathway [7-10], even though the vast majority of ERVs in mice and humans have lost their ability to

retrotranspose [1-6] , and deletion of entire clusters of KRAB-ZFP factors does not lead to elevated transposition rates in mice [16]. Previous studies suggested that de-repressed ERVs can act as enhancers and alternative promoters of cellular genes [1,23-25], but why ERV de-repression may be lethal in early embryos and pluripotent cells is a mystery.

Resolving the direct consequences of ERV de-repression has in part been impeded by limitations of classic gene disruption strategies and essentiality of the TRIM28-HP1α pathway in embryonic cells [7-10]. Therefore, we first generated a murine embryonic stem cell (ESC) line that encodes degradation-sensitive TRIM28-FKBP alleles using the dTAG technology (Fig. 1c, Extended Data Fig. 2a-b) [26]. Endogenously tagged TRIM28 experienced reversible, ligand-dependent proteolysis with near-complete degradation after 6h exposure to the dTAG-13 ligand (Fig. 1d, Extended Data Fig. 2b). Quantitative mass-spectrometry confirmed that TRIM28 degradation was highly selective up to 24h of dTAG-13 treatment (Extended Data Fig. 2c). Of note, short term (up to 24h) TRIM28 degradation did not substantially alter the protein levels of pluripotency markers (e.g. OCT4, SOX2, SSEA-1) (Fig. 1d, Extended Data Fig. 2d-e) suggesting that acute TRIM28 degradation did not considerably alter the pluripotent state.

To monitor changes in transcriptional activity upon acute TRIM28-degradation, we used TT-SLAM-Seq, a recently developed high-sensitivity genome-wide nascent transcription readout [27]. TT-SLAM-Seq combines metabolic labeling and chemical nucleoside conversion (SLAM-Seq) [28] with selective enrichment of newly synthesized RNA (TT-Seq) [29] to detect nascent RNA transcription with high temporal resolution and sensitivity (Extended Data Fig. 3a-c). Consistent with previous reports [7,24], we observed de-repression of major classes of ERVs, including intracisternal A-type particles (IAPs), MMERVK10s and MMERVK9s in TRIM28-degraded ESCs (Fig. 1e, Extended Data Fig. 4a-f), and loss of H3K9me3 at these sites (Extended Data Fig. 4g). De-repression of ERVs was also confirmed with extended TRIM28-degradation for 96h and RNA-Seq (Fig. 1e, Extended Data Fig. 4a-e). The TT-SLAM-Seq data revealed ~300 genes whose transcription was significantly induced, and ~300 genes whose transcription was significantly reduced upon 24h of TRIM28 degradation (>2-fold, FDR<0.05) (Fig. 1f-g). The downregulated genes were enriched for super-enhancer –associated pluripotency genes (NES=-1.6, $p<10^{-3}$) (Fig. 1f-g, Extended Data Fig. 5a-d). Downregulation of these genes was associated

with the reduction of nascent transcription at the super-enhancers, that tended to precede the reduction of transcription at the super-enhancer -driven gene (Fig. 1f, 1h, Extended Data Fig. 5a-c). These results were unexpected, as TRIM28 binds to ERVs in the genome of ESCs and is not bound at enhancers or super-enhancers (Fig. 1b, Extended Data Fig. 1i-j), and because previous studies using genetic disruption and long term cultures (96 hours) reported 8-times more upregulated than downregulated genes in TRIM28 KO ESCs [24]. These data reveal the direct transcriptional response to the loss of TRIM28, and suggest that acute TRIM28 degradation leads to reduction of super-enhancer transcription in ESCs.

Components of the transcription machinery, e.g. RNA Polymerase II (RNAPII) and the Mediator co-activator form biomolecular condensates that associate with super-enhancers in ESCs [30-33], and the presence of RNAPII condensates at genomic sites correlates with elevated transcriptional activity [32]. We thus hypothesized that reduction of super-enhancer transcription in TRIM28-degraded ESCs may be caused by dissociation of transcriptional condensates from super-enhancer loci. To test this idea, we visualized the genomic region containing the super-enhancer at the *MiR290-295* locus using nascent RNA-FISH, and transcriptional condensates using immunofluorescence against RNAPII. RNAPII puncta consistently co-localized with the *MiR290-295* locus in control ESCs, and the co-localization was diminished after 24h of TRIM28 degradation (Fig. 1i, Extended Data Fig. 6a), while the overall level of RNAPII did not change (Extended Data Fig. 6b). These data indicate that transcriptional condensates less frequently associate with the *MiR290-295* super-enhancer locus in TRIM28-degraded ESCs.

To investigate whether transcriptional condensates co-localize with de-repressed ERVs, we visualized IAP ERV loci with RNA-FISH. Nuclear *IAP* foci became progressively apparent after 24-48h of TRIM28-degradation (Fig. 2a, Extended Data Fig. 7a), and some nuclear *IAP* foci co-localized with RNAPII puncta visualized with IF [Mean Mander's co-localization coefficient (MCC): 0.1, n=98 cells] (Fig. 2a, Extended Data Fig. 7a). Co-localization of *IAP* foci was similarly observed with Mediator puncta visualized with IF using antibodies against the MED1 and MED23 Mediator subunits (Fig. 2b, Extended Data Fig. 7b), and with the NFY transcription factor (TF), whose motif is significantly enriched in the Long Terminal Repeat (LTR) of IAPs (Fig. 2c, Extended Data Fig. 8a). Consistent with the co-localization of transcriptional

condensates with *IAP* foci and their loss of co-localization with super-enhancers, the occupancy of RNAPII, Mediator and the transcription-associated H3K27Ac chromatin footprint increased at various ERV families already after 24h TRIM28-degradation, while their occupancy was reduced at super-enhancers (Fig. 2d, Extended Data Fig. 9a-b).

As newly formed transcriptional condensates could influence their genomic environment, we explored the possibility that transcriptional condensates associating with ERVs may incorporate ERV-proximal genes. To this end, we visualized the *Cthrc1* locus using nascent RNA-FISH, as *Cthrc1* was among the most upregulated genes in the TT-SLAM-Seq data after 24h of TRIM28-degradation, and is located within 100kb of three ERVs (Fig. 2e). We found that the *Cthrc1* locus co-localized with RNAPII and NFY puncta in TRIM28-degraded ESCs (Fig. 2f-g, Extended Data Fig. 8b-c), but not with puncta of a control TF NRF1 (Extended Data Fig. 8d). Transient (30min) treatment of the cells with 1.5% 1-6 hexanediol (1-6 HD), a short chain aliphatic alcohol that dissolves various biomolecular condensates including RNAPII condensates [32] (Extended Data Fig. 7c-d), reduced the level of *Cthrc1* nascent RNA (2-fold, $p<0.05$, t-test) in TRIM28-degraded cells, indicating that RNAPII condensates contribute to the upregulation of this gene (Extended Data Fig. 7e). We then used CRISPR/Cas9 to delete the three ERVs at the *Cthrc1* locus, and found that in the absence of the three ERVs, *Cthrc1* was not induced upon TRIM28-degradation (Extended Data Fig. 7f-g). To further probe contacts between de-repressed ERVs and genes, we performed in situ Hi-C in control and TRIM28-degraded ESCs. 24h of TRIM28 degradation did not lead to major genome-wide changes in chromatin contacts (Extended Data Fig. 10a-b), but did lead to a shift of the most-induced ERV taxa from the inactive "B" to the active "A" compartment (Fig. 2h, Extended Data Fig. 10c), and an increase of the contact frequency between ERVs and transcribed genes (Fig. 2i, Extended Data Fig. 10b, 10d). These results demonstrate that transcriptional condensates may incorporate genes and de-repressed ERVs.

RNA is a key component of numerous biomolecular condensates [34] , and nascent RNA can enhance phase separation of transcriptional regulatory proteins [35]. Therefore, we hypothesized that RNA produced at ERV loci may contribute to the genomic localization of RNAPII-containing condensates. To test this model, we knocked-down various ERV RNAs in TRIM28-

degraded cells. Expression of shRNAs targeting the four most-induced ERV families (IAPs, MMERVK10s, MMERVK9s, MMETns) partially rescued the downregulation of super-enhancers and their associated genes after 24h of TRIM28 degradation while knocking down IAPs alone did not (Extended Data Fig. 11a-d). However, expression of the shRNAs for 24h before inducing TRIM28 degradation (for 24h) almost entirely rescued ERV transcript levels (Fig. 3a-b, Extended Data Fig. 11e-f), the appearance of IAP RNA FISH foci (Extended Data Fig. 11g), transcription at super-enhancers and their associated genes (Fig. 3b, Extended Data Fig. 11e-f), and the association of RNAPII-condensates at the *Mir290-295* super-enhancer locus (Fig. 3c).

To further dissect the relationship between ERV RNA and transcriptional condensates, we performed in vitro reconstitution experiments. We purified recombinant, mCherry-tagged C-terminal domain (CTD) of RNAPII, which was previously shown to form condensates in vitro [30,33], and mixed it with fluorescein-labeled in vitro-transcribed IAP RNA fragments. The IAP RNA fragments facilitated RNAPII CTD droplet formation in a dose-dependent manner (Fig. 3d-3e, Extended Data Fig. 12a), and the IAP RNA was enriched within RNAPII CTD droplets in a dose-dependent manner (Fig. 3d, f). IAP RNA also facilitated condensation of the intrinsically disordered region (IDR) of the MED1 Mediator subunit Extended Data Fig. 12b-f), a frequently used in vitro model of Mediator [31,33,36]. Furthermore, IAP RNA enhanced droplet formation of purified recombinant HP1α (an in vitro model of heterochromatin [37]), but the optimal concentration of the RNA was about 5-fold lower for HP1α than MED1 IDR in this in vitro system (Fig. 3g, Extended Data Fig. 13a-b). As expected, various other RNAs e.g. super-enhancer RNA [35] and RNA from major satellite repeats [38] also enhanced droplet formation of MED1 IDR and HP1α in vitro, but the difference in the optimal RNA concentration stayed consistently about 5-fold (Extended Data Fig. 13a-b). Moreover, IAP RNA fragments facilitated partitioning of both the MED1 IDR and NFYC IDR into IAP-RNA -containing heterotypic droplets (Fig. 3h-i, Extended Data Fig 12h-j). These results indicate that IAP RNA can enhance droplet formation of key transcriptional regulatory proteins, and suggest a mechanistic basis for the difference of the effect of RNA on heterochromatin and transcriptional condensates.

RNAPII and Mediator-containing condensates are thought to be anchored at super-enhancers by transcription factors (TFs) that are enriched at these sites [36]. One would thus expect that the loss of transcriptional condensates at super-enhancers in TRIM28-degraded cells is rescued by overexpression of super-enhancer –enriched transcription factors. To test this idea, we generated degradation-sensitive TRIM28-FKBP alleles in an induced pluripotent stem cell (iPSC) line that contains integrated transgenes encoding OCT4, SOX2, KLF4 and MYC under a doxycycline-inducible promoter (Fig. 4a-b, Extended Data Fig. 14a-b) [39]. OCT4, SOX2 and KLF4 are highly enriched at super-enhancers in ESCs [40]. TRIM28 degradation in the iPSCs led to the appearance of IAP foci as revealed by IAP RNA-FISH (Fig. 4c-d). Overexpression of OCT4, SOX2, KLF4 and MYC significantly reduced the fraction of iPSCs containing IAP foci (Fig. 4c-d), and overall *IAP* RNA level in the cell population (Fig. 4e). Furthermore, RNAPII puncta co-localized with the *MiR290-295* super-enhancer locus more frequently in TRIM28-degraded cells that overexpressed OCT4, SOX2, KLF4 and MYC (Fig. 4f) while the overall levels of RNAPII subunits did not change (Extended Data Fig. 14c-d). OCT4, SOX2, KLF4 and MYC overexpression also partially rescued the downregulation of the *MiR290-295* super-enhancer RNA and *Pri-MiR290-295* transcript in TRIM28-degraded cells (Fig. 4g) and nascent transcript levels at the *Klf4*, *Fgf4*, *Oct4* and *Mycn* super-enhancer loci (Extended Data Fig. 14e-h). These results suggest that forced expression of super-enhancer –binding TFs prevents the loss of transcriptional condensates at the *MiR290-295* super-enhancer locus in TRIM28-degraded cells.

De-repressed ERV loci appear to compete for transcriptional condensates with super-enhancers, in part through producing RNA that facilitates condensation of transcriptional activators. To directly probe this competition, we investigated whether simultaneously activated transcription at repetitive loci (e.g. ERVs) could compromise transcription at super-enhancers. First, we attempted to activate IAPs using CRISPRa [41], but targeting a dCas9-VP64 protein to IAPs with several guide RNAs failed to produce meaningful transcription at those elements (not shown). We then mimicked the effect of simultaneous ERV induction by generating three independent mESC lines that contain different copy numbers of an integrated Piggyback transposon encoding a Dox-inducible GFP transgene (Fig. 4h, Extended Data Fig 15a-c). Induction of the GFP transgenes in cells that contain the highest copy number of the transposon lead to a significant downregulation of super-enhancer transcription (Fig. 4i, Extended Data Fig 15d), and even

compromised *IAP* and *Cthrc1* induction in TRIM28-degraded cells (Fig 4i, Extended Data Fig. 15c). These results demonstrate that simultaneous activation of model transposon loci may compromise super-enhancer transcription in mESCs.

The above results suggest that pluripotent stem cells fail to maintain transcription of super-enhancer -driven pluripotency genes when ERV repression is compromised, and predict that the amount of ERV products would correlate with the inability of embryos to maintain a pluripotent compartment. To test this model in vivo, we used our recently developed zygotic perturbation platform (Fig. 5a) [42]. We generated zygotic deletion mutants of TRIM28, SETDB1, HP1α and other epigenetic regulators implicated in ERV repression, and assayed the timing and amount of the GAG protein produced by IAPs (Fig. 5b-c). IAP GAG foci were detected in E3.5 blastocysts of TRIM28, SETDB1 and KDM1A KO mutants, and these mutations were lethal at ~E6.5 in embryos. In contrast, IAP GAG foci appeared ~3 days later in DNMT1 KO embryos (Fig. 5b-c), and earlier appearance of IAP GAG foci correlated with the onset of embryonic lethality (Fig. 5c-d).

To probe which cell types are affected by the presence of IAP foci in blastocysts, we used scRNA-Seq. We created a reference cell state map of an early mouse embryo spanning a window of developmental stages (E5.5 to E7.0) that encompass the reported onset of lethality in TRIM28 knock-out mice (~E6.5) (Extended Data Fig. 16a-e) [7]. We then generated TRIM28-deficient embryos using Cas9/sgRNA delivery into zygotes, and mapped cell states using scRNA-Seq (Fig. 5e, Extended Data Fig. 16f). qRT-PCR data confirmed that IAPs were upregulated in TRIM28 KO embryos (Fig. 5f, Extended Data Fig. 16g). The E6.5 scRNA-Seq data revealed a dramatic scarcity of epiblast cells normally derived from pluripotent cells of the inner cell mass (ICM) (Fig. 5e), and an abundance of extraembryonic lineages (e.g. parietal endoderm) in TRIM28 KO embryos (Fig.5e, Extended Data Fig. S17a-b). Immunofluorescence imaging showed that the pluripotency factors NANOG, OCT4, SOX2 and KLF4 were already virtually absent in the ICM of TRIM28 KO E3.5 blastocysts (Fig. 5g, Extended Data Fig. 18a-b), and the inner part of the blastocysts was instead populated by cells expressing the endoderm marker GATA6 (Fig. 5g). This phenotype is reminiscent of NANOG knockout embryos, in which the pluripotent ICM is replaced by GATA6-expressing parietal endoderm cells (Fig. 5g) [43]. These

data are consistent with upregulation of endoderm markers observed in TRIM28-degraded ESCs (Extended Data Fig. 19a-b). Overall the findings suggest that extended ERV de-repression results in the loss of expression of pluripotency genes and consequent depletion of pluripotent cells in early mouse embryos.

The results presented here support a model that ERV retrotransposons have the capacity to hijack biomolecular condensates formed by key transcriptional regulatory proteins in pluripotent cells (Fig. 5h). This model may help explain why thousands of retrotransposition-incapable ERVs are repressed in mammals and how their reactivation could alter cellular fates in the absence of retrotransposition [7,9,14-16,23,44,45]. De-repressed ERVs appear to compete with super-enhancers for transcriptional condensates in pluripotent cells, and this competition may explain why simultaneous transcription of multiple ERV loci are associated with embryonic lethality. Furthermore, recent studies reported that m6A methylation plays important roles in repressing ERV transcripts [11-13], and the contribution of ERV RNA to the hijacking of transcriptional condensates from super-enhancers may explain the in vivo role and importance of ERV RNA modifications. Many nuclear non-coding RNAs are known to localize to the loci where they are produced [46], but their functions are mysterious. We propose, that such RNAs – like ERV RNAs – may contribute to the genomic distribution of nuclear condensates. Sets of ERVs are transcribed in various disease contexts include leukemias [47] and neurological diseases [48-51]; suggesting that the ERV RNA produced in those contexts may contribute to the pathologies through altering nuclear condensate homeostasis.

Acknowledgments:

We thank Tugce Aktas and Aydan Bulut-Karslioglu for comments on the manuscript, Charles Haggerty for S2 cells, Georg Winter and Martin Jäger for sharing RNAPII antibody, Alessandra Dall'Agnese and Eric Guo for advice on FISH-IF, Alexandra Mattei for help with IF, Shaon Basu for helping with co-localization analysis, Sebastian Mackowiak for help with code, Charlotte Althoff for help with Western blots, Beate Lukaszewsa-McGreal for help with Mass-

spectrometry, Celine Hillgardt, Christin Franke and the MPIMG transgenic facility for help with mouse work, and the MPIMG Sequencing core for sequencing. **Funding**: This work was funded by the Max Planck Society, and partially supported by grants from the NIH (1P50HG006193 to A.M.) and the Deutsche Forschungsgemeinschaft (DFG) Priority Program Grant HN 4/1-1 (to D.H.); the Austrian Academy of Sciences and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC-CoG-866166, RiboTrace) (to S.L.A.). H.N is supported by fellowships from the Emil Aaltonen Foundation and Orion Research Foundation. **Author contributions**: Conceptualization: V.A., S.H., A.S.K., H.N., A.M., D.H.; Investigation: V.A., A.S.K., H.N., C.R., J.N., N.F., Z.D.S., B.T., L.W.; Methodology: N.F., N.P., S.H., H.N.; Software: S.H., N.P.; Formal analysis: V.A., S.H., H.N., R.B.; Resources: S.L.A., A.M., D.H.; Data curation: S.H., Visualization: V.A., S.H., A.S.K., H.N., C.R., J.N., Writing: V.A., S.H., A.S.K., H.N., A.M., D.H.; Supervision: A.M., D.H., Project administration: D.H.; Funding acquisition: S.L.A., A.M., D.H. **Competing interests**: S.L.A. is founder and advisor of QUANTRO Therapeutics GmbH. A patent related to SLAMseq has been awarded to S.L.A. (EP3589748B1) and licensed to Lexogen GmbH. **Data and materials availability**: All data is available in the supplementary materials. Sequence data were deposited at GEO under the accession number GSE159468. Mass spectrometry data were deposited at ProteomeXchange under the accession ID PDX021895. **Code availability:** All custom code used in this study is available upon request.

Reviewer access to deposited data:
Go to https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE159468
Enter token upgxequwxzylhgx into the box

A UCSC track hub for reviewers:
http://ngs.molgen.mpg.de/ngsuploads/dept_meissner/Trim28_Submission.txt

http://proteomecentral.proteomexchange.org/cgi/GetDataset
Username: reviewer_pxd021895@ebi.ac.uk
Password: zbELCqWf

Figures

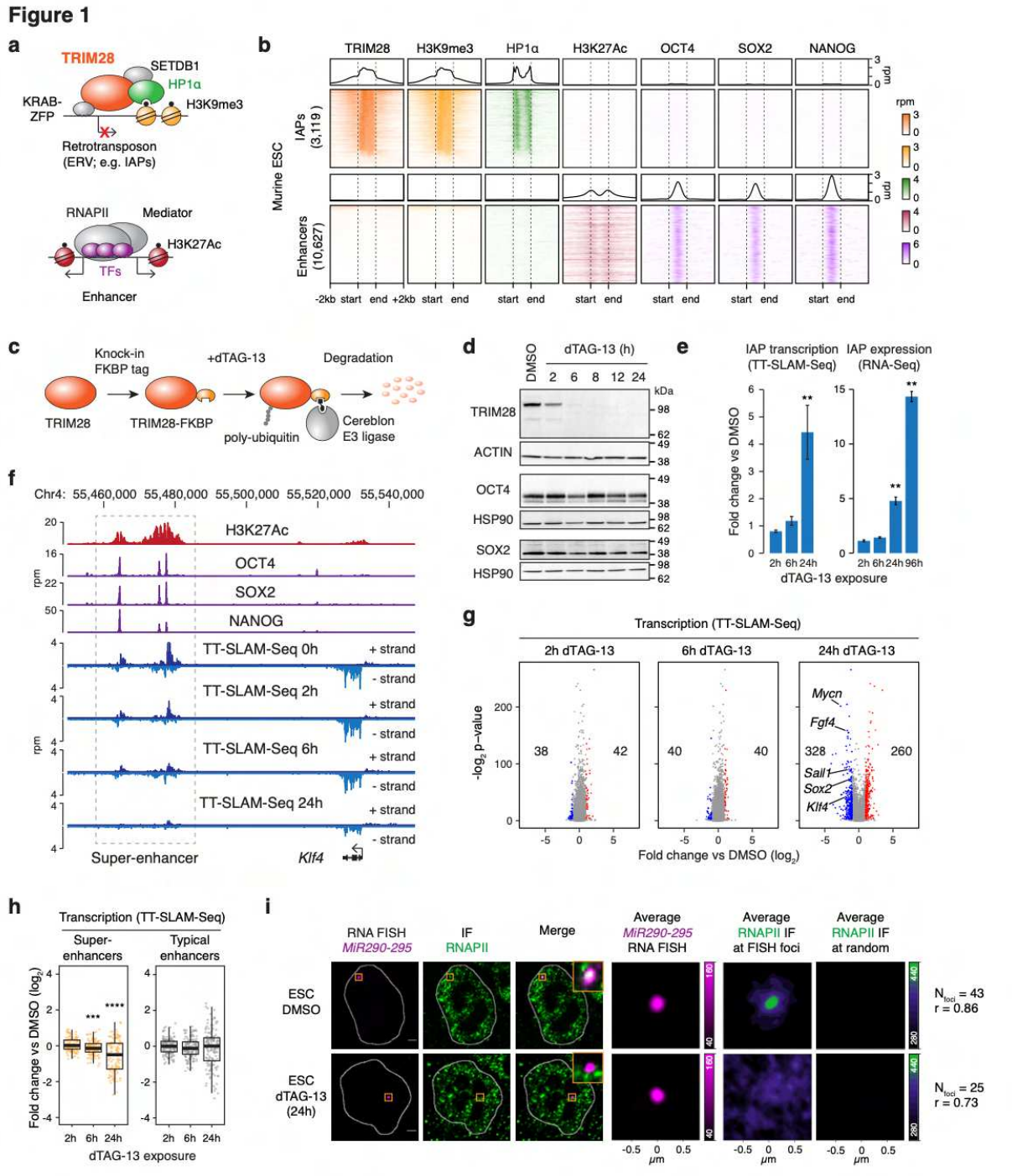


Figure 1. TRIM28 degradation leads to the reduction of super-enhancer transcription and loss of transcriptional condensates at super-enhancers in mESCs

a. (top) Model of TRIM28/HP1α pathway and its major regulators at endogenous retroviruses (ERVs). KRAB-ZFP: KRAB-Zinc Finger Protein, TRIM28: Tripartite motif -containing protein 28, HP1α: heterochromatin protein 1α, IAP: intracisternal A-type particle. (bottom) Model of enhancers. Enhancers are bound by transcription factors (TFs) that recruit the Mediator complex and RNA Polymerase II (RNAPII).

b. Heatmap representation of ChIP-seq read densities of the indicated factors within a 2kb window around full-length IAP ERVs and enhancers in mESC. The genomic elements (the middle part of the plots) were length normalized. Enhancers in mESCs are bound by the OCT4, SOX2, NANOG TFs, and are devoid of components of the TRIM28/HP1α pathway that are bound at IAP ERVs. Enhancers include the constituent enhancers of super-enhancers and typical enhancers. Rpm: reads per million.

c. Scheme of the dTAG system to degrade TRIM28 in mESCs.

d. Western blot validation of the FKBP degron tag and its ability to degrade TRIM28. The protein levels of the pluripotency factors OCT4 and SOX2 do not change.

e. TRIM28-degradation leads to de-repression of IAP ERVs. Displayed is fold change in read density from three replicates of TT-SLAM-Seq and RNA-Seq data after the indicated duration of dTAG-13 treatment, normalized to the level in DMSO control.

f. Acute reduction of transcription at the *Klf4* super-enhancer locus upon TRIM28-degradation. Displayed are genome browser tracks of ChIP-Seq data (H3K27Ac, OCT4, SOX2, NANOG) in control mESCs and TT-SLAM-Seq data upon 0h, 2h, 6h and 24h dTAG-13 treatment at the *Klf4* locus. Co-ordinates are mm10 genome assembly co-ordinates.

g. Volcano plots displaying fold change of gene transcription (TT-SLAM-Seq data) upon dTAG-13 treatment. The number of significantly de-regulated genes, and genes implicated in the control of pluripotency are highlighted.

h. Reduced transcription at super-enhancers in TRIM28-degraded cells. The box plots show $\log_2$ fold change in transcription (TT-SLAM-Seq read density) at super-enhancers and typical enhancers upon dTAG-13 treatment normalized to untreated control mESCs. P values are from two tailed, one sample t tests. ****: P<10-4, ***: P<10-3, *: P<0.05.

i. Co-localization between the nascent RNA of *MiR290-295* and RNAPII puncta in mESCs. The co-localization is lost upon TRIM28-degradation. Displayed are separate images of individual z-slices (same z) of the RNA-FISH and IF signal, and an image of the merged channels. The nuclear periphery determined by DAPI staining (not shown) is highlighted as a white contour. Also shown are averaged signal of either RNA FISH or RNAPII IF centered on the *MiR290-295* FISH foci or randomly selected positions. Scale bars: 2.5μm.

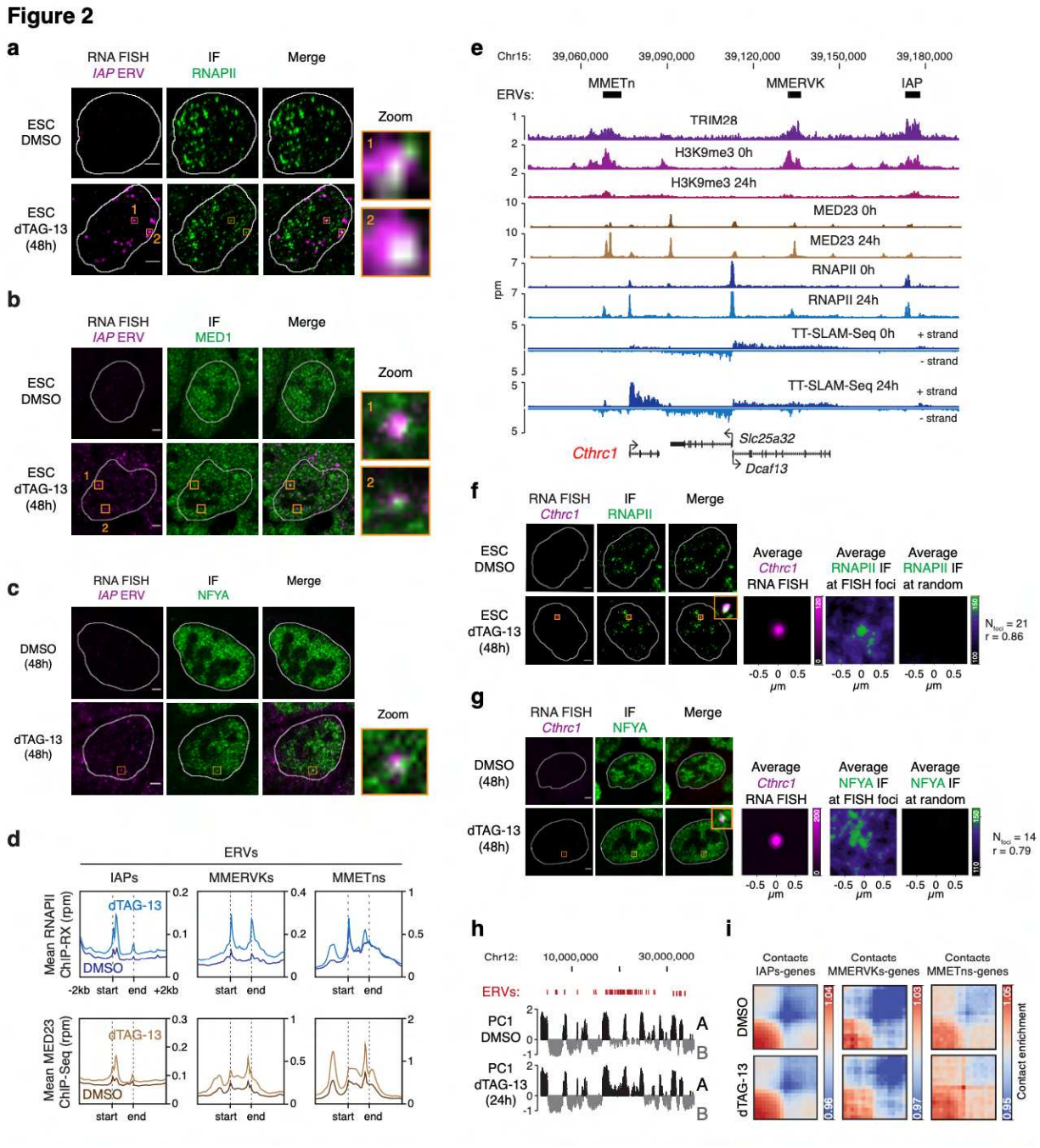


Figure 2. De-repressed IAPs form nuclear foci that associate with RNAPII condensates, and incorporate nearby genes

a-c. Co-localization between the *IAP* RNA and **(a)** RNAPII puncta **(b)** Mediator puncta and **(c)** NFYA puncta in TRIM28-degraded mESCs. Displayed are separate images of individual z-slices (same z) of the RNA-FISH and IF signal, and an image of the merged channels. The nuclear

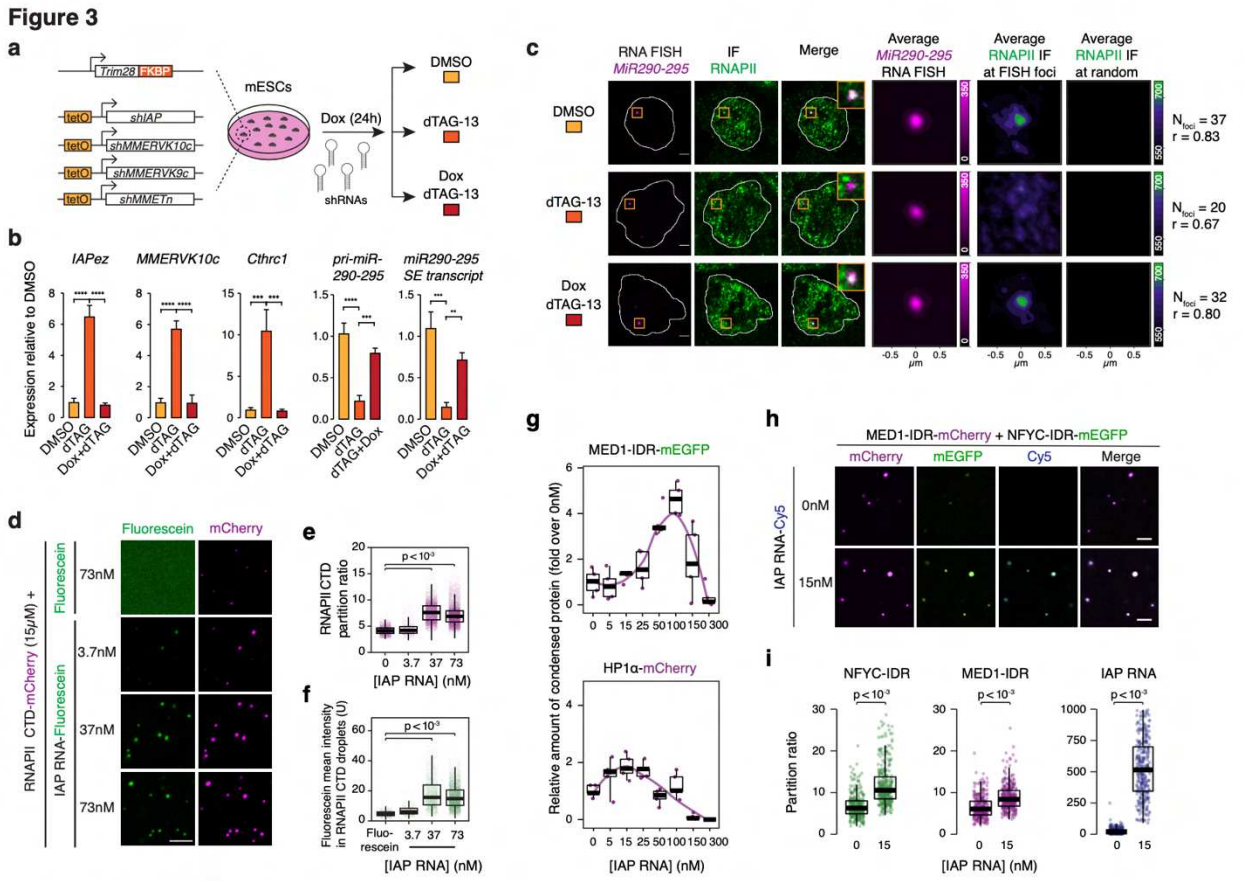
periphery determined by DAPI staining (not shown) is highlighted as a white contour. The zoom column displays the region of the images highlighted in a yellow box zoomed in for greater detail. Scale bars: 2.5µm.

d. Elevated RNAPII and Mediator occupancy at various ERV classes in TRIM28-degraded ESCs. Displayed are meta representations of RNAPII ChIP-RX read densities at IAP, MMERVK and MMETn ERVs in control (DMSO) and dTAG-13 (24h) treated mESCs. The mean read densities are displayed +/-2kb around the indicated elements. The genomic elements (the middle part of the meta plot) were length normalized.

e. Genome browser tracks at the *Cthrc1* locus. The *Cthrc1* gene is surrounded by three ERVs that lose H3K9me3 signal and gain occupancy of transcription-associated marks upon TRIM28-degradation. Note the independent transcription initiation events at *Cthrc1* and MMETn, ruling out that MMETn acts as an alternative *Cthrc1* promoter. Rpm: reads per million. Co-ordinates are mm10 genome assembly co-ordinates.

f-g. Co-localization between the nascent RNA of *Cthrc1* and **(f)** RNAPII puncta and **(g)** NFY puncta in TRIM28-degraded mESCs. Displayed are separate images of individual z-slices (same z) of the RNA-FISH and IF signal, and an image of the merged channels. The nuclear periphery determined by DAPI staining (not shown) is highlighted as a white contour. Also shown are averaged signal of either RNA FISH or IF centered on the *Cthrc1* FISH foci or randomly selected positions. Scale bars: 2.5µm.

h. Principal component (PC) plot of Hi-C interactions at an ERV-rich locus on chromosome 12.
i. Pile-up analysis of contacts between IAPs, MERVKs, MMETns and transcribed genes in wild type and TRIM28-degraded mESCs.



372

373

374 **Figure 3. Contributions of IAP RNA to condensate localization in vivo and condensate**

375 **formation in vitro**

376 **a.** Schematic model of the ERV shRNA knockdown experiments

377 **b.** Rescue of Mir290-295 super-enhancer and primary transcript levels upon ERV RNA-

378 knockdown in TRIM28-degraded cells. The bar plots show qRT-PCR data as fold change

379 normalized to the DMSO treatment control. P values are from two-tailed t tests. ****: $P<10^{-4}$,

380 ***: $P<10^{-3}$, **: $P<10^{-2}$.

381 **c.** Co-localization between the nascent RNA of *MiR290-295* and RNAPII puncta in TRIM28-

382 degraded mESCs upon ERV RNA-knockdown. Displayed are separate images of individual z-

383 slices (same z) of the RNA-FISH and IF signal, and an image of the merged channels. The

384 nuclear periphery determined by DAPI staining (not shown) is highlighted as a white contour.

385 Also shown are averaged signal of either RNA FISH or RNAPII IF centered on the *MiR290-295*

386 FISH foci or randomly selected positions. Scale bars: 2.5μm.

d. *IAP* RNA is enriched within RNAPII CTD droplets. Displayed are representative images of mixtures of fluorescein-labeled *IAP* RNA and purified recombinant RNAPII CTD-mCherry in droplet formation buffer. Scale bar: 5μm.

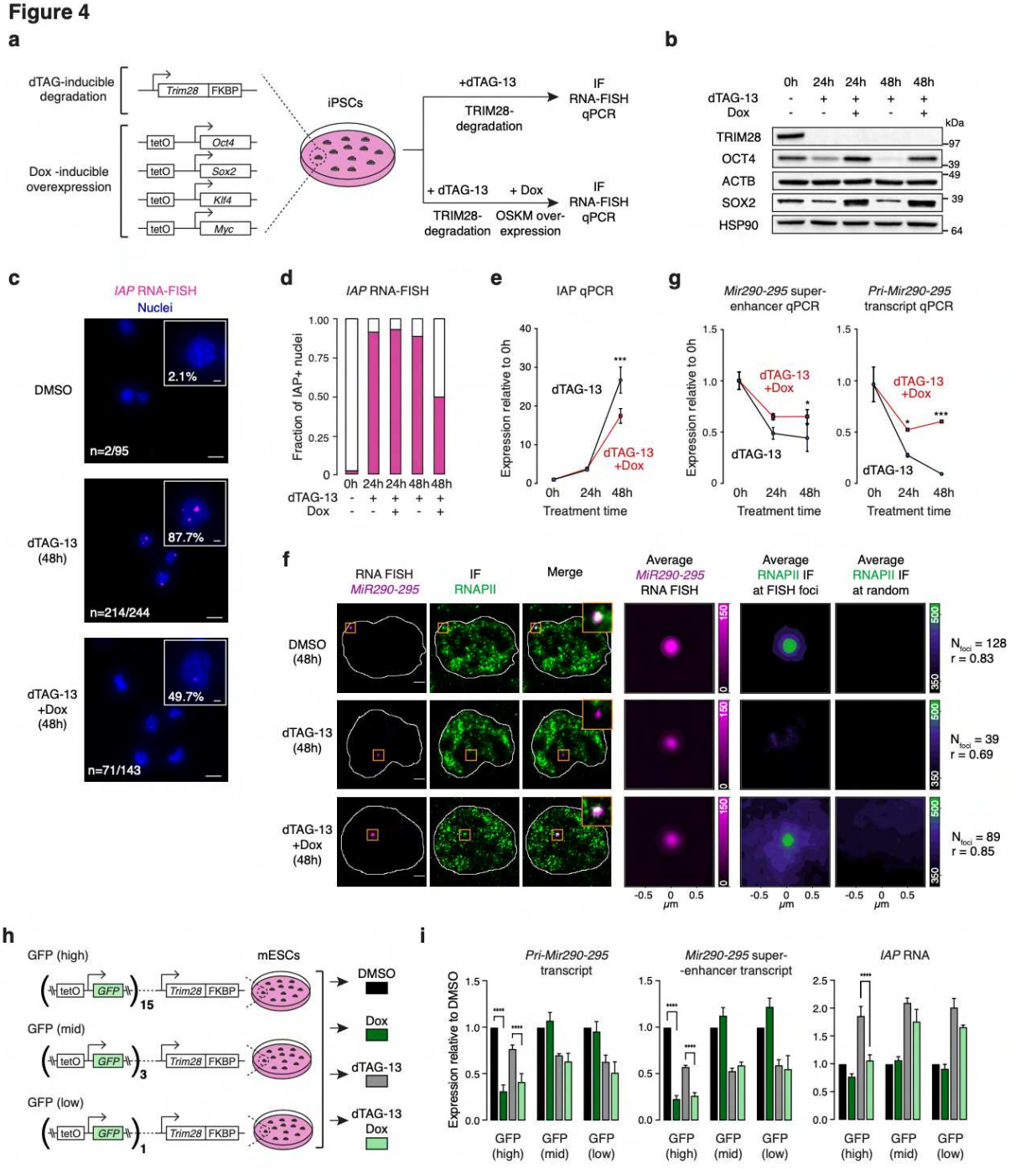
e. Partitioning ratio of RNAPII CTD-mCherry into droplets at the indicated IAP RNA concentrations. Every dot represents a detected droplet.

f. Quantification of the enrichment of fluorescein-labeled *IAP* RNA in RNAPII CTD-mCherry droplets.

g. Quantification of the partitioning of (top) MED1 IDR and (bottom) HPIα into droplets in the presence of *IAP* RNA. Values are normalized against the partition ratio at no RNA added. Corresponding images are found in Extended Data Fig. S13a.

h. The NFYC IDR forms heterotypic droplets with MED IDR in vitro, and IAP RNA facilitates partitioning of both proteins into heterotypic droplets. Displayed are representative images of droplet formation by purified NFYC IDR-mEGFP (1μM) and MED1 IDR-mCherry (5μM) fusion proteins in the presence of in vitro transcribed Cy5-labelled IAP RNA fragment. Scale bar: 5μm.

i. Partitioning ratio of NFYC IDR-mEGFP, MED1 IDR-mCherry and IAP RNA into droplets at the indicated IAP RNA concentrations. Every dot represents a detected droplet. All pairwise P-vales $< 2.2*10^{-16}$ (Welch's t-test).



405

406 **Figure 4. Super-enhancer -enriched transcription factors and repetitive RNA-producing**

407 **loci compete for transcriptional activators**

408 **a.** Genotype of the iPSC line and scheme of the experimental setup. The iPSC cell line contains

409 degradation-sensitive *Trim28-FKBP* alleles, and Doxycycline-inducible *Oct4*, *Sox2*, *Klf4*, *c-Myc*

410 (OSKM) transgenes.

b. Western blot validation of the FKBP degron tag and OSKM ectopic expression in iPSCs.

c. OSKM ectopic expression reduces the number of IAP-expressing cells in the absence of TRIM28. Displayed are representative images of IAP RNA FISH staining. The number and percentage refer to cells with detectable IAP foci. Scale bars: 10µm, inset scale bars: 2µm.

d. Quantification of cells with detectable IAP foci (i.e. IAP+ cells) at the indicated treatment regimes.

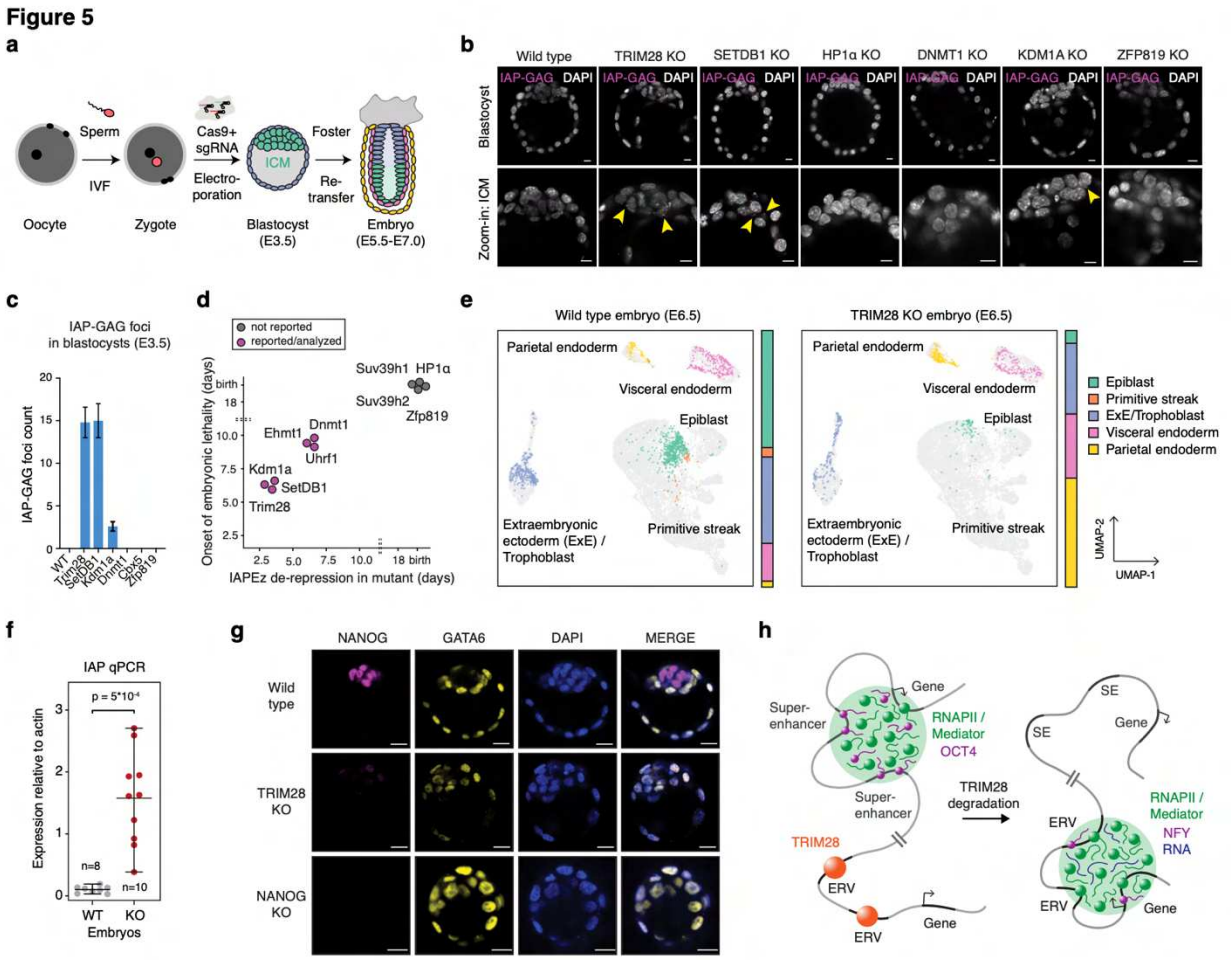
e. *IAP* RNA expression is reduced in TRIM28-degraded iPSCs that ectopically express OSKM factors. The line plot shows qRT-PCR data of IAP RNA levels normalized to 0h of dTAG-13 treatment. *** $p<10^{-3}$ two-tailed t test.

f. Co-localization between the nascent RNA of *MiR290-295* and RNAPII puncta in TRIM28-degraded iPSCs that ectopically express OSKM factors. Displayed are separate images of individual z-slices (same z) of the RNA-FISH and IF signal, and an image of the merged channels. The nuclear periphery determined by DAPI staining (not shown) is highlighted as a white contour. Also shown are averaged signal of either RNA FISH or RNAPII IF centered on the *MiR290-295* FISH foci or randomly selected positions. Scale bars: 2.5µm.

g. Elevated levels of *MiR290-295* super-enhancer transcript and *Pri-MiR290-295* nascent transcript in TRIM28-degraded iPSCs that ectopically express OSKM factors. The line plots show qRT-PCR data normalized to the 0h of dTAG-13 treatment. * $p<0.05$, *** $p<10^{-3}$ two-tailed t test.

h. Schematic model of the experiment mimicking ERV transcription. mESC lines harboring degradation-sensitive TRIM28-FKBP alleles and 15, 3 or 1 copies of a Piggyback transposon encoding a Dox-inducible GFP transgene were treated with Dox (to induce GFP expression) and/or dTAG-13 (to induce TRIM28 degradation).

i. Simultaneous transcription of a high copy number of GFP transgenes leads to a reduction of transcription at the *MiR290-295* super-enhancer and its associated gene. The bar plots show qRT-PCR data as fold change normalized to the DMSO control treatment. P values are from two-tailed t tests. ****: $P<10^{-4}$, ***: $P<10^{-3}$.



438

439

440 **Figure 5. Early activation of IAP ERVs correlates with onset of lethality and depletion of**
441 **pluripotent cells in mouse embryos**

442 **a.** Scheme of the zygotic CRISPR/Cas9-perturbation platform.

443 **b.** IF images of mouse E3.5 blastocysts stained for the GAG protein produced by IAPs. Nuclei
444 are counterstained with DAPI. Note the purple IAP GAG foci highlighted with yellow
445 arrowheads. Scale bar: 10 µm.

446 **c.** Quantification IAP GAG foci in multiple embryos of the indicated genotypes.

447 **d.** The timing of IAP de-repression correlates with the onset of lethality. The plots show the day
448 when the indicated mutations are embryonic lethal versus when IAP foci are first detected in
449 mutant embryos.

e. Epiblast cells are depleted in TRIM28 KO embryos. UMAP of E6.5 wild type and E6.5 TRIM28 KO embryos mapped on the combined reference cell state map. The proportions of cells that belong to the individual cell states are indicated as a bar on the right of the UMAP plots.

f. Elevated *IAP* RNA level in TRIM28 KO mouse embryos. The plot shows qRT-PCR data of *IAP* expression in E3.5 wild type and TRIM28 KO embryos. "n" indicates the number of embryos included in the experiment. P value is from a two-tailed t-test.

g. The inner part of TRIM28 KO blastocysts is populated by GATA6-expressing, NANOG-negative cells. Displayed are representative immunofluorescence images of NANOG and GATA6 in E3.5 wild type, TRIM28 KO and NANOG KO blastocysts. Scale bars: 20μm.

h. Condensate hijacking model. In pluripotent cells, transcriptional condensates associate with super-enhancers bound by pluripotency transcription factors (e.g. OCT4). In the absence of TRIM28, transcriptional condensates are lost from super-enhancers, and associate with de-repressed ERVs.

Materials and Methods

Cell culture

V6.5 mouse Embryonic Stem Cells (mESCs) and induced Pluripotent Stem Cells (iPSCs) were cultured on irradiated primary Mouse Embryonic Fibroblasts (MEFs) under standard Serum/LIF conditions [knockout DMEM containing 15% fetal bovine serum (FBS), supplemented with 1X penicillin/streptomycin, 1X GlutaMAX supplement, 1X non-essential amino acids, 0.05mM ß-mercaptoethanol (all from Gibco) and 1000 U/ml leukemia inhibitory factor (LIF)].

For ChIP-Seq, TT-SLAM-Seq and RNA-Seq experiments, mESCs were depleted from MEFs by incubating them on gelatin-coated cell culture plates for 45 min at 37ºC allowing MEFs to attach while mESCs remain in suspension. MEF depletion was performed two times after which mESCs were seeded on gelatin-coated plates and maintained in Serum/LIF conditions with 2000 U/ml LIF.

For RNA FISH combined with immunofluorescence, MEF-depleted cells were grown on round 18mm glass coverslips (Roth LH23.1). Coverslips were coated with 5 µg/ml of poly-L-ornithine (Sigma-Aldrich, P4957) for 30 min at 37°C and with 5µg/ml of Laminin (Corning, 354232) overnight at 37°C.

To perturb RNAPII condensates, cells were treated 30 minutes with 1.5% 1.6-hexanediol (Sigma) in Serum/LIF conditions with 2000U/ml LIF (Extended Data Fig. 7c-e).

Generation of the TRIM28-FKBP ESC line

To knock in the degradation-sensitive $FKBP^{F36V}$ tag at the N-terminus of TRIM28, a repair template containing homology arms spanning upstream and downstream of the target site was cloned into a pUC19 vector (NEB) (Extended Data Fig. 2a). The repair template included a mRuby2 fluorescent protein sequence, P2A linker and the FKBP tag sequence (Extended Data Fig. 2a) [26]. mRuby2 sequence was amplified from the mRuby2-N1 plasmid (Addgene #54614) and the P2A-FKBP sequence was amplified from the PITCh dTAG donor vector (Addgene #91792). A guide RNA (Supplementary Table 1b) targeting the N-terminus of TRIM28 was cloned into the

sgRNA-Cas9 vector pX458 (Addgene: 48138). The repair template and the sgRNA-Cas9 vector were transfected into V6.5 mESCs and iPSCs by nucleofection using Amaxa 4D Nucleofector X Unit (Lonza) according to manufacturer's instructions. To screen for positive integrations, the transfected cells were sorted for mRuby2 fluorescent protein expression with flow cytometry. The sorted cells were seeded as single-cells and expanded for few days. Single colonies were picked and genotyped for the correct integration with PCR.

Generation of theTRIM28-FKBP iPSC line

To generate mouse secondary induced pluripotent stem cells (iPSCs), secondary MEFs that harbor Doxycycline (Dox) -inducible *Oct4*, *Sox2*, *Klf4* and *c-Myc* (OSKM) transgenes, and a NANOG::GFP reporter [39], were reprogrammed. MEFs were seeded at a low density of about 10,000-25,000 cells per well of a 12-well plate that was coated with 0.2% gelatin. Prior to induction of the transgenes, cells were synchronized with a 12-16h incubation in 2.5% FBS containing medium. Afterwards, the medium was switched to 15% FBS/LIF with 2µg/ml Doxycycline. Media was changed every day; colonies emerged after about a week and clonal iPSC lines were resolved within 2 weeks. The iPSC line with the TRIM28 FKBP knock-in alleles was generated as described above.

Inactivation of NANOG::GFP in the TRIM28-FKBP iPSC line

For immunofluorescence combined with RNA FISH experiments (Fig. 4f, Extended Data Fig. 14c), GFP at the *Nanog::GFP* locus was inactivated by targeting with a guide RNA (Supplementary Table 1b) against the GFP sequence. CRISPR-Cas9-guide RNA Ribonucleoprotein complex (RNPs) was assembled following manufacturer's guidelines (Alt-R CRISPR; IDT). Briefly, 2µl each of 200µM of crRNA and 200µM of tracrRNA was mixed and denatured at 95°C for 5 minutes, and allowed to anneal to room temperature for 20 minutes. 3µl of crRNA-tracrRNA was mixed with 1µl of 61µM Cas9 20 minutes prior to transfection. TRIM28-FKBP iPSCs were transfected with the assembled RNPs by nucleofection using Amaxa 4D Nucleofector X Unit (Lonza) according to manufacturer's instructions. The cells were then sorted for non-GFP cells using flow cytometry. The sorted cells were expanded and cultured as a clonal line.

Generation of shRNA knockdown mESC line for IAPEz and IAPEz/MMERVK10c/ /MMERVK9c/ MMETn

shRNA against the repeat elements was designed using the consensus sequence of the inner part of each element. shRNA sequences are listed in Supplementary Table 1b. shRNA sequences were cloned into an inducible plasmid following the standard cloning protocol recommended for this vector. Lentivirus with the shRNA constructs was generated by transfecting HEK293 cells along with packaging and envelope plasmids. TRIM28-FKBP mESCs were transduced with the virus with 8µg /mL polybrene for 48 hours followed by treatment with 2µg/ml puromycin to select for the transduced cells. Single colonies were picked, characterized and expanded as a clonal line for IAPEz and as a bulk line for the quadruple knockdown line for IAPEz, MMERVK10c, MMERVK9c, MMETn. shRNA expression was induced by treatment with 2µg/ml Doxycycline. For the experiment with pre-induction of shRNAs (Fig. 3a-c, Extended Data Fig. 11e-g), the cells were treated for 24 hours with Doxycycline followed by DMSO/dTAG-13/Dox+dTAG-13 for another 24 hours. Tet-pLKO-puro was a gift from Dmitri Wiederschain (Addgene plasmid # 21915). psPAX2 was a gift from Didier Trono (Addgene plasmid # 12260). pCMV-VSV-G was a gift from Bob Weinberg (Addgene plasmid # 8454).

Integration of Piggyback transposon encoding a Dox-inducible GFP transgene

The PB-tetO-GFP construct (Dox-inducible GFP) was created by digesting an "all-in-one" Piggybac, TREG/Tet-3G plasmid (Addgene plasmid # 97421; a gift from Steven Carr & Samuel) with restriction enzymes NcoI and KpnI and cloning following oligonucleotides to insert a start-codon "ATG" in front of the GFP (PG_ATG_GFP: 5´cttcctaccctcgtaaaggtctagagctagccaccatgggtaccgtgagcaagggcgaggagctgttcaccggggtggtg 3´) via NEBuilder HiFi Assembly. The construct was randomly integrated in the mESC V6.5 Trim28-FKBP subclone by co-transfecting $4x10^5$ cells with 8.5µg PB-tetO-GFP and 1.5µg Super PiggyBac transposase expression vector (SBI, PB210PA-1) using FuGENE HD Transfection Reagent (Promega). After 4 days of Puromycin selection (2µg/ml) single colonies were picked, expanded and tested for doxycycline-inducibility of the GFP construct monitored by GFP detection with FACS. Three clones expressing low, intermediate and high levels of GFP were selected and expanded for subsequent experiments.

Deletion of ERVs in the TRIM28-FKBP mESC line

To generate mESC line with three ERV deletions at the *Cthrc1* locus, deletions of MMETn, MMERVK and IAP sequences were performed sequentially with CRISPR/Cas9. Guide RNAs (Supplementary Table 1b) flanking the individual ERV elements were cloned into the sgRNA-Cas9 vector pX458 and the two vectors for each cut were delivered to cells with Lipofectamine 2000 (Thermo Fisher, 11668027) according to manufacturer's instructions. Transfected cells were sorted based on GFP expression two days later with flow cytometry. Single colonies were picked, genotyped (Extended Data Fig. 7f, Supplementary Table 1b) and clones with homozygous deletion were selected for the next ERV deletion. Deletions were also confirmed by sequencing the PCR products.

TRIM28 degradation

Prior to treatment, cells were seeded onto 0.2% gelatin coated plates after two rounds of MEF depletion. For degradation of TRIM28, 500nM of dTAG-13 compound [26] was mixed with mESC media (Supplemented with 2000 U/mL LIF) and incubated for the indicated time duration.

Western blot

Cultured cells were lysed in RIPA buffer for 30 minutes at 4ºC, and centrifuged for 20 minutes at maximum speed. The supernatant was then transferred to a new tube and quantified by BCA assay (Thermo Scientific). 10µg of extracted protein was run on a 4-12% NuPAGE SDS gel and transferred onto a PVDF membrane using iBlot2 Dry Gel Transfer Device (Invitrogen) according to manufacturer's instructions. The blots were blocked with 5% skim milk in TBST and incubated with primary antibodies. Primary antibodies used in this study include TRIM28 (ab22553; 1:500), ACTB (ab8226; 1:1000), OCT4 (ab19857; 1:500), OCT4 (sc-5279; 1:500), HSP90 (BD610419; 1:4000), SOX2 (ab79351; 1:500). HRP-conjugated secondary antibodies were used against the host species at 1:3000 – 1:5000 dilution and visualized with HRP substrate SuperSignal West Dura (Thermo Scientific) (Fig. 1d, 4b, Extended Data Fig. 2b, 14a).

Proteomics Sample Preparation and LC-MS/MS Instrument Settings for Shotgun Proteome Profiling and Data Analysis

Proteomics sample preparation was done according to a published protocol with minor modifications [52]. In brief, three biological replicates of dTAG-13 -treated samples with 4 million cells per replicate were lysed under denaturing conditions and sequentially digested with LysC and Trypsin (Roche, Basel, Switzerland). Peptide desalting was performed according to the manufacturer's instructions (Pierce C18 Tips, Thermo Scientific, Waltham, MA). Desalted peptides were further separated into four fractions by strong cation exchange chromatography (SCX, 3M Purification, Meriden, CT). LC-MS/MS was carried out by nanoflow reverse phase liquid chromatography (Dionex Ultimate 3000, Thermo Scientific) coupled online to a Q-Exactive HF Orbitrap mass spectrometer (Thermo Scientific), as reported previously [53]. Briefly, the LC separation was performed using a PicoFrit analytical column (75 µm ID × 50 cm long, 15 µm Tip ID; New Objectives, Woburn, MA) in-house packed with 3-µm C18 resin (Reprosil-AQ Pur, Dr. Maisch, Ammerbuch, Germany). Raw MS data were processed with MaxQuant software (v1.6.10.43) and searched against the mouse proteome database UniProtKB with 55,471 entries, released in May 2020. The MaxQuant processed output files can be found in (Supplemental Table 3), showing peptide and protein identification, accession numbers, % sequence coverage of the protein, q-values, and label-free quantification (LFQ) intensities. The mass spectrometry data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE [54] partner repository with the dataset identifier PDX021895. The correlation analysis of biological replicates and the calculation of significantly different proteins were done with Perseus (v1.6.14.0). Only groups with valid values in at least one group were used, missing values were replaced by values from the normal distribution. Statistical analysis was done by a two-sample t-test with Benjamini-Hochberg (BH, FDR of 0.05) correction for multiple testing (Extended Data Fig. 2c, 14d).

RNA isolation and quantitative Real-Time PCR (qRT-PCR)

RNA from cultured cells was isolated using RNeasy kit (Qiagen) according to manufacturer's instructions. 1µg of RNA was used for cDNA synthesis using RevertAid First Strand cDNA synthesis kit (Thermo Scientific) with random hexamer primers according to manufacturer's instructions. RNA from single blastocysts (E3.5) was isolated using PicoPure RNA isolation kit (Thermo Fischer Scientific) and cDNA was synthesized with High Capacity cDNA Reverse Transcription kit (Applied Biosystems). qRT-PCR was performed with primers (Supplementary

Table 1b) and 2X PowerUP SYBR green master mix (Applied Biosystems). Note that in Fig. 3a-b, and Extended Data Fig 11a-f, the qPCR primers used for MMETn and MERVK9c elements fail to detect induction after dTAG-13 treatment, which is explained by these elements having degenerate genomic sequences.

Immunofluorescence (IF) in mouse embryos and blastocysts

E3.5 blastocysts and E6.5 embryos were dissected from the uterus in 1X HBBS and fixed in 4% Paraformaldehyde (PFA) at 4ºC overnight. The embryos were washed three times in 1X PBS with 0.1% TritonX-100 and then permeabilized in 1X SuperBlock PBS with 0.5% TritonX-100 for one hour at room temperature. Blocking against unspecific binding was done by incubating embryos in blocking buffer (1X SuperBlock PBS with 0.2% TritonX-100, 10% Donkey Serum, 3% Bovine Serum Albumin) for two hours at room temperature. Primary antibodies were diluted in antibody buffer (1X SuperBlock PBS with 0.2% TritonX-100, 10% Donkey Serum, 0.3% Bovine Serum Albumin) and incubated at 4 ºC overnight. Primary antibodies used include TRIM28 (ab22553; 1:200), OCT4 (ab19857; 1:200), NANOG (REC-RCAB002P-F; 1:400), KLF4 (AF3158; 1:200), SOX2 (ab79351; 1:200), GATA6 (AF1700; 1:200), GATA4 (sc-25310; 1:200), SOX17 (AF1924; 1:200), IAP-GAG (MBS8566075; 1:100). Embryos were washed in blocking buffer three times for 30 minutes each and incubated with host-specific secondary antibodies conjugated to fluorescent dyes for three hours in the dark at room temperature. After washing with blocking buffer, nuclei were counterstained with 0.5µg/mL DAPI for 30 minutes in the dark at room temperature. Embryos were mounted in drops of 1X PBS and covered with mineral oil prior to imaging. Images were acquired with LSM880 (Zeiss) microscope with a 40X objective. Images were processed with ZEN 3.1 (Zeiss) and ImageJ (Fig. 5b, 5g Extended Data Fig. 18a-b).

SSEA-1 staining

Cultured cells were dissociated into single-cells with Accutase enzyme mix, washed in 1X PBS and fixed with 4% PFA for 15 minutes at room temperature. Following three washes with 1X PBS, the cells were incubated with fluorescent conjugated SSEA-1 antibody (BioLegend 125608, 125609; 1:1000) for one hour in the dark at 37ºC. After three washes in 1X PBS, nuclei were counterstained with 0.24 µg/ml DAPI and imaged with LSM880 confocal microscope at 63X oil

objective (Extended Data Fig. 2d-e). For FACS analysis, the cells were washed two times in 1X PBS+10%FBS and analyzed with BD FACS Celesta (Extended Data Fig. 19a).

RNA-FISH

TRIM28-FKBP iPSCs were seeded onto 0.2% gelatin-coated plates prior to treatment. For the treatment, 500nM of dTAG-13 and/or 2µg/ml Doxycycline was mixed with the media and incubated for the indicated durations. RNA-FISH was performed on glass coverslips coated with Poly-L-Lysine (Sigma-Aldrich) prior to seeding. Cells were dissociated with Accutase enzyme mix (Sigma-Aldrich), washed in 1X PBS, resuspended in minimal volume of ESC media. 10-20µL drops of cells were added to the coverslips and were allowed to attach for 10 minutes at room temperature. 4% PFA fixation for 15 minutes was followed by two 1X PBS washes and permeabilized in 70% ethanol at 4ºC overnight. RNA-FISH was performed with reagents from the Stellaris RNA-FISH method (LGC Technologies) following manufacturer's instructions. Hybridization was performed with an oligo probe that was labeled with Cy5 dye at the 5'end (TTCTTGATGTCCTAACCCTTTTCCTTC). After the final wash, the coverslips were mounted onto glass slides with ProLong Gold Anti-fade (Invitrogen) mounting media. Images were acquired with Z1 Observer (Zeiss) microscope at 100X magnification. Images were processed with ZEN 3.1 (Zeiss) and ImageJ software (Fig. 4c-d).

RNA-FISH combined with Immunofluorescence (IF)

RNA-FISH combined with Immunofluorescence was performed essentially as described [31]. For immunofluorescence, dTAG-13 or DMSO treated cells were fixed in 4% PFA for 10 minutes at RT and stored in PBS at 4°C. All buffers and antibodies were diluted in RNase-free PBS (Thermo Fisher, AM9624). Cells were permeabilized with 0.5% Triton X-100 (Thermo Fisher, 85111) for 10 min at RT, followed by three consecutive 5 min PBS washes. Cells were then incubated in the primary antibody (RNAPII (abcam, ab817) at 1:500, NFY-A (Santa Cruz, sc-17753 X) at 1:250, NRF1 (abcam, ab55744) at 1:500, MED1 (abcam, ab64965) at 1:500 and MED23 (Bethyl Labs, A300-425A) in PBS overnight. After two 5 min PBS washes, cells were incubated in the secondary antibody (Invitrogen, goat anti-mouse Alexa 488 (A-11001) or goat anti-rabbit Alexa 488 (A-11008)) at 1:500 in PBS for 60 min at RT. Cells were washed twice in PBS for 5 min and re-fixed with 4% PFA in PBS for 10 min at RT. Following two 5 min PBS washes, cells were washed once

with 20% Stellaris RNA FISH Wash Buffer A (Biosearch Technologies, Inc., SMF-WA1-60), 10% Deionized Formamide (EMD Millipore, S4117) in RNase-free water (Invitrogen, 10977035) for 5 min at RT. Cells were hybridized with 90% Stellaris RNA FISH Hybridization Buffer (Biosearch Technologies, SMF-HB1-10), 10% Deionized Formamide and 12.5 or 25 µM Stellaris RNA FISH probes. Probes were hybridized in a humidified chamber O/N at 37°C. Cells were washed with Wash Buffer A for 30 min at 37°C and stained with 0.24µg/mL DAPI in Wash Buffer A for 3 min at RT. Cell were washed with Stellaris RNA FISH Wash Buffer B (Biosearch Technologies, SMF-WB1-20) for 5 min at RT, mounted onto glass microscopy slides with Vectashield mounting medium (Vector Laboratories, H-1900) and sealed using transparent nail polish. Images were acquired with LSM880 Airyscan microscope equipped with a Plan-Apochromat-63x/1.40 oil DIC objective or Z1 Observer (Zeiss) microscope with 100X magnification. Images were processed with ZEN 3.1 (Zeiss) and ImageJ software (Fig. 1i, 2a-c, 2f-g, 3c, 4f, Extended Data Fig. 6a, 7a-b, 7d, 8b, 11g). RNA FISH probes were designed and generated by Biosearch Technologies Stellaris RNA FISH to target introns of *MiR290-295* primary transcript and *Cthrc1*, and *IAPez* transcripts. Sequences of RNA-FISH probes are available in the Supplementary Table 1a.

TrueSeq Stranded mRNA-seq

mESCs were cultured with either DMSO or 500 nM dTAG-13 for 2, 6, 24 and 96 hours. RNA was isolated using RNeasy kit (Qiagen) and 1µg of RNA was used for preparing the libraries. TrueSeq Stranded mRNA capture kit was used to prepare the libraries (KAPA biosystems) according to manufacturer's instructions. Unique Dual-Indexed (UDI; KAPA biosystems) adapters were ligated and the library was amplified for 8 cycles. The libraries were then sequenced as Paired-end 100 (PE100) on a Novaseq6000 with 50 million fragments per library.

TT-SLAM-Seq

TT-SLAM-Seq was performed as described previously [27]. Briefly, cells were treated with DMSO or 500 nM dTAG-13 for 2, 6 or 24 hours and subjected to 15 minutes of 4-Thiouridine (4sU) labeling using 500 µM 4sU. Total RNA was extracted with Trizol (Ambion) and 24:1 chloroform:isoamylalcohol (Sigma), while using 0.1 mM DTT in isopropanol precipitation and ethanol washes. For each sample, 50 µg of total RNA was fragmented with Magnesium RNA

Fragmentation Module (NEB), and fragmentation buffer was removed from samples with ethanol precipitation in presence of 0.1 mM DTT. RNA was then resuspended in 350 µl RNAse free water, diluted in biotinylation buffer (200 mM HEPES pH 7.5, and 10 mM EDTA) and topped up with 5 µg MTS-Biotin (previously diluted to 50 µg/ml in dimethylformamide) to reach a final volume of 500 µl. Biotinylation reaction was incubated for 30 minutes at room temperature while keeping samples in rotation and protected from light. Unbound biotin was removed with Acid-Phenol:Chloroform extraction (124:24:1, Ambion) and isopropanol precipitation. Biotinylated RNA was resuspended in 100 µl RNAse-free water, denatured in 65 °C for 10 minutes and then cooled on ice for 5 minutes. The biotinylated RNA was captured with 100 µl µMACS streptavidin beads (Miltenyi) by incubating for 15 minutes in rotation while keeping samples protected from light. µMACS columns were equilibrated on magnetic stand with nucleic acid equilibration buffer and two times with biotinylation buffer (20 mM HEPES, 1 mM EDTA, pH 8). Beads were transferred to columns, washed three times with wash buffer (100 mM Tris-HCl pH7.5, 10 mM EDTA, 1 M NaCl and 0.1 % Tween 20) and labeled RNA was eluted two times with total 200 µl of 100 mM DTT. RNA was cleaned up with RNEasy Minelute columns (Qiagen) and eluted to RNAse-free water with 1 mM DTT. 4sU residues of RNA were alkylated with iodoacetamide treatment (10 mM iodoacetamide in 50 mM $NaPO_4$, pH 8, and 50 % DMSO) by incubating samples in 50 °C for 15 minutes, followed by quenching with 20 mM DTT. RNA samples were purified with ethanol precipitation and treated with Turbo DNAse (Invitrogen). Sequencing libraries were prepared with NEBNext Ultra II Directional RNA Library Prep Kit and NEBNext Multiplex Oligos (NEB), according to manufacturer's instructions, except using 8 minutes incubation time in fragmentation step.

H3K27Ac and H3K9me3 Chromatin immunoprecipitation -sequencing (ChIP-Seq)

For ChIP-Seq experiments, DMSO and dTAG-13 treated cells were detached with TrypLE Express (Gibco), washed once with PBS, fixed in rotation with 1 % formaldehyde for 10 minutes in room temperature followed by 5 minutes of quenching with 125 mM glycine. For H3K27Ac and H3K9me3 ChIPs, three million mESCs were used per replicate sample, and 750,000 S2 cells were added for exogenous genome spike-in normalization [55]. Cells were lysed in LB1 (50 mM HEPES-KOH, 140 nM NaCl, 1 mM EDTA, 10 % glycerol, 0.5 % Igepal CA-630 and 0.25 % Triton X-100, 5 mM Na-butyrate and 1x protease inhibitor cocktail) and collected by

centrifugation. Lysis was continued in LB2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 5 mM Na-butarate and 1x protease inhibitor cocktail) followed by centrifugation. Nuclei were lysed in LB3 (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1 % Na-deoxycholate, 0.5 % N-Lauroylsarcosine, 5 mM Na-butyrate and 1x protease inhibitor cocktail) and chromatin was fragmented with Bioruptor NextGen for 35 cycles (high setting). Lysates were clarified and 10 % of the sample was set aside as input. The remaining sample was split in two to capture protein-DNA complexes with 1µg H3K27Ac (ab4729; Abcam) and 1µg H3K9me3 (ab8898; Abcam) antibodies by incubating them in rotation overnight in 4 °C. This was followed by 24 hours of incubation with Protein A Dynabeads (Invitrogen) that had been washed three times with 0.25 % BSA in PBS. Beads from immunoprecipitation were washed 7 times with RIPA buffer (50 mM HEPES-KOH pH7.5, 1 mM EDTA, 1 % Igepal CA-630, 0.7 % Na-deoxycholate, 500 mM LiCl, 5 mM Na-butyrate and 1 x protease inhibitor cocktail), once with TE buffer (10 mM Tris-HCl pH8.0, 1 mM EDTA and 50 mM NaCl) and eluted from beads with Elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA and 1% SDS). Samples were decrosslinked for 16 hours at 65 °C in presence of 550 mM NaCl and proteinase K (Ambion), treated with RNAse A (Thermo Scientific) and DNA was extracted with Phenol:Chloroform:Isoamylalcohol followed by chloroform extraction and ethanol precipitation. Sequencing libraries were prepared from 10 ng of DNA with KAPA HyperPrep Kit (Roche) and paired-end sequenced with NovaSeq 6000 (Illumina) to produce ~100 million fragments for each library.

RNAPII and MED23 ChIP-Seq

Cells were treated with DMSO or 500 nM dTAG-13 for 24 hours. RNAPII and MED23 ChIP-Seq samples were prepared as described above, except for MED23 samples, protein-protein crosslinking was performed by incubating cells in rotation with 2 mM disuccimidyl glutarate (Thermo Scientific; Pierce) in PBS, followed by two washes with PBS and formaldehyde fixation as above. 30 million and 10 million cells were used for RNAPII and MED23 IPs, respectively. RNAPII samples included an exogenous genome spike-in of 7.5 million S2 cells. Cell lysis was performed as above and samples were sonicated in sonication buffer (50 mM HEPES-KOH pH 7.5, 140 mM EDTA, 1 mM EDTA, 1 mM EGTA, 1 % Triton X-100, 0.1 % Na-deoxycholate, 0.1 % SDS and 1x protease inhibitors (Roche)) for 35 cycles in case of RNAPII samples and 50 cycles for MED23 samples by using Bioruptor NextGen (high setting). When preparing beads for IPs, 50

µl of Protein A Dynabeads (Invitrogen) was washed three times in blocking buffer (0.5 % BSA in PBS) and then bound to 5 µg of RNAPII antibody (8WG16; Biolegend) and 2 µg of Spike-in antibody (61686, Active Motif) or 5 µg of MED23 antibody (A300-425A; Bethyl Laboratories) during an overnight incubation in blocking buffer. Antibody-conjugated beads were washed three times with blocking buffer, resuspended to 100 µl of blocking buffer prior to adding them to sheared chromatin in a total volume of 2.5 ml sonication buffer for RNAPII and 1 ml for Med23. Immunoprecipitation was performed in rotation overnight in 4 °C. Beads were washed twice with sonication buffer, once with sonication buffer containing 500 nM NaCl, once with LiCl wash buffer (20 mM Tris pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5 % Igepal CA-630, 0.5 % Na-deoxycholate, 1x protease inhibitors) and once with TE buffer (10 mM Tris pH 8.0, 1 mM EDTA, 1x protease inhibitors). Samples were eluted from beads with 200 µl elution buffer as above and crosslinks were reversed by incubating samples in 65 °C for 16 hours. Samples were topped up with 200 µl TE buffer and 8 µl of RNAse A, incubated in 37 °C for 2 hours followed by addition of 7 µl of 300 mM $CaCl_2$, 2 µl Proteinase K (Ambion) and 30 minutes incubation in 37 °C. DNA purified from samples with phenol:chloroform:isoamyl alcohol extraction followed by ethanol precipitation. Libraries were prepared and sequenced as above for total ~50 million fragments for each library.

In-situ Hi-C

In-situ Hi-C experiments were performed for two biological replicates of TRIM28-FKBP mESCs after 24h DMSO or dTAG-13 treatments as described previously [56], with minor modifications. Briefly, 3 million cells per sample were fixed in suspension with 1 % formaldehyde for 10 minutes and quenched with 0.125 M glycine for 5 minutes in rotation. Cells were washed with PBS and lysed in Hi-C lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5% Igepal CA-630, 1.15 % Triton X-100 and 1x protease inhibitors (Roche)) for 10 minutes on ice. Nuclei were washed with cold PBS, NEBuffer 2 (NEB) and resuspended in 50 µl of 0.5 % SDS, incubated in 62 °C for 7 minutes and SDS was quenched with 145 µl water and 25 µl 10 % Triton X-100 followed by 15 min incubation in 37 °C. For restriction enzyme digestion, 25 µl of 10X NEBuffer2 and 250 U of MboI (NEB) were added and samples were incubated overnight in 37 °C, followed by additional 1 hour incubation after adding another 250 U of MboI. Enzymes were inactivated for 20 minutes in 65 °C. Samples were spinned in 5000 g, 200 µl of supernatant was removed and

sample was resuspended in 200 µl of 1.2X NEBuffer 2. To fill-in DNA overhangs, 50 µl of fill-in buffer (37.5 µl of 0.4 mM biotin-14-dATP (Thermo Fisher, 19524016), 1.5 µl of dCTP, 1.5 µl of dGTP, 1.5 µl of dTTP, 10 mM each, and 8 µl of 5U/µl DNA PolI Large (Klenow) Fragment (NEB, M0210L)) was added and the samples were incubated in 37 °C for 90 minutes. For ligation reaction, samples were topped up with 663 µl of water, 120 µl of NEB T4 DNA ligase buffer (NEB), 100 µl of 10 % Triton X-100 and 12 µl of 10 mg/ml BSA (NEB, B9000S) and 2000 U of T4 DNA ligase (NEB, M0202) and incubated overnight in 18 °C in rotation. Crosslinks were reversed with Proteinase K and 0.88 % SDS in 10 mM Tris-HCl (pH 7.5) during 30 min in 55 °C, followed by incubation in 68 °C in presence of 2.3 M NaCl for 2 hours. For library preparation, DNA was extracted from samples with ethanol and sodium acetate precipitation, sheared with Covaris E220 and pulled down with 150 µl of T1 Streptavidin Dynabeads (Thermo Fisher). The end-repair for captured DNA was performed in room temperature for 30 minutes using 100 µl reaction with 0.5 U of T4 Polynucleotide Kinase (NEB), 12 U of T4 DNA polymerase (NEB) 5 U of DNA pol I Large (Klenow) Fragment in NEB T4 DNA ligase buffer with 10 mM ATP and 0.5 mM dNTP. After washing the bead-bound DNA, samples were resuspended to NEBuffer 2 with 0.5 mM dATP and 25 U of Klenow exo minus (NEB, M0212) and incubated in 37 °C for 30 minutes. After washing the beads, universal sequencing adapters from NEBnext Ultra DNA Library Kit (NEB) were ligated in NEB Quick ligation reaction buffer according to manufacturer's guidelines. Libraries were amplified for 6 PCR cycles and purified using Ampure XP beads (Beckman Coulter). Libraries were paired-end sequenced with NovaSeq 6000 for 400 million fragments for each library.

Generation of DNA constructs for protein purification

For the purification of fluorescently labeled fusion proteins we amplified the C-terminal domain of POLR2A (Fwd: 5'-TACTCACCGACATCACCAGCCTATG-3', Rev: 5'-GTTCTCCTCATCGCTGTCA-3'), full length CBX5 (HP1α) (Fwd: 5'-ATATGGGAAAGAAGACCAAGAGGACAG-3', Rev: 5'-GCTCTTCGCGCTTTCTTT -3'), the intrinsically disordered region of MED1 (Fwd: 5'-ATTGCAGAGACAGTTGAAGACATGG-3', Rev: 5'-ATTGCCAATCAGGGCCACA-3') and the intrinsically disordered region of NFYC (Fwd: 5'-CCTGTCCAGTACTACTTCACG-3', Rev: 5'-GTCTCCAGTCACCTGGGG-3') from cDNA generated from v6.5 mouse embryonic stem cells. The amplified fragments were cloned

into pET45-mCherry (Addgene Cat# 145279) or pET45-mEGFP (this paper) backbones by Gibson assembly using NEBuilder® HiFi DNA Assembly Master Mix (NEB). All sequences of interest were cloned C-terminally to the respective fluorescence marker.

Protein purification

Protein overexpression in BL21 (DE3) (NEB M0491S) was performed as described [57]. Pellets were resuspended in 25 mL of ice-cold Buffer A (50 mM Tris pH 7.5, 500 mM NaCl, 20 mM Imidazole) supplemented with cOmplete protease inhibitors (Sigma, 11697498001) and sonicated for 10 cycles (15 s ON, 45 s OFF) on a Qsonica Q700 sonicator. The resulting lysate was cleared by centrifugation at 15,500 g for 30 minutes at 4°C. For protein purification we used the Äkta avant 25 chromatography system. The supernatant was loaded onto a cOmplete His-Tag purification column (Merck, 6781543001) pre-equilibrated in Buffer A. The loaded column was washed with 15 column volumes (CV) of Buffer A. Bound fusion protein was eluted in 10 CV of Elution Buffer (50 mM Tris pH 7.5, 500 mM NaCl, 250 mM Imidazole) and immediately diluted 1:1 in Storage Buffer (50 mM Tris pH 7.5, 125 mM NaCl, 1 mM DTT, 10% Glycerol). The resulting eluate was concentrated by centrifugation at 10,000 g for 30 minutes at 4°C using 3000 MWCO Amicon Ultra centrifugal filters (Merck, UFC803024). The concentrated fraction was diluted 1:100 in Storage Buffer, re-concentrated and stored at -80°C.

In vitro transcription of IAP RNA

To generate *in vitro* transcribed RNA for the *gag* region of IAPez, template PCR amplicon was generated from cDNA using primers listed in Supplementary Table 1b. The regions of interest cloned for in vitro transcription of the Mir290-295 super-enhancer RNA and Major Satellite Repeat RNA were previously described [35,38], and the primers used are listed in Supplementary Table 1b. 1µg of DNA was used for *in vitro* transcription using Fluorescein 12-UTP or Cy5-UTP RNA labeling kit (Jena Biosciences) and incubated at 37ºC for 4 hours followed by 15mins incubation with TURBO DNaseI. Labeled RNA was purified using Clean and Concentrator Kit-5 (Zymo research) and eluted in Tris-HCl pH8.0. For *Mir290-295* super-enhancer RNA, transcribed RNA of positive and negative strand was mixed 1:1 at equimolar concentrations.

In vitro droplet assay

For the *in vitro* droplet formation assay (Fig. 3d-i, Extended Data Fig. 12, Extended Data Fig. 13), the purified mCherry- or mEGFP-fusion proteins were measured for concentration and subsequently diluted to the desired concentration in Storage Buffer (50 mM Tris pH 7.5, 125 mM NaCl, 1 mM DTT, 10% Glycerol). Protein solutions were mixed with Fluorescein- or Cy5-labeled RNA and then diluted 1:1 with 20% PEG-8000 in de-ionized water (w/v). After 30 minutes of incubation at room temperature, 10 µl of this mix was pipetted onto a chambered coverslip (Ibidi, 80826-90). Images were acquired using an LSM880 confocal microscope equipped with a Plan-Apochromat-63x/1.40 oil DIC objective with a 2.5x post-magnification if indicated. Data for the quantification of condensate formation in *in vitro* mixing assays was acquired from at least 5 images of two independent image series per condition. For the detection of droplet regions, we used the ZEN blue 3.1 Image Analysis and Intellesis software packages. Image segmentation was achieved by use of a previously trained ZEN Intellesis algorithm for classification of each individual pixel into foreground (droplet area) or background (image background). Generated probability maps with a minimal confidence of at least 90%, a minimum area of 3 pixels and watershed for primary objects were implemented into the ZEN Image Analysis module to classify regions of interest. For the calculation of the partition ratio, the mean intensity values of droplets and background for the respective channel was used. Enrichment of RNA in droplets was calculated by background subtraction from the mean intensity values of droplet areas in the respective RNA channel (Fig. 3f, 3i, Extended Data Fig. 12e) or as partition ratio as described above. All figures were generated using R-Studio.

Fluorescence recovery after photobleaching (FRAP)

MED1 IDR droplets for photobleaching experiments were formed as described above at a concentration of 10µM (Extended Data Fig. 12f). Instead of a preassembly incubation for 30 minutes at room temperature, the droplets were immediately imaged after they were pipetted on the slide. Droplets were bleached with 488 nm light, and fluorescence recovery was imaged every second over a total of 60 seconds. Bleaching was performed on a central region of a settled single droplet using 10 iterations of 100% laser power. The resulting recovery curves were fitted to a power law model. All figures were generated using GraphPad PRISM9.

Generating wild type and mutant mouse embryos

Zygotes were generated by *in vitro* fertilization (IVF) as previously described [58]. Briefly, B6D2F1 female mice of 6-8 weeks of age were superovulated with two rounds of hormone injections (5IU of PMSG followed by 5IU of HCG after 46 hours). Oocytes were isolated and cultured in pre-gassed KSOM prior to IVF. F1 (C57BL/6J x Castaneous) sperm isolated from the cauda epididymis was thawed and used for IVF. 6 hours after fertilization, zygotes were washed in M2 media for multiple rounds and then prepared for electroporation. Alt-R CRISPR-Cas9 and guide RNAs Ribonucleoproteins (RNP) were prepared as previously described [42]. Guide RNAs used to target the genes are listed in Supplementary Table 1b. Zygotes were washed in 3 drops of OptiMEM Reduced Serum Media (Thermo Fisher Scientific) prior to electroporation. NEPA21 electroporator (NEPAgene) was used for electroporating zygotes with the following settings for a small chamber: Four poring pulses of 34 V for 2.5ms with an interval of 50ms was used to generate pores in the Zona pellucida layer. Voltage decay was set at 10% and (+) polarity. To enable intake of the RNPs, Five Transfer pulses of 5V was applied for 50ms with an interval of 50ms. Voltage decay for the transfer was set at 40% with an alternating polarity of (+) and (-). Electroporated zygotes were washed in 3 drops of KSOM medium and cultured in pre-gassed KSOM drops until blastocyst stage under standard embryo culture conditions. Blastocysts were scored for viability and morphology, and re-transferred bilaterally in a clutch of 15 blastocysts per uterine horn into day 2.5 pseudopregnant CD-1 surrogate female mice. E6.5 embryos were dissected from the uterus in 1X HBSS and used for further analysis. E5.5 wild type embryos were generated with the setup and mock electroporation with guide targeting GFP sequence was used.

All procedures have been performed in our specialized facility, following all relevant animal welfare guidelines and regulations, approved by the Max Planck Institute for Molecular Genetics and the local authorities LAGeSo, Berlin (license number: G0247/18-SGr1).

Single-cell RNA sequencing of embryos

E5.5 wild type and E6.5 TRIM28 mutant embryos were dissected from the decidua in 1X HBSS and then washed in 1X PBS. Reichert's membrane was carefully removed with sharp forceps and glass capillaries and the embryos were washed in 1X PBS with 0.4% BSA. The embryos were disaggregated with TrypLE Express (Gibco) with gentle pipetting every 10 minutes up to a total of 40 minutes at 37ºC. The dissociated cells were counted for viability and then washed in 1X PBS

with 0.4% BSA for a total of three washes at 4ºC and 1,200rpm for 5 minutes. The cells were subjected to single-cell RNA sequencing using 10X Genomics Chromium Single Cell 3' v2 kit. Single-cell libraries were generated following manufacturer's instructions with an exception to the cycle number used. Libraries were sequenced on a Novaseq6000 with asymmetric reads and a depth of 300-350 million fragments per library.

Average image and radial distribution analysis

Image analysis pipeline used for the colocalization analysis of RNA FISH combined with IF was described previously [31]. Briefly, MATLAB scripts were used to identify RNA FISH foci in z stacks through intensity thresholding (same threshold was used for image sets shown on the same figure panels) and create RNA FISH signal centroids (*x, y, z*) that were stitched together and positioned in a box of size $l = 1.5\ \mu m$. For identified FISH foci, signal from corresponding location in the IF channel was collected in the *l x l* square centered at the RNA FISH focus at every corresponding z-slice. The IF signal centered at FISH foci for each FISH and IF pair were then combined to calculate an average intensity projection, providing averaged data for IF signal intensity within a *l x l* square centered at FISH foci. The same process was carried out for the FISH signal intensity centered on its own coordinates, providing averaged data for FISH signal intensity within a *l x l* square centered at FISH foci. As a control, this same process was carried out for IF signal centered at random positions. These average intensity projections were then used to generate 2D contour maps of the signal intensity or radial distribution plots. Contour plots are generated using in-built functions in MATLAB. The intensity radial function *((r))* is computed from the average data. For the contour plots of the IF channel, an intensity colormap consisting of 14 bins with gradients of black, violet and green was generated. For the FISH channel, black to magenta was used. The generated colormap was employed to 14 evenly spaced intensity bins for all IF plots. The averaged IF centered at FISH or at randomly selected nuclear locations were plotted using the same color scale. For the radial distribution plots, the Spearman correlation coefficients r were computed and reported between the FISH and IF (centered at FISH) signal. A two-tailed student's t-test, comparing the Spearman correlation calculated for all pairs, was used to generate p values (Fig. 1i, 2f-g, 3c, 4f, Extended Data Fig. 6a).

Bioinformatics

All analyses were carried out using R 3.6.3 if not stated otherwise.

RNA-Seq processing

Raw reads were subjected to adapter and quality trimming with cutadapt [59] (version 2.4; parameters: --nextseq-trim 20 --overlap 5 --minimum-length 25 --adapter AGATCGGAAGAGC -A AGATCGGAAGAGC), followed by poly-A trimming with cutadapt (parameters: --overlap 20 --minimum-length 25 --adapter "A[100]" --adapter "T[100]"). Reads were aligned to the mouse reference (mm10) using STAR [60] (version 2.7.5a; parameters: --runMode alignReads --chimSegmentMin 20 --outSAMstrandField intronMotif --quantMode GeneCounts) and transcripts were assembled using StringTie [61] (version 2.0.6; parameters: -e) with GENCODE annotation (VM19). For the repeat expression quantification, reads were re-aligned with additional parameters '--outFilterMultimapNmax 50'.

TT-SLAM-Seq processing

Raw reads were trimmed by quality, Illumina adapter content and poly-A content analogous to the RNAseq samples and aligned with STAR with parameters '--outFilterMultimapNmax 50 --outReadsUnmapped Fastx' to the SILVA database [62] (download: March 6, 2020) in order to remove rRNA content. Unaligned reads were afterwards reverse-complemented using the seqtk 'seq' command (https://github.com/lh3/seqtk, version 1.3-r106; parameters: -r). Reverse-complemented reads were processed using SLAM-DUNK [63] with the 'all' command (version 0.4.1; parameters: -rl 100 -5 0) with the GENCODE gene annotation (VM19) as '-b' option. Reads with a 'T>C' conversion representing nascent transcription were filtered from the BAM files using alleyoop (provided together with SLAM-DUNK) with the 'read-separator' command. Counts per gene were quantified based on the 'T>C'-converted reads using htseq-count (version 0.11.4; parameters: --stranded=yes, --nonunique=all) [64]. FPKM values were calculated based on the resulting counts. For genome-wide coverage tracks, technical replicates were merged using samtools 'merge' [65]. Coverage tracks for single and merged replicates were obtained using deepTools bamCoverage [65] (version: 3.4.3; parameters: --normalizeUsing CPM) separately for the forward and reverse strand based on the 'T>C'-converted reads.

Public ChIP-Seq data

Fastq files of public ChIP-Seq data for H3K27Ac [66], H3K9me3 [21], HP1α [22], TRIM28 [21], NANOG [40], OCT4 [40] and SOX2 [40] were downloaded from GEO (Supplementary Table 1c).

ChIP-Seq processing

Raw reads of treatment and input samples were subjected to adapter and quality trimming with cutadapt (version 2.4; parameters: --nextseq-trim 20 --overlap 5 --minimum-length 25 --adapter AGATCGGAAGAGC -A AGATCGGAAGAGC). Reads were aligned separately to the mouse genome (mm10) and to the fly genome (*D. Melanogaster*, dm6) using BWA [67] with the 'mem' command (version 0.7.17, default parameters). A sorted BAM file was obtained and indexed using samtools with the 'sort' and 'index' commands (version 1.10). Duplicate reads were identified and removed using GATK [68] (version 4.1.4.1) with the 'MarkDuplicates' command and default parameters. Technical replicates of treatment and input samples were merged respectively using samtools 'merge'. Peaks were called with reads aligning to the mouse genome only using MACS2 [69] 'callpeak' (version 2.1.2; parameters --bdg --SPMR) using the input samples as control samples. For H3K9me3 only, the '--broad' option was used. Genome-wide coverage tracks for single and merged replicates normalized by library size were computed using deepTools bamCoverage (parameters: --normalizeUsing CPM --extendReads) and in addition normalized by the spike-in factor obtained from the reads aligning to the *Drosophila* genome as described [55].

Publicly available paired-end ChIP-Seq data (HP1α) were processed as described above except spike-in and input normalization.

Publicly available single-end ChIP-Seq data (H3K9me3, TRIM28, H3K27ac, NANOG, OCT4, SOX2) were trimmed using cutadapt (version 2.4; parameters: --nextseq-trim 20 --overlap 5 --minimum-length 25 --adapter AGATCGGAAGAGC), and aligned and de-duplicated analogous to the paired-end data. In order to estimate the fragment size, only reads with a mapping quality of at least 15 were considered and used as input for spp [70] (version 1.2.2). All reads were then used to create coverage tracks using bamCoverage and to call peaks with MACS2 'callpeak' (version 2.1.2; parameters --bdg --SPMR --nomodel) with '--extsize' set to the estimated fragment length divided by two and input samples used as control for H3K27ac, NANOG, OCT4 and SOX2. For H3K9me3 only, the '--broad' and '--nolambda' options were used.

Enhancer and super-enhancer annotation

The annotation of super-enhancers, enhancers and enhancer constituents was taken from Whyte et al. [71]. Coordinates were lifted from mm9 to mm10 using UCSC liftOver. These coordinates were used throughout this study for all enhancer-associated analyses (Supplementary Table 2).

Detection of eRNA expression and preparation of TT-SLAM-Seq histograms

TT-SLAM-Seq signal was quantified at typical enhancer and SE constituents. Enhancers within 3kb of TSS were excluded and intragenic enhancers were only considered on opposing strand relative to genes (GENCODE VM19, excluding "bidirectional_promoter_lncRNA") to measure eRNA without interference from gene transcription. Gene coordinates were extended for 10kb from transcription termination site to prevent overlap of transcriptional readthrough signal. The detection of eRNA was performed at 2 kb regions centered by the constituent coordinates. TT-SLAM-Seq counts were quantified from .bam files with htseq-count (parameters: --stranded = yes, --nonunique = all). Differential expression analysis was performed on enhancer elements with DEseq2 [72] (parameters: test="LRT", reduced=~1) and excluding regions with less than 10 counts across samples. For visualizing fold changes, regions with low eRNA expression (FPKM < 1) were excluded (Fig. 1h).

Quantifications for eRNA histograms were done with HOMER software [73] (version 4.10) and only intergenic enhancers were considered. Reads from .bam files were prepared for quantifications with 'makeTagDirectory', and histograms for enhancer and SE constituents were prepared with 'annotatePeaks.pl' (parameters: -size 4000, -hist 20) and metagene plots for SEs with 'makeMetaGeneProfile.pl' (parameters: -min 500, -size 4000). Histograms were smoothened by taking a rolling mean with window size 5 before plotting.

To visualize TT-SLAM-Seq signal across genes, protein-coding genes were considered and metagene plot was prepared using 'makeMetaGeneProfile.pl' after removing outliers (FPKM <0.5 or >100). Histograms were smoothened by taking a rolling mean with window size 3 before plotting (Extended Data Fig. 3c).

Differential gene expression analysis

For the differential gene expression analysis only protein coding genes were considered. Differential expression for RNAseq and TT-SLAM-Seq samples was measured using DESeq2 (parameters: test="LRT", reduced=~1) based on the raw counts per gene considering all time points per experiment type in one design. Only genes with at least 10 reads across all samples of the same experiment type were considered for the analysis. Genes with an absolute log2 fold change greater than 1 in comparison to the DMSO control and an adjusted p-value of less than 0.05 were termed differentially expressed per time point. Lowly expressed genes across all time points (average TPM across all RNAseq samples of less than 0.5 or average FPKM across all TT-SLAM-Seq samples of less than 0.25) were excluded from the analysis.

Gene set enrichment (GSEA) analysis

GSEA was performed to test if super-enhancer (SE) associated genes were overrepresented in the sets of TT-SLAM-Seq differentially regulated genes. All protein-coding genes were ranked based on log2 fold change. To prepare a gene set for the GSEA, Refseq IDs of SE-associated genes obtained from [71] were converted to (version free) Ensembl gene IDs using biomaRt R package [74]. GSEA was run with the fgsea R package [75] using 100000 permutations (Extended Data Fig. 5d).

Retrotransposon element definition

The genome-wide retrotransposon annotation of LTR, LINE and SINE elements was downloaded from Repbase [76]. Based on the Repbase classification system, we used the element annotation as LTR, LINE or SINE as the retrotransposon classes. Retrotransposon families considered in this study were L1 and L2 elements (LINE), ERV1, ERV3, ERVK, ERVL and MALR (LTR), as well as Alu, B2, B4 and MIR elements (SINE). Repeat subfamilies used in this study were subdivided into IAP, MMERVK and MMETn (ERVK) elements. IAPs and MMERVKs consist of multiple different subfamilies as annotated by Repbase (Extended Data Fig. 1a, 4a) which we summarized under these broader terms. The classification is consistent with retrotransposon classification described in previous studies [1,77,78].

Full length retrotransposons were defined based on the Repbase repeat annotation. For full length ERVK elements, we required the element to consist of an inner part with two flanking LTRs. First,

elements annotated as inner parts (containing the keyword 'int') were merged if they belonged to the same subfamily and were located within maximal 200 base pairs of each other. Second, only the merged inner parts with an annotated ERVK LTR within a distance of at most 50 base pairs on each side were selected as full length element candidates. For IAPs specifically only LTRs were considered that belonged to an IAP subfamily. No size restrictions were applied on the inner parts or LTRs which could lead to potential false positive candidates that are too truncated to be able to be transcribed but on the other hand provides an unbiased definition of full length repeat elements. The subfamily per element was defined based on the inner part.

Inner parts flanked by only one LTR were termed half length elements. LTRs without an inner part were termed solo LTRs.

In order to provide a broad overview of potential full length L1 elements only annotated elements with a size of more than 6 kb were shown.

The genomic co-ordinates of retrotransposons are listed in Supplementary Table 2a-e.

Mappability

The genome-wide mappability was calculated using GenMap [79] (parameters: -K 100 -E 0 -t -w -bg) with the mouse genome (mm10) as input (Extended Data Fig. 1c-g).

ChIP-Seq enrichment analysis

Bigwig tracks of ChIP-Seq signal were imported into R with the package rtracklayer [80]. Enriched heatmaps and metaprofile plots of ChIP-Seq signal were generated using the R package EnrichedHeatmap [81]. For this purpose, the signal was normalized to genomic features using the function 'normalizeToMatrix' (parameters: extend = c(2000, 2000), mean_mode = "w0", w = 50, target_ratio = 0.25). The resulting data matrix was visualized using the function 'EnrichedHeatmap'.

Hi-C data processing and analysis

Raw reads were trimmed for adapters and quality as described above. Reads were further trimmed at MboI restriction site using cutadapt (version 2.4; parameters: -a GATC, --minimum-length 25) and separately aligned to the mouse genome (mm10) using BWA [67] 'mem' (version 0.7.17, default parameters). Replicates were processed individually and then merged with samtools 'merge' [65]. Remainder of Hi-C data processing was performed with HOMER software [73] (version 4.10), unless specified otherwise. Paired-end reads were matched and prepared for analysis with 'makeTagDirectory' and PCR duplicates were removed (parameters: -tbp 1). Reads were filtered to remove continuous genomic fragments, re-ligation events, self-ligations and tags from regions with unusually high tag density while keeping only paired-end reads where both reads have a MboI restriction site within the fragment length estimate 3' to the read (parameters: -removePEbg -restrictionSite GATC -both -removeSelfLigation -removeSpikes 10000 5).

Analysis of compartments was performed with HOMER 'runHiCpca.pl' using resolution of 25 kb, 50 kb analysis window and using H3K27Ac peaks to assign active (A) and inactive (B) compartments. Hi-C matrices were prepared for visualization with 'analyzeHiC' by normalizing counts to interactions per hundred square kilobases per billion (default) or displaying log2 observed/expected counts normalized for interaction distance (parameters: -distNorm). Resolutions used were 150 kb for whole chromosome view (Extended Data Fig. 10a), 150kb resolution and 300 kb 'window' for region chr12:3,600,000-35,800,000 (Extended Data Fig. 10b, left) and in displaying differential (dTAG-13 - DMSO) of observed/expected matrix (Extended Data Fig. 10b, right). Images were prepared with Treeview 3.0

To perform pile-up analysis from ERV-gene and ERV-super-enhancer contacts, HOMER formatted Hi-C data was converted first to .hic format with 'tagDir2hicFile.pl' and then to .cool format with 'hic2cool convert' (version 0.8.3). Pile-up analysis was performed with 'coolpup.py' software [82] using Knight-Ruiz balanced matrices with 50 kb resolution and analyzing all cis interactions between ERVs and super-enhancers and between ERVs and protein coding genes, excluding genes with low expression (FPKM < 0.25 across TT-SLAM-Seq samples).

Motif enrichment

Enrichment of motifs was calculated using ame (version 5.3.0, default parameters) [83]. The sequences of 5' full length IAP LTRs and full length IAP inner parts were analyzed separately using the sequences of super-enhancer constituents as control (Extended Data Fig. 8a). In Extended Data Fig. 8a., the top TFs whose motifs show enrichment in the IAP LTRs or inner parts are shown. Also displayed is the expression level of the TFs calculated from the RNA-Seq data. For further functional tests, NFY was selected as its motif is highly enriched in IAPs, and it is expressed above 50 TPM in mESCs.

scRNA-seq processing

Fastq files for the wild-type time points E6.5 and E7.0 were downloaded from GEO (Supplementary Table 1c) [84]. For the wild-type time point E5.5 and the Trim28 KO, raw reads (fastq) were generated using Cell Ranger [https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest] (version 4) from 10x Genomics Inc. with the command 'cellranger mkfastq'. Reads from all time points were aligned against the mouse genome (mm10), and barcodes and UMIs were counted using 'cellranger count'. Multiple sequencing runs were combined using 'cellranger aggr'.

Genotyping

Genotyping was performed as described [42]. Briefly, samples were aligned against a hybrid mouse assembly (mm10 with positions of SNPs with the CAST/Eij strain masked as 'N' using SNPsplit [85] using STAR (parameters: --outSAMattributes NH HI NM MD). Reads were sorted by reference genome using SNPsplit. SNPs identified in [42] to be covered by reads from both genomic backgrounds were used to determine the allele composition for each cell. Individual cells were then assigned to embryos using the autosomal distribution of CAST/Eij SNPs. In order to determine embryo sex, marker genes *Xist*, *Erdr1*, *Ddx3y* and *Eif2s3y* were used to evaluate X and Y chromosome transcription based on the counts produced by Cell Ranger. Embryos were determined to be female if they had a high fraction of cells expressing *Xist*, while embryos with a high faction of cells expressing *Erdr1*, *Ddx3y* and *Eif2s3y* were determined to be male. Cells that could not confidently be assigned to an embryo were discarded from the analysis.

Wild-type reference

In order to define clusters of cells for the wild-type embryos, the R package Seurat [86] was used. For each time point (E5.5, E6.5 and E7.0), the 3000 most variable genes were detected across all cells and embryos. All time points were integrated using the functions 'FindIntegrationAnchors' and 'IntegrateData' (parameters: dims = 1:20). The combined wild-type time points were log2-normalized ('NormalizeData') and scaled while accounting for potential bias from cell cycle or mitochondrial gene counts using the function 'ScaleData' with parameters 'vars.to.regress = c('percent.mt', 'S.Score', 'G2M.Score')'. A UMAP was used to represent the cells in two dimensions using the function 'RunUMAP' (parameters: reduction = 'pca', dims = 1:20) based on the PCA computed by the function 'RunPCA' (parameters: npcs = 30). Six clusters of cells were identified using the functions 'FindNeighbors' (parameters: reduction = 'pca', dims = 1:20) and 'FindClusters' (parameters: resolution = 0.1). Marker genes per cluster were identified with the 'FindAllMarkers' function (parameters: only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25). Based on the detected marker genes per cluster and literature-based markers of cell states we assigned clusters to the following cell states: epiblast, primitive streak, extraembryonic ectoderm (ExE)/trophoblast, visceral endoderm, mesoderm and parietal endoderm (Supplementary Table 6). The heatmap of marker genes per cell state was generated based on the z-score transformed log2-normalized counts using the function 'DoHeatmap' (Extended Data Fig. 16e).

Knockout assignment

The E6.5 TRIM28 knockout expression data were log2-normalized analogous to the wild-type time points and transferred to the combined wild-type reference using the functions 'FindTransferAnchors' (parameters: dims = 1:30) and 'TransferData' (parameters: dims = 1:30) based on the cell states identified in the wild-type. Wild-type and knockout samples were integrated using 'FindIntegrationAnchors' and 'IntegrateData' (parameters: dims = 1:20, k.filter = 50). Analogous to the wild-type reference, the combined dataset was scaled and a UMAP was used in order to represent the data in two dimensions (same parameters for all functions as for the wild-type reference).

Cell state proportions

Per embryo, the proportion of a cell state was computed as the number of cells assigned to a cell state divided by the total number of cells assigned to the respective embryo. Per time point, cell

state proportions were represented by the median of proportions of the respective cell state across all embryos of a developmental time point (Fig. 5e, Extended Data Fig. 17a-b).

RNA velocity

RNA velocity was computed using velocyto [87], visualized using scanpy [88], and projected on the wild-type reference UMAP (Extended Data Fig. 16d).

Retrotransposon expression quantification

Global repeat expression quantification from RNAseq, TT-SLAM-Seq and scRNA-seq (Fig. 1e, Extended Data Fig. 4b-e) was carried out as described [42]. Briefly, in order to estimate the expression for each retrotransposon subfamily without bias due to gene expression, only reads not overlapping any gene were considered for the analysis. Reads overlapping splice sites as well as reads with a high poly-A content were removed. The remaining reads were counted per subfamily only if they aligned uniquely or multiple times to elements of the same subfamily. Here, any annotated element of a specific subfamily from Repbase was considered independent of our full length ERVK annotations. Reads aligning to multiple elements were only counted once. For scRNA-seq samples, reads were counted per subfamily, sample and cell state. The number of reads per subfamily were normalized by library size for RNAseq and TT-SLAM-Seq samples, and normalized by reads aligning to genes and repeats for scRNA-seq samples. Fold changes were calculated with respect to the DMSO or wild-type samples.

Pluripotency and differentiation marker quantification

For the pluripotency and differentiation marker genes, TPMs were z-score transformed and visualized using the R package and function pheatmap [Kolde (2019) pheatmap: Pretty Heatmaps. R package version 1.0.12. https://CRAN.R-project.org/package=pheatmap] (Extended Data Fig. 19b-c).

Statistical tests

The statistical significance of the difference of IAP expression between DMSO control and dTAG time points for TT-SLAM-Seq and RNAseq was calculated using an unpaired two-sided t-test (Fig. 1e). The significance for the difference in eRNA expression after 24h dTAG treatment (Fig. 1h)

was calculated with one-sample t-test. Statistical significance of differences in fold changes (vs. DMSO) in control vs. 1.6-hexanediol -treated cells was estimated with unpaired two-sided t-test (Extended Data Fig. 7e). All other tests are described in the figure legends.

Definition of boxplot elements

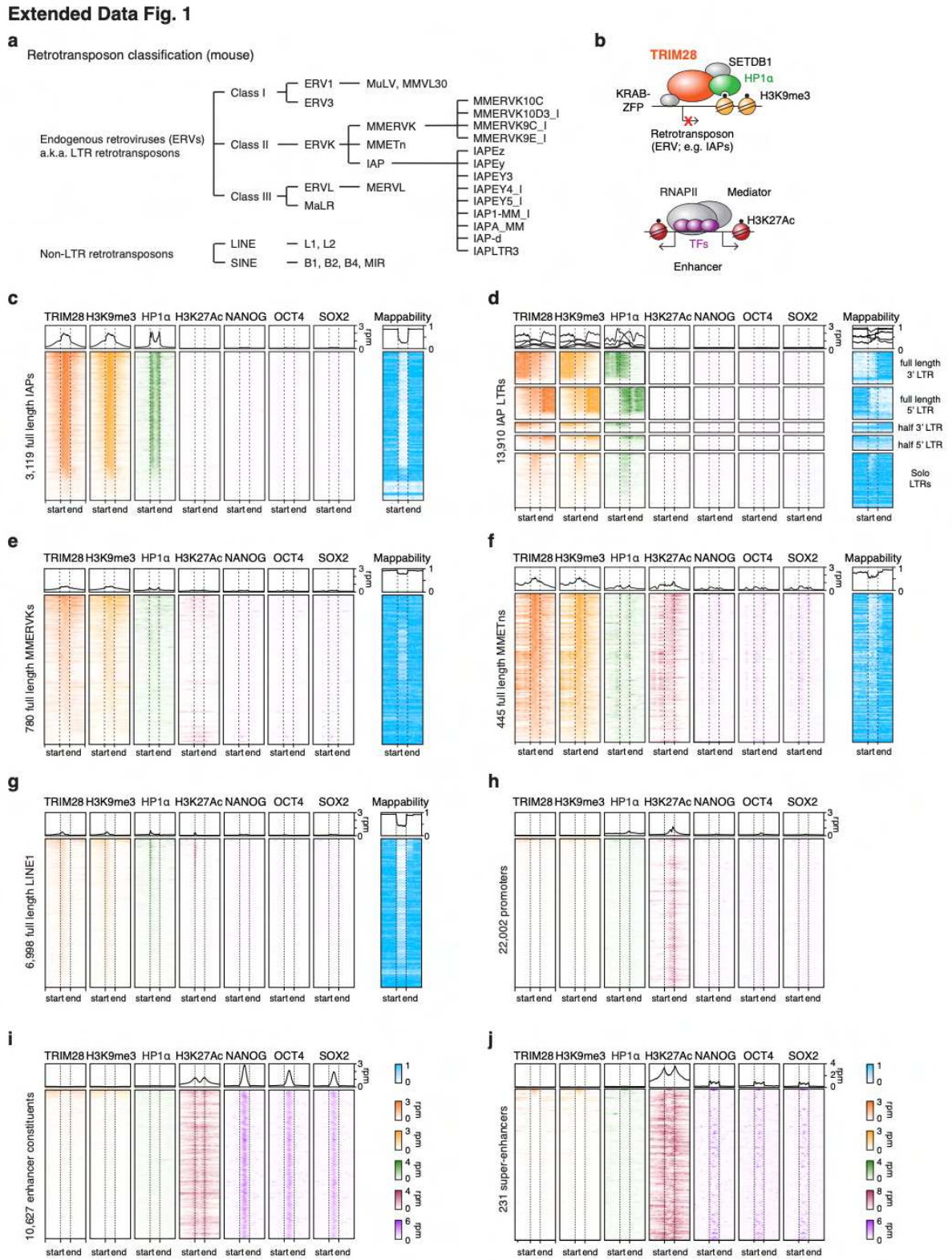
Fig. 1h: Elements depicted in boxplot: middle line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range.

Fig. 5d: Elements depicted in Scatter dot plot: middle line, mean; scatter limits, upper and lower range.

Fig. 3e, 3f, 3g, 3i, Extended Data Fig. 12c, 12e, 12j, 13b: Elements depicted in boxplots: middle line, median; box limits, upper and lower quartile; whiskers, 1.5x interquartile range.

Extended Data Fig 6b, 14c: Elements depicted in dot plots: middle line, mean; whiskers, standard deviation; points, all data points.

Extended Data Figures



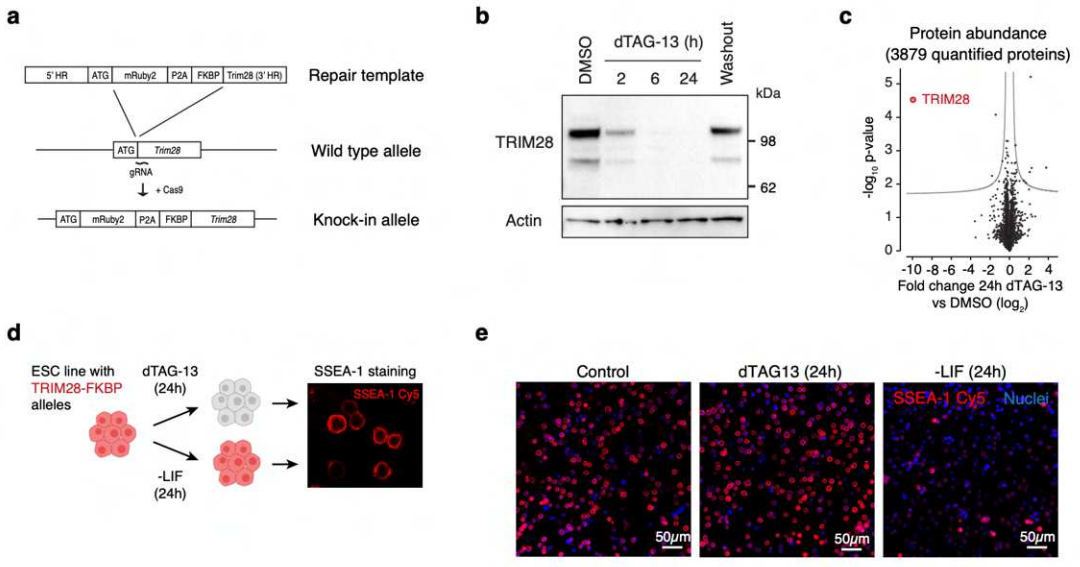
Extended Data Fig. 1. TRIM28 occupies ERVs but not enhancers and promoters in mESCs

a. Classes of retrotransposons in the mouse

b. (top) Model of TRIM28/HP1α pathway and its major regulators at endogenous retroviruses (ERVs). KRAB-ZFP: KRAB-Zinc Finger Protein, TRIM28: Tripartite motif -containing protein 28, HP1α: heterochromatin protein 1α. (bottom) Model of enhancers. Enhancers are bound by transcription factors (TFs) that recruit the Mediator complex and RNA Polymerase II (RNAPII).

c-j. Heatmap representation of ChIP-seq read densities of the indicated factors within a 2kb window around the indicated genomic features. The genomic features (the middle part of the plot) were length normalized. Meta-analyses of the mean binding profile of the indicated factors are displayed above the heatmaps. Enhancers in mESCs are bound by the OCT4, SOX2, NANOG TFs, and are devoid of components of the TRIM28/HP1α pathway that are bound at ERVs. For transposable elements, the mappability information is displayed on the right. Rpm: reads per million.

Extended Data Fig. 2



1264

1265 **Extended Data Fig. 2. Extended characterization of TRIM28-FKBP ESCs**

1266 **a.** Scheme of FKBP knock-in strategy at the *Trim28* locus.

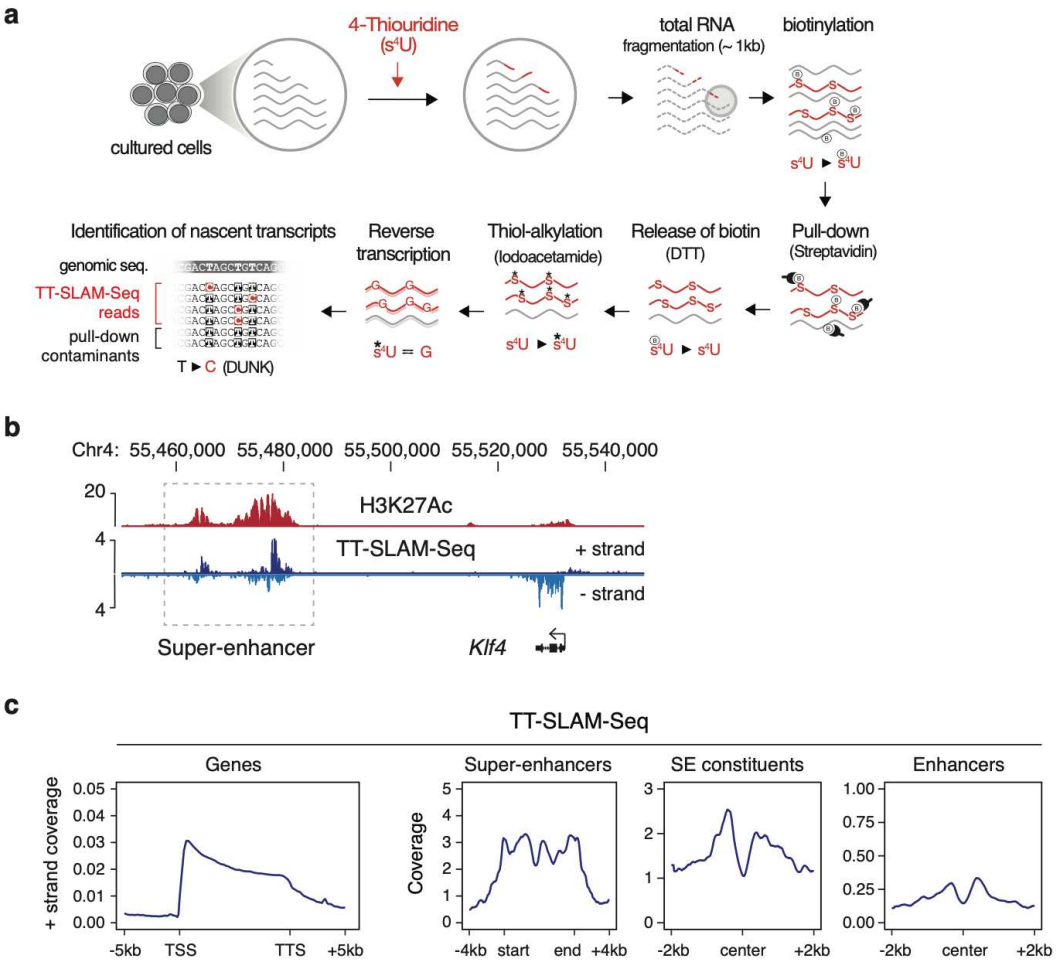
1267 **b.** Western blot validation of the FKBP degron tag and its ability to degrade TRIM28. Washout
1268 of the dTAG-13 ligand (24h) indicates reversibility of degradation. Actin is shown as the loading
1269 control.

1270 **c.** Mass spectrometry analysis of protein levels after 24h of dTAG-13 treatment. Every dot
1271 corresponds to a quantified protein. The degradation appears highly selective for TRIM28.

1272 **d.** Scheme of the SSEA-1 staining experiments. SSEA-1 is a marker of pluripotent cells.
1273 Withdrawal of LIF in the cell culture medium leads to differentiation, and is included as a
1274 control.

1275 **e.** 24h of dTAG-13 treatment does not substantially change the fraction of cells expressing the
1276 SSEA-1 pluripotency marker. Displayed are representative images of SSEA-1-Cy5
1277 immunofluorescence experiments. Scale bars: 50μm.

Extended Data Fig. 3

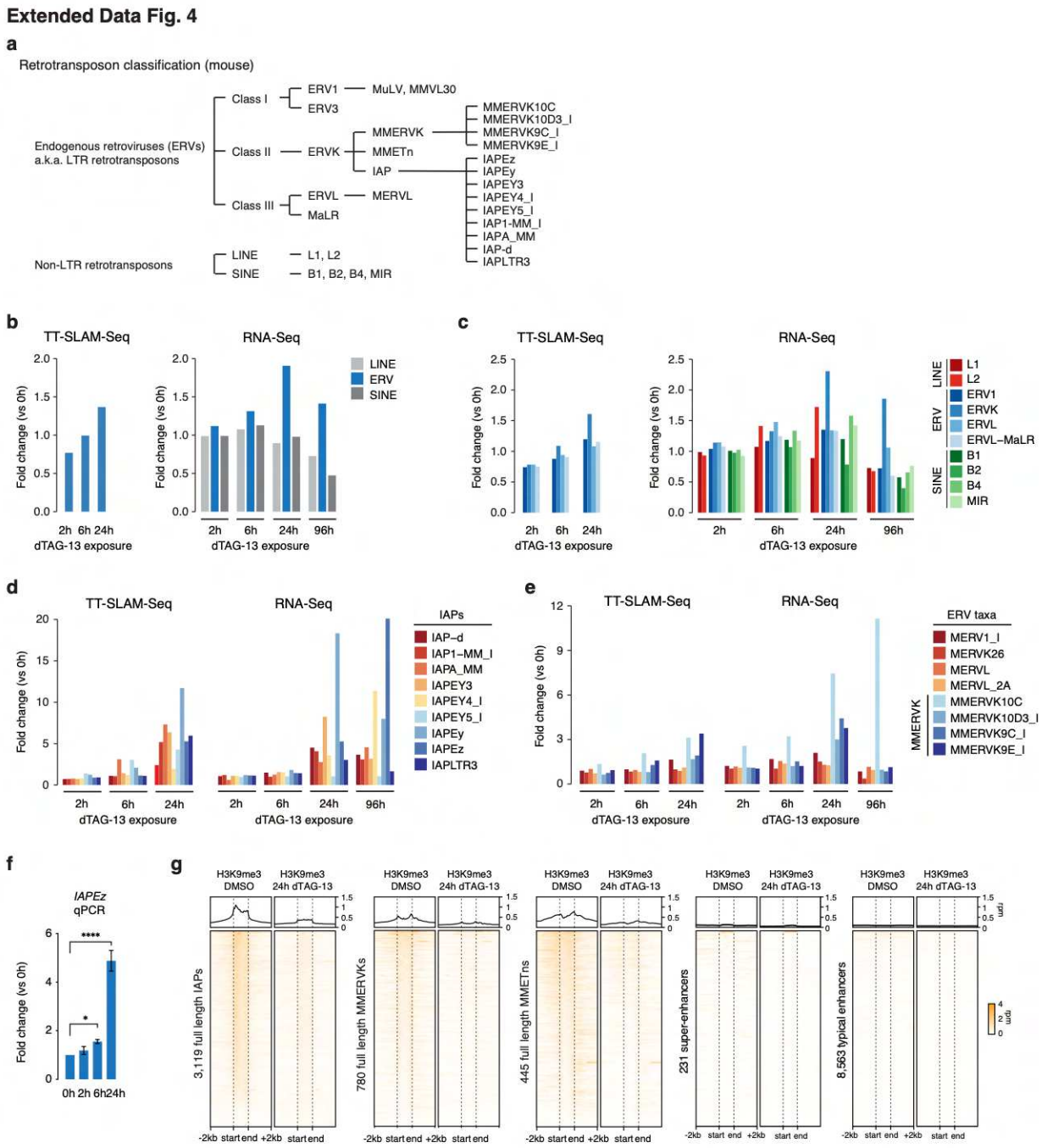


Extended Data Fig. 3. TT-SLAM-Seq detects nascent transcription

a. Schematic overview of the TT-SLAM-Seq experimental and analytical procedure.

b. TT-SLAM-Seq and H3K27Ac ChIP-Seq browser tracks at the *KLF4* super-enhancer locus. Rpm: reads per million. Co-ordinates are mm10 genome assembly co-ordinates.

c. TT-SLAM-Seq detects nascent transcription at genes, enhancers and super-enhancers. Displayed are meta representations of TT-SLAM-Seq read densities at genes, super-enhancers, constituent enhancers of super-enhancers, and enhancers in mESCs. Enhancers include all constituent enhancers of super-enhancers and typical enhancers. The genomic features at genes and super-enhancers were length normalized.



1288

1289 **Extended Data Fig. 4. Retrotransposon de-repression in TRIM28-degraded ESCs**

1290 **a.** Classes of retrotransposons in the mouse.

1291 **b.** Transcript levels of ERVs, LINEs and SINEs detected with TT-SLAM-Seq and RNA-Seq

1292 upon dTAG-13 treatment in mESCs. Values are normalized to the levels detected at 0h.

c. Transcript levels of ERV (sub)classes detected with TT-SLAM-Seq and RNA-Seq upon dTAG-13 treatment in mESCs. Values are normalized to the levels detected at 0h.

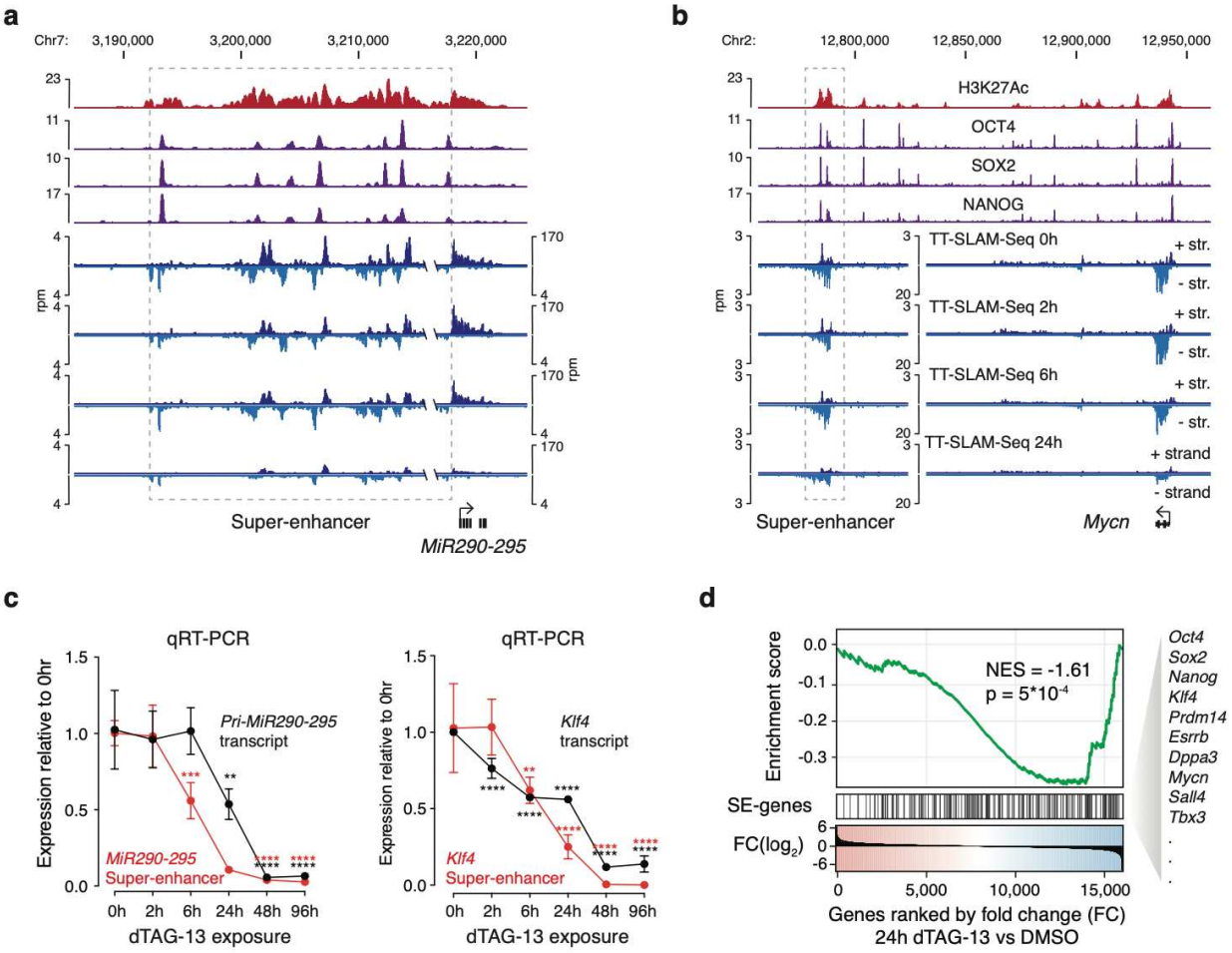
d. Transcript levels of IAP subfamilies detected with TT-SLAM-Seq and RNA-Seq upon dTAG-13 treatment in mESCs. Values are normalized to the levels detected at 0h.

e. Transcript levels of the indicated ERV taxa detected with TT-SLAM-Seq and RNA-Seq upon dTAG-13 treatment in mESCs. Values are normalized to the levels detected at 0h.

f. qRT-PCR analysis of IAPez RNA levels upon dTAG-13 treatment in mESCs. Values are normalized to the levels detected at 0h. ****: $P<10^{-4}$, *: $P<0.05$ (t test).

g. Reduced H3K9me3 at IAPs, MMERVKs and MMETns in dTAG-13 treated (24h) mESCs. Displayed are heatmap representations of H3K9me3 ChIP-seq read densities within a 2kb window around the indicated genomic features. The genomic features (the middle part of the plot) were length normalized. Meta-analyses of the mean binding profile of the indicated factors are displayed above the heatmaps. Rpm: reads per million.

Extended Data Fig. 5



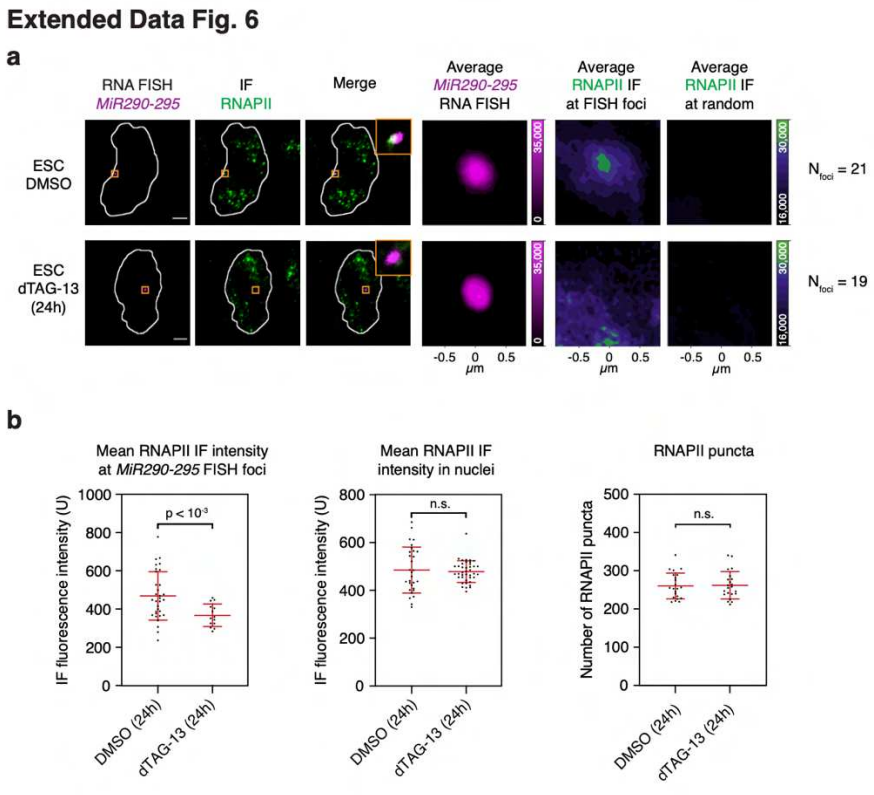
Extended Data Fig. 5. Reduction of super-enhancer transcription and the pluripotency circuit in TRIM28-degraded mESCs

a. Acute reduction of transcription at the *MiR290-295* super-enhancer locus upon TRIM28-degradation. Displayed are genome browser tracks of ChIP-Seq data (H3K27Ac, OCT4, SOX2, NANOG) in control mESCs, and TT-SLAM-Seq data upon 0h, 2h, 6h and 24h dTAG-13 treatment at the *MiR290-295* locus. Rpm: reads per million. Co-ordinates are mm10 genome assembly co-ordinates.

b. Acute reduction of transcription at the *Mycn* super-enhancer locus upon TRIM28-degradation. Displayed are genome browser tracks of ChIP-Seq data (H3K27Ac, OCT4, SOX2, NANOG) in control mESCs, and TT-SLAM-Seq data upon 0h, 2h, 6h and 24h dTAG-13 treatment at the *Mycn* locus. Rpm: reads per million. Co-ordinates are mm10 genome assembly co-ordinates.

c. qRT-PCR validation of the TT-SLAM-Seq data at the *MiR290-295* and *Klf4* loci. Displayed are transcript levels after the indicated duration of dTAG-13 treatment. Values are displayed as mean+/-SD from three independent experiments, and are normalized to the level at 0h. P values are from two-tailed t tests. ****: $P<10^{-4}$, ***: $P<10^{-3}$, **: $P<10^{-2}$, *: $P<0.05$

d. Genes downregulated after 24h dTAG-13 treatment are enriched for super-enhancer associated genes. Displayed is a Gene Set Enrichment Analysis (GSEA). Genes are ranked according to their fold change in transcription (TT-SLAM-Seq) after 24h dTAG-13 treatment. Genes associated a super-enhancer are marked with a black tick mark. Some of the genes implicated in the control of pluripotency are highlighted.

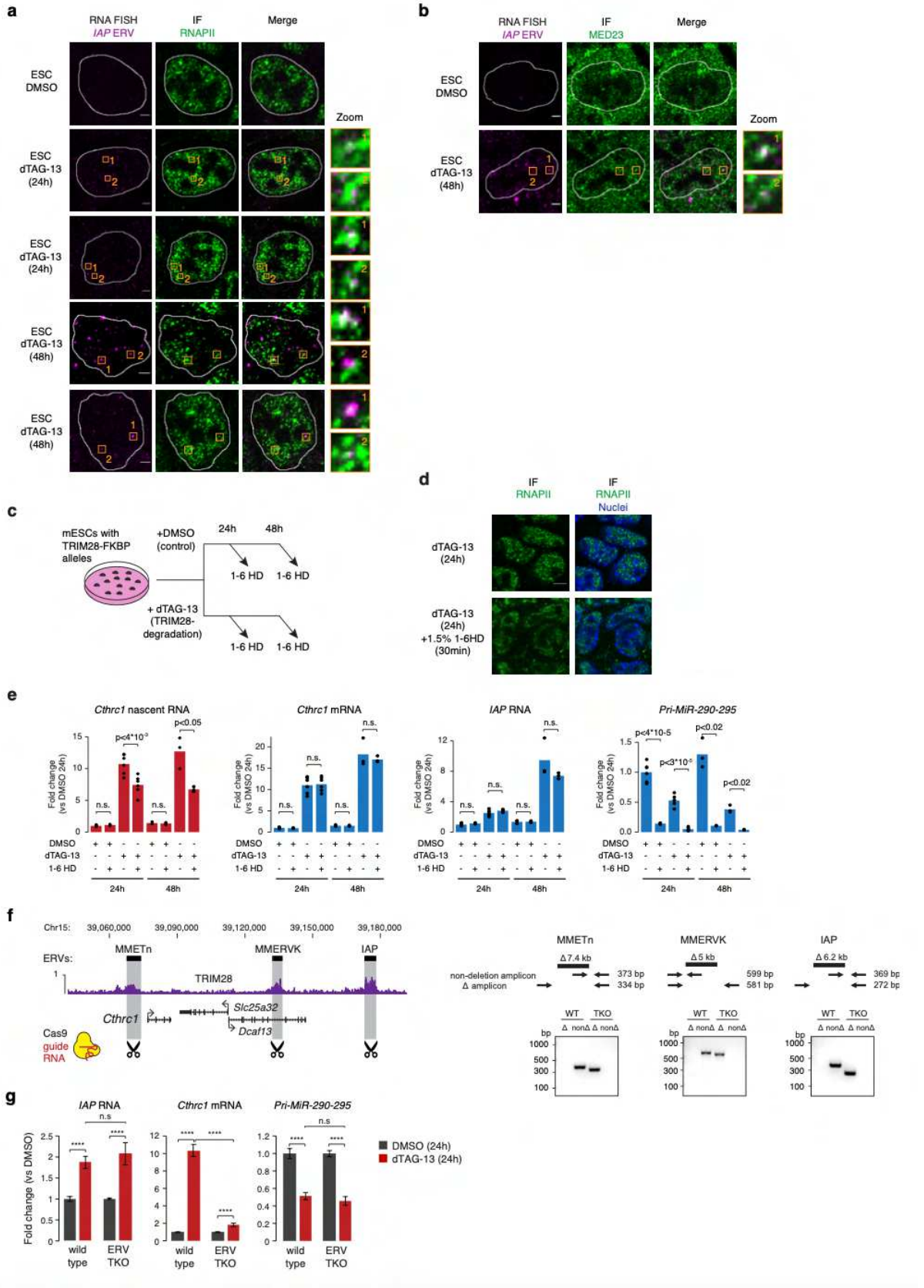


Extended Data Fig. 6. Loss of SE-association with RNAPII puncta at the *Mir290-295* super-enhancer

a. Co-localization between the nascent RNA of *MiR290-295* and RNAPII puncta in mESCs. The co-localization is lost upon TRIM28-degradation. Displayed are separate images of the RNA-FISH and IF signal, and an image of the merged channels. The nuclear periphery determined by DAPI staining (not shown) is highlighted as a white contour. Also shown are averaged signal of either RNA FISH or RNAPII IF centered on the *MiR290-295* FISH foci or randomly selected nuclear positions. These data were collected as an independent experimental verification of data displayed in Fig. 1i. Scale bars: 2.5μm.

b. dTAG-13 treatment leads to reduced RNAPII immunofluorescence signal at *MiR290-295* FISH foci, but does not change overall RNAPII levels. (left) quantification of RNAPII IF intensities at the *MiR290-295* FISH foci detected in the cells used in Fig. 1i. (center) quantification of RNAPII mean fluorescence intensity in the cells used in Fig. 1i. (right) quantification of the number of RNAPII puncta in the cells used in Fig. 1i. P values are from Mann Whitney tests n.s.: not significant.

Extended Data Fig. 7



1343

Extended Data Fig. 7. Condensate hijacking additional data

a-b. Co-localization between the *IAP* RNA and **(a)** RNAPII puncta, **(b)** MED23 puncta in TRIM28-degraded mESCs. Displayed are separate images of the RNA-FISH and IF signal, and an image of the merged channels. The nuclear periphery determined by DAPI staining (not shown) is highlighted as a white contour. The zoom column displays the region of the images highlighted in a yellow box zoomed in for greater detail. After 24h dTAG-13 treatment, small nuclear puncta appear, after 48h of dTAG-13 treatment large nuclear foci are visible. Scale bars: 2.5µm.

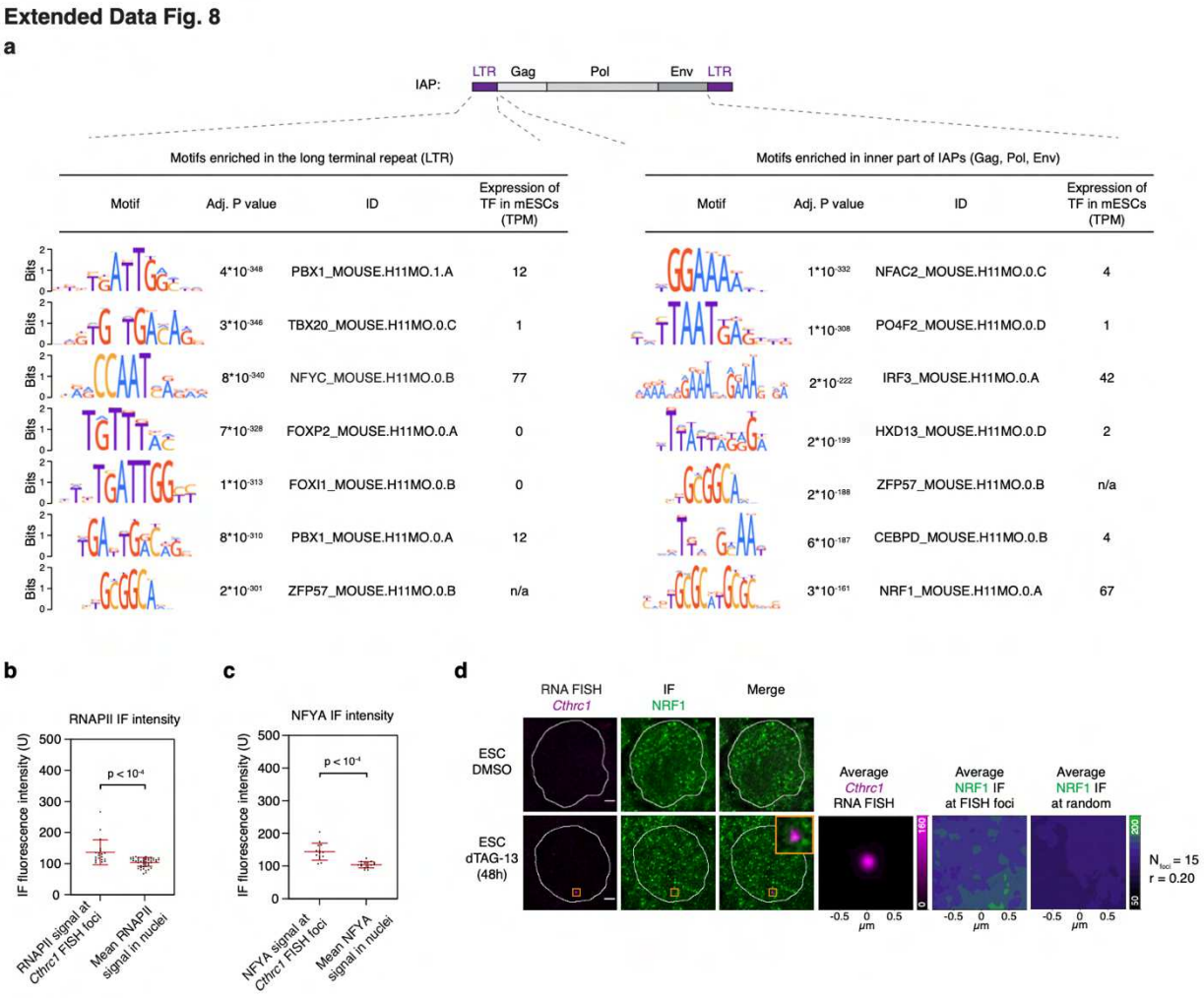
c. Scheme of the 1-6 hexanediol (1-6 HD) treatment experiments.

d. Representative images of RNAPII immunofluorescence in control and 1-6 HD-treated cells. 1-6 HD partially dissolved the punctate localization of RNAPII. Scale bars: 5µm.

e. Transcription of the nascent *Cthcr1* RNA is reduced by 30min 1% 1-6 hexanediol-treatment in TRIM28-degraded cells. The bar plots show qRT-PCR data as fold change normalized to the DMSO control. Note that the *IAP* RNA does not contain introns, thus the *IAP* RNA qRT-PCR detects the steady state pool of *IAP* RNAs. P values are from two-tailed t tests.

f. (left) The *Cthrc1* locus, highlighting the three ERVs that are deleted in the ERV TKO mESC line. TRIM28 ChIP-Seq data is shown above the gene models. (right) Genotyping PCR of the ERV TKO cell line.

g. Absence of *Cthrc1* induction in the ERV TKO cell line upon TRIM28- degradation. The bar plots show qRT-PCR data for *IAP* RNA, *Cthrc1* mRNA and the *Pri-MiR-290-295* transcript. Data are from biological duplicate experiments. P values are from two-tailed t tests. ****: $P<10^{-4}$



Extended Data Fig. 8. Additional characterization NFY and NRF1

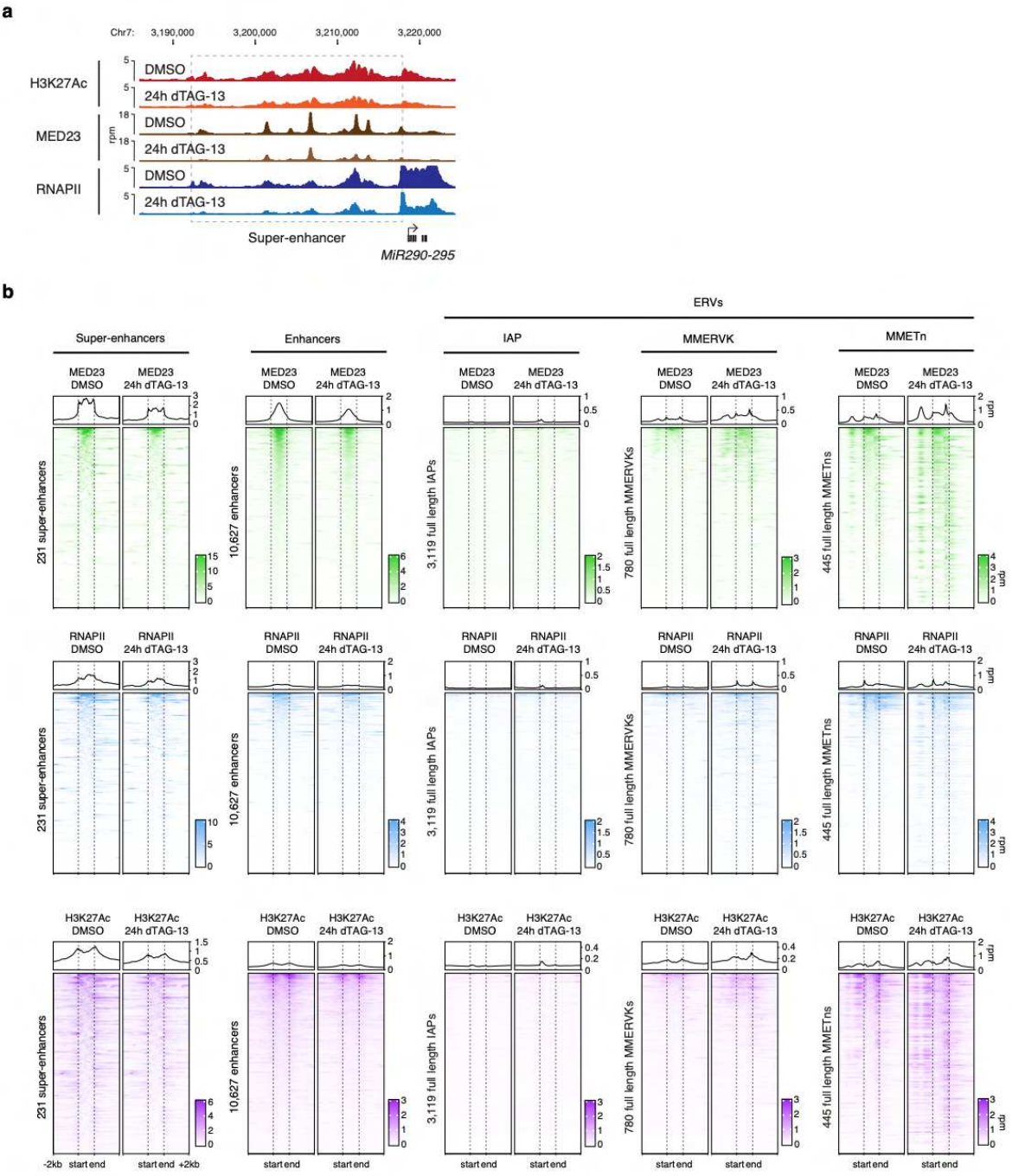
a. The sequence of IAPs is enriched for various TF binding motifs, including the motif of NFY. Top: schematic of an IAP element; Bottom: motif images, adjusted P-values and motif IDs, and the expression level of the TF in mESC RNA-Seq data. Displayed are the top scoring motifs based on adjusted P-value. Motifs were filtered for redundancy.

b. The RNAPII IF signal at *Cthrc1* FISH foci is higher than the average nuclear signal. Quantification of the RNAPII IF intensity at Cthrc1 FISH foci and nuclei in the cells analyzed in Fig. 2f is showed. P value is from a Mann Whitney test.

c. The NFYA IF signal at *Cthrc1* FISH foci is higher than the average nuclear signal. Quantification of the NFYA IF intensity at Cthrc1 FISH foci and nuclei in the cells analyzed in Fig. 2g is showed. P value is from a Mann Whitney test.

d. NRF1 puncta do not co-localize with the nascent RNA *Cthrc1* in TRIM28-degraded mESCs. Displayed are separate images of the RNA-FISH and IF signal, and an image of the merged channels. The nuclear periphery determined by DAPI staining (not shown) is highlighted as a white contour. Also shown are averaged signal of either RNA FISH or NRF1 IF centered on the *Cthrc1* FISH foci or randomly selected nuclear positions. Scale bars: 2.5μm.

Extended Data Fig. 9

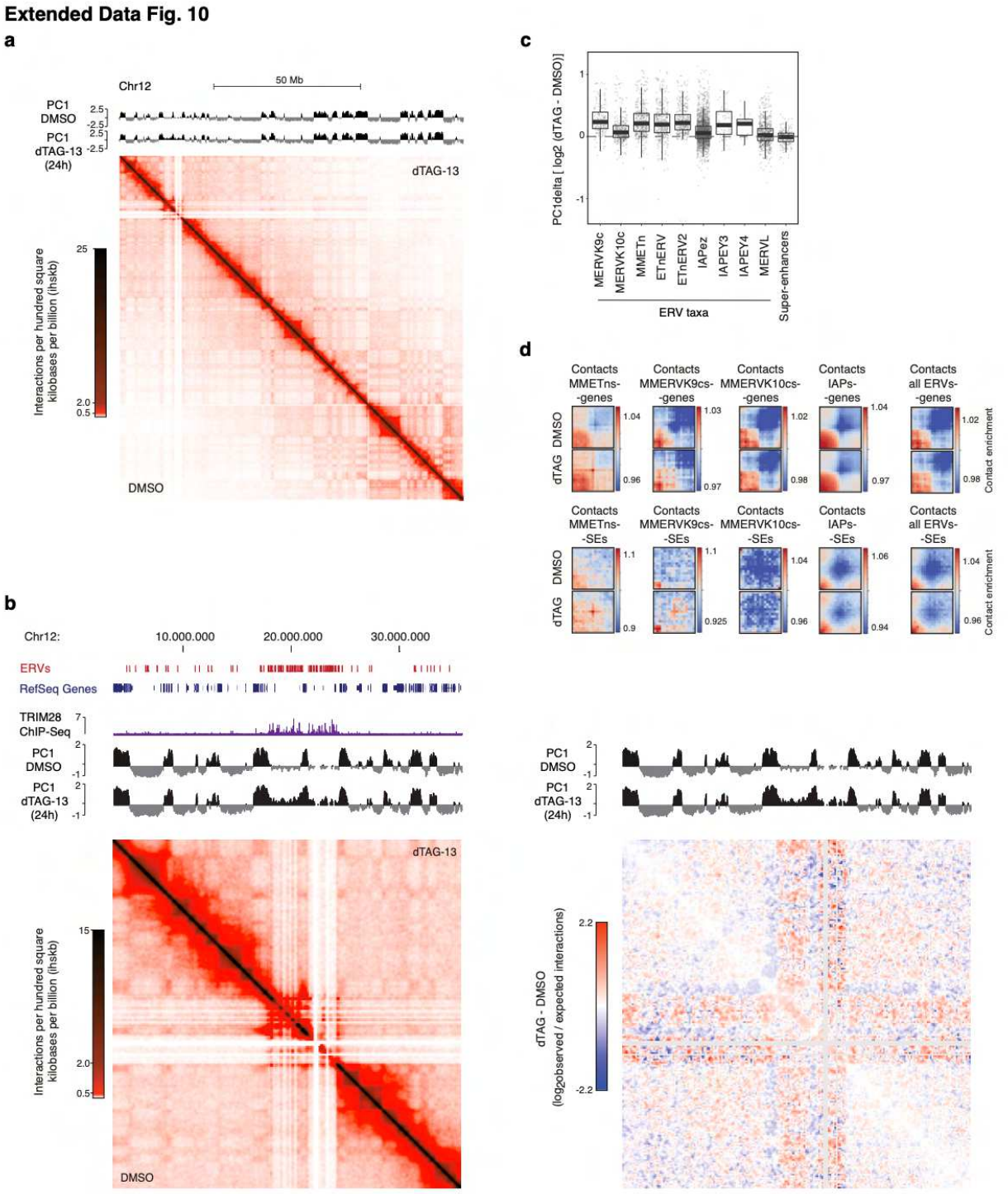


1382

1383 **Extended Data Fig. 9. Reduced occupancy of transcription-associated factors at super-**
1384 **enhancers and increased occupancy at ERVs in TRIM28-degraded mESCs**

1385 **a.** Genome browser tracks of ChIP-Seq data (H3K27Ac, RNAPII, MED23) in control mESCs
1386 and upon 24h dTAG-13 treatment at the *MiR290-295* locus. Rpm: reads per million.

b. Heatmap and meta representations of H3K27Ac, MED23 and RNAPII ChIP-Seq read densities at IAP, MMERVK and MMETn ERVs in control (DMSO) and dTAG-13 (24h) treated mESCs. The mean read densities are displayed +/-2kb around the indicated elements. The genomic elements (the middle part of the meta plot) were length normalized.



Extended Data Fig. 10. TRIM28-dagradation leads to shift of ERVs from the inactive "B" towards the active "A" compartment

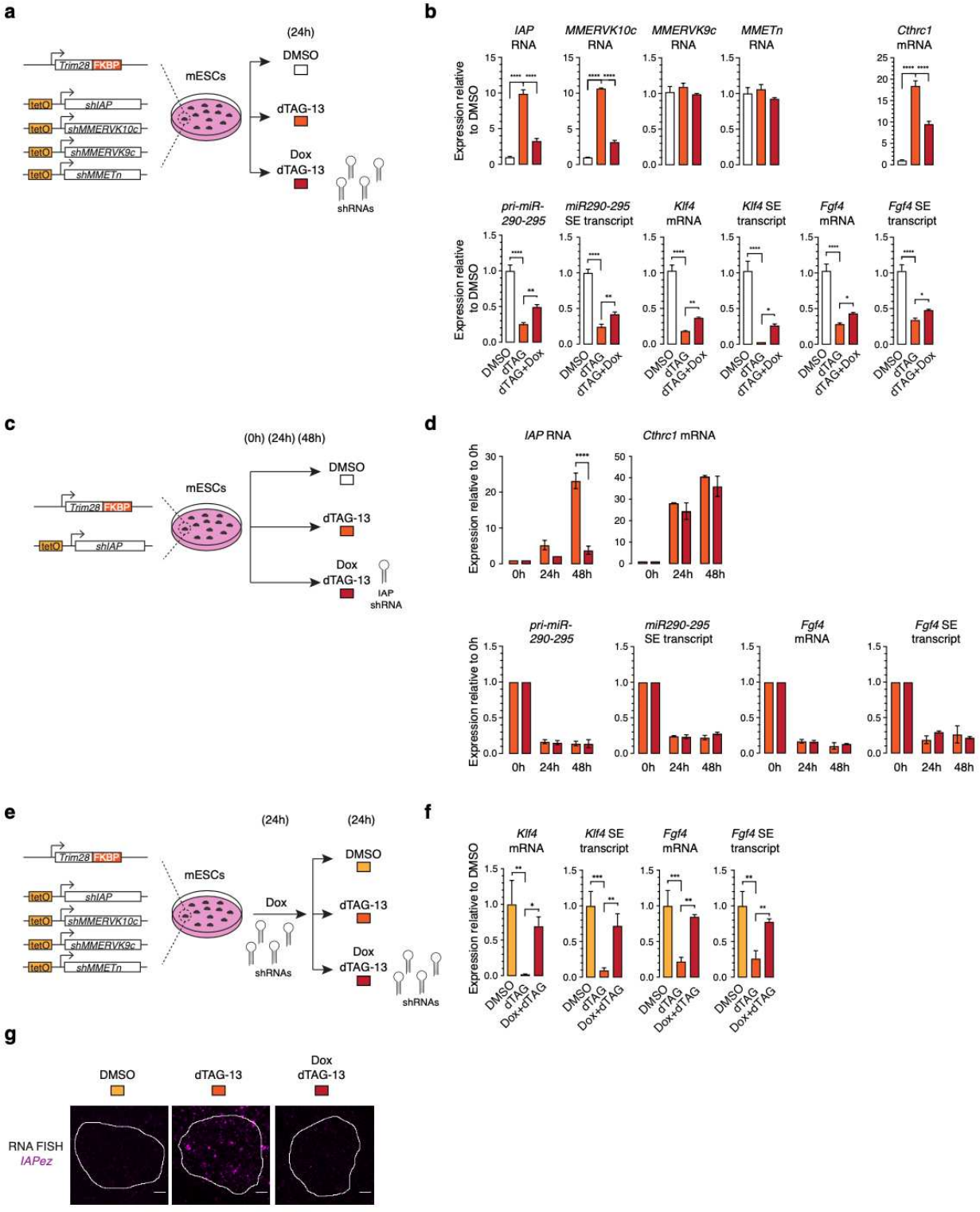
a. TRIM28-degradation does not lead to major genome-wide changes in chromatin contacts. Displayed are the Hi-C interaction matrices for chromosome 12, and the Principal Component 1 (PC1) of the normalized interaction matrix of the DMSO control and dTAG-treated samples.

b. (left) Hi-C interaction matrices for chromosome 12, zoomed-in on an ERV cluster, and the Principal Component 1 (PC1) of the normalized interaction matrix of the DMSO control and dTAG-treated samples. (right) differential interaction matrix of the region shown in the left.

c. TRIM28-degradation leads to shift of ERVs from the inactive "B" towards the active "A" compartment. Quantification of the mean compartment score (value of the PC1) of the indicated genomic regions. The values are displayed as the log2 fold change between the PC1 value in the dTAG-treated sample normalized against the values in the DMSO control sample.

d. Pile-up analysis of contacts between IAPs, MERVKs, MMETns and (top) transcribed genes or (bottom) super-enhancers (SEs) in wild type and TRIM28-degraded mESCs. The "all ERVs set" includes the combination of IAPs, MERVKs and MMETns.

Extended Data Fig. 11



Extended Data Fig. 11. Knockdown of ERV RNA rescues super-enhancer transcription in TRIM28-degraded cells

a. Schematic model of the experiment knocking down IAP, MMERVK10c, MMERKV9c and MMETn transcripts simultaneously with TRIM28 degradation (24h).

b. qRT-PCR analysis of IAP, MMERVK10c, MMERKV9c, MMETn, *Cthrc1*, *MiR290-295* primary transcript, *MiR290-295* super-enhancer RNA, *Klf4* mRNA, *Klf4* super-enhancer RNA, *Fgf4* mRNA, *Fgf4* super-enhancer RNA. Note that the qPCR primers used for MMETn and MERVK9c elements fail to detect induction after dTAG-13 treatment, which is explained by these elements having degenerate genomic sequences. Nevertheless, the combined knockdown in part rescues the induction of *Cthrc1* and downregulation of super-enhancers and their associated genes in TRIM28-degraded cells. P values are from two-tailed t tests. ****: $P<10^{-4}$, ***: $P<10^{-3}$, **: $P<10^{-2}$, *: $P<0.05$

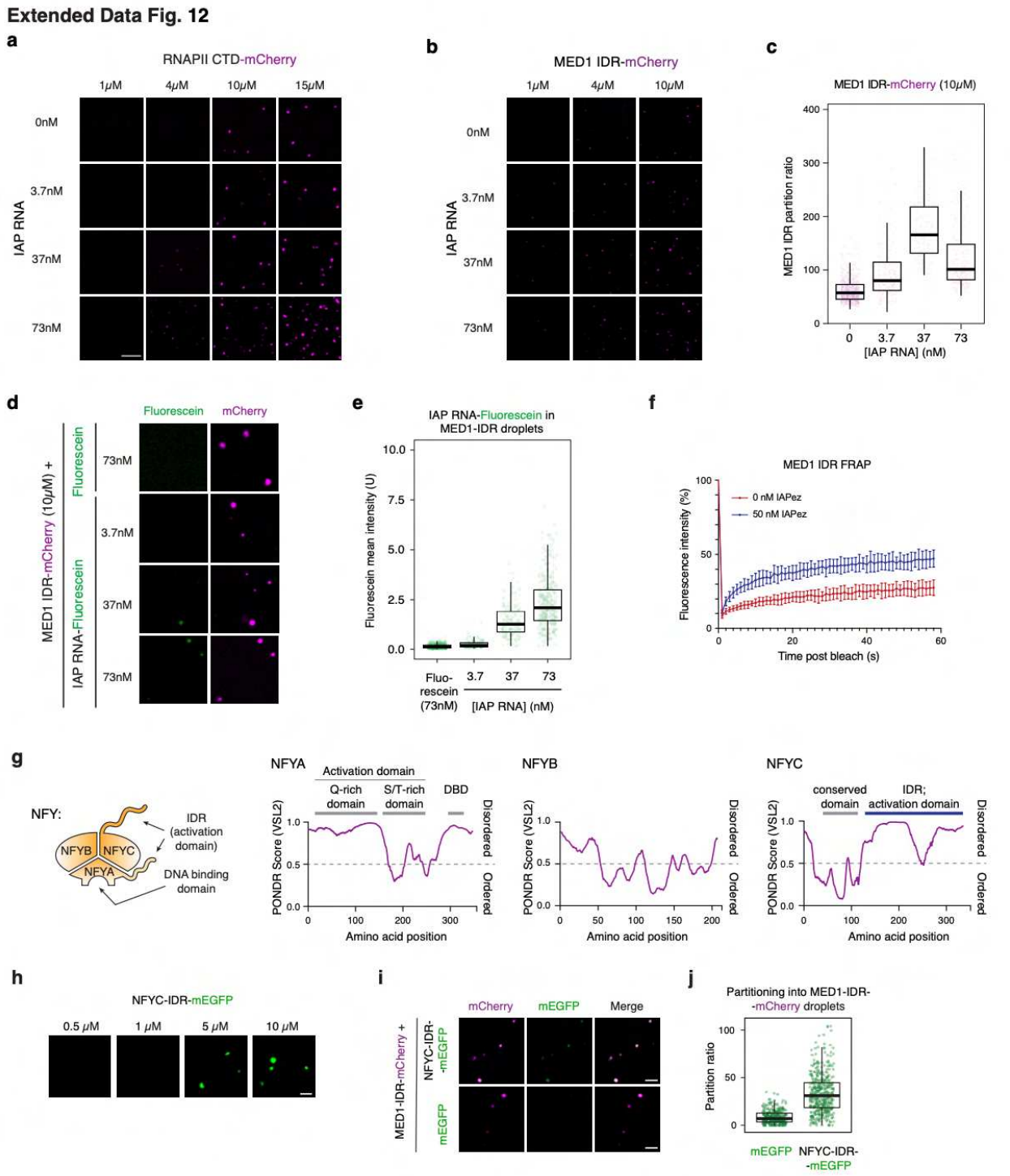
c. Schematic model of the experiment knocking down IAP transcripts simultaneously with TRIM28 degradation (24h).

d. IAP knockdown alone failed to rescue downregulation of super-enhancers and their associated transcripts The bar plots show qRT-PCR data as fold change normalized to the 0h control. P values are from two-tailed t tests. ****: $P<10^{-4}$, ***: $P<10^{-3}$, **: $P<10^{-2}$, *: $P<0.05$

e. Schematic model of the experiment in which shRNAs targeting IAP, MMERVK10c, MMERKV9c and MMETn transcripts are induced for 24h of Dox treatment, and the cells are treated either with DMSO (yellow), dTAG-13 (orange) or with dTAG-13 and Dox (maroon) for additional 24h.

f. Rescue of *Klf4* super-enhancer and primary transcript levels, and of *Fgf4* super-enhancer and primary transcript levels upon ERV RNA-knockdown in TRIM28-degraded cells. The bar plots show qRT-PCR data as fold change normalized to the Dox (24h) treatment control. P values are from two-tailed t tests. ****: $P<10^{-4}$, ***: $P<10^{-3}$, **: $P<10^{-2}$, *: $P<0.05$.

g. Representative images of *IAP* RNA FISH in cells described in panels (**e-f).** Scale bar: 2.5μm.



1435

1436 **Extended Data Fig. 12. IAP RNA facilitates droplet formation of transcriptional activators**
1437 **in vitro**

1438 **a.** IAP RNA facilitates RNAPII CTD droplet formation in vitro. Displayed are representative
1439 images of droplet formation by purified RNAPII CTD-mCherry fusion proteins in the presence

of in vitro transcribed IAP RNA fragments. As expected, fusion events between droplets were also observed. Scale bar: 10μm.

b. IAP RNA facilitates MED1 IDR droplet formation in vitro. Displayed are representative images of droplet formation by purified MED1 IDR-mCherry fusion proteins in the presence of in vitro transcribed IAP RNA.

c. Partitioning ratio of MED1 IDR-mCherry into droplets at the indicated IAP RNA concentrations. Every dot represents a detected droplet.

d. *IAP* RNA is enriched within MED1 IDR droplets. Displayed are representative images of the enrichment of fluorescein-labeled *IAP* RNA in MED1 IDR-mCherry droplets.

e. Quantification of the enrichment of fluorescein-labeled *IAP* RNA in MED1 IDR-mCherry droplets.

f. Fluorescence recovery after photobleaching (FRAP) experiments of MED1 IDR droplets in the absence or presence of 50nM *IAP* RNA.

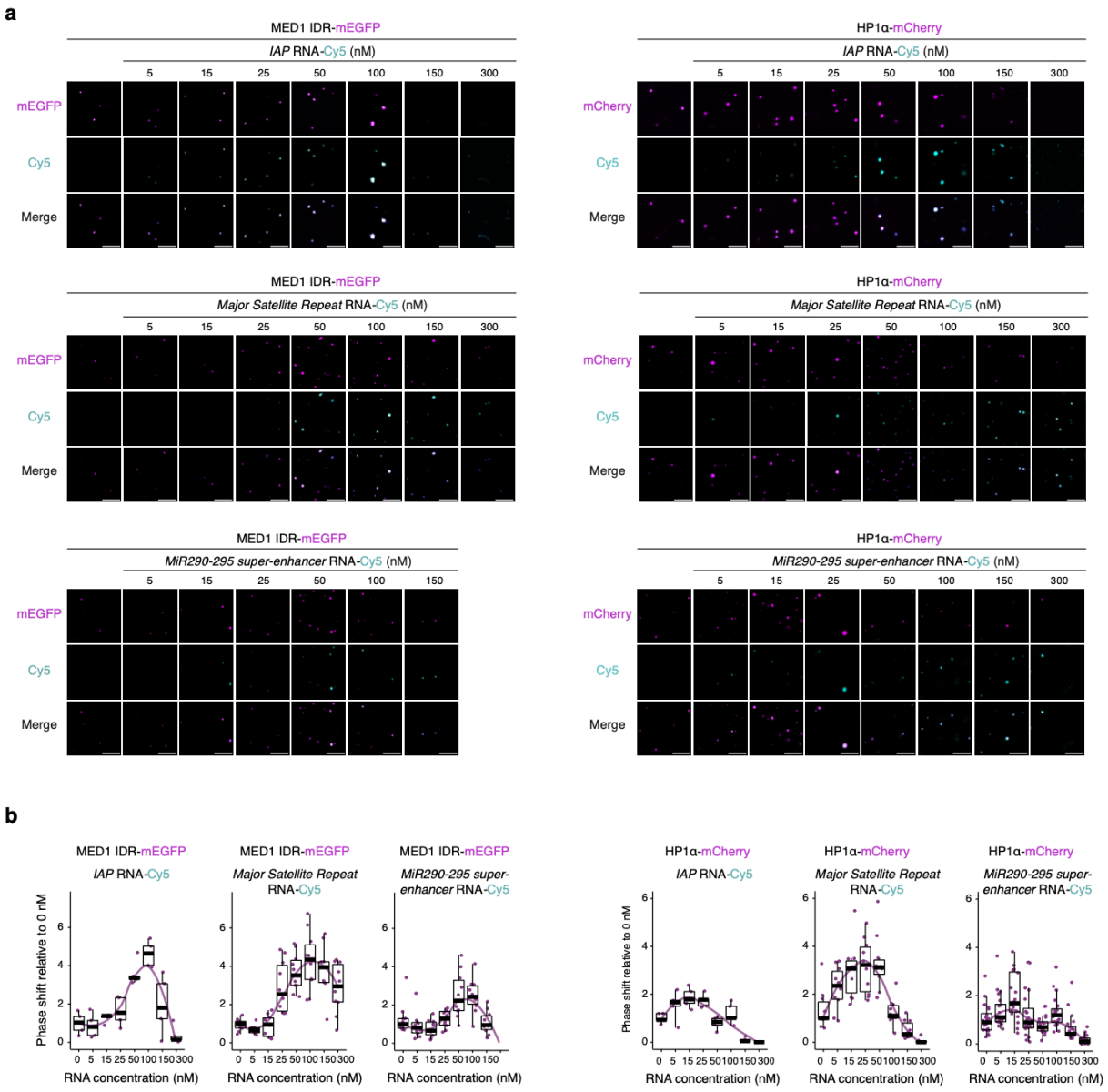
g. (left) Schematic model of the heterotrimeric NFY transcription factor. (right) Graphs plotting intrinsic disorder in the NFYA, NFYB and NFYC proteins. The NFYC IDR cloned for subsequent experiments is highlighted with a blue bar.

h. Concentration-dependent droplet formation by purified recombinant NFYC IDR-mEGFP. Scale bar: 10μm.

i. The NFYC IDR forms heterotypic droplets with MED IDR in vitro. Displayed are representative images of droplet formation by purified NFYC IDR-mEGFP and MED1 IDR-mCherry fusion proteins. MED1 IDR-mCherry mixed with purified mGFP is included as a control. Scale bar: 10μm.

j. Partitioning ratio of NFYC IDR-mEGFP or mEGFP in MED1 IDR-mCherry droplets. Every dot represents a detected droplet. P-value $< 2.2*10^{-16}$ (Welch's t-test).

Extended Data Fig. 13

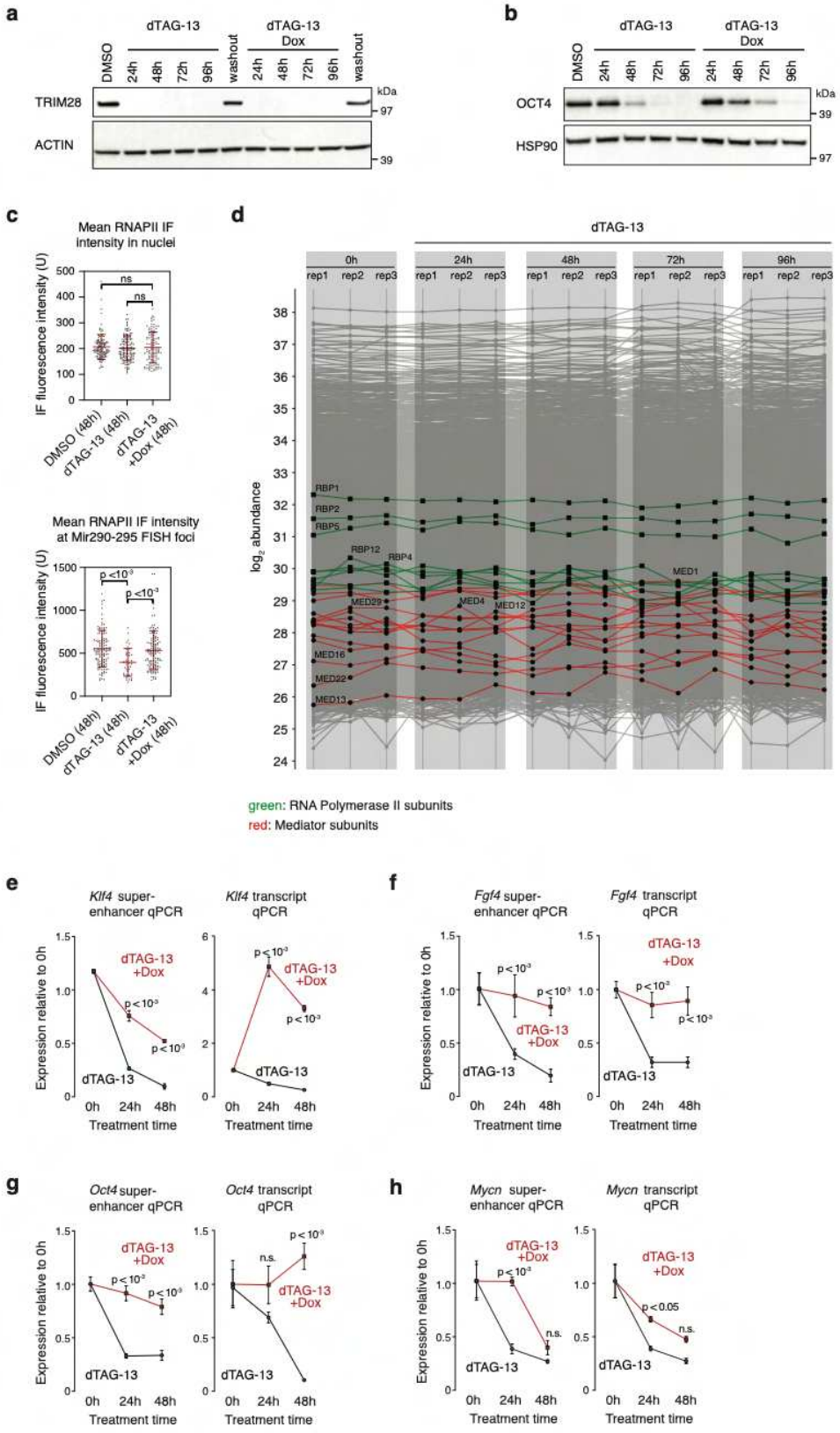


Extended Data Fig. 13. Effects of various RNAs on MED1 IDR and HPα droplets in vitro

a. RNA facilitates MED IDR and droplet HPα formation in vitro. Displayed are representative images of droplet formation by purified (left) MED1 IDR-mEGFP fusion protein and (right) HPα-mCherry fusion protein in the presence of in vitro transcribed RNA fragments. Scale bar: 10μm.

b. Quantification of the partitioning of (left) MED1 IDR and (right) HPIα into droplets in the presence of the indicated RNA species. Values are normalized against the partition ratio at no RNA added.

Extended Data Fig. 14



1473

Extended Data Fig. 14. Additional characterization of the OSKM/dTAG-13 experiments

a. Western blot validation of the FKBP degron tag and its ability to degrade TRIM28 in iPSCs. Washout of the dTAG-13 ligand (24h) indicates reversibility of degradation. Actin is shown as the loading control.

b. Western blot validation of the OSKM ectopic expression in the iPSC line. HSP90 is shown as the loading control.

c. dTAG-13 treatment leads to reduced RNAPII immunofluorescence signal at *MiR290-295* FISH foci which is rescued by OSKM ectopic expression, while overall RNAPII levels do not change. (top) Quantification of RNAPII mean fluorescence intensity in the cells used in Fig. 3f. (bottom) Quantification of RNAPII IF intensities at the *MiR290-295* FISH foci detected in the cells used in Fig. 3f. P value is from Mann Whitney test.

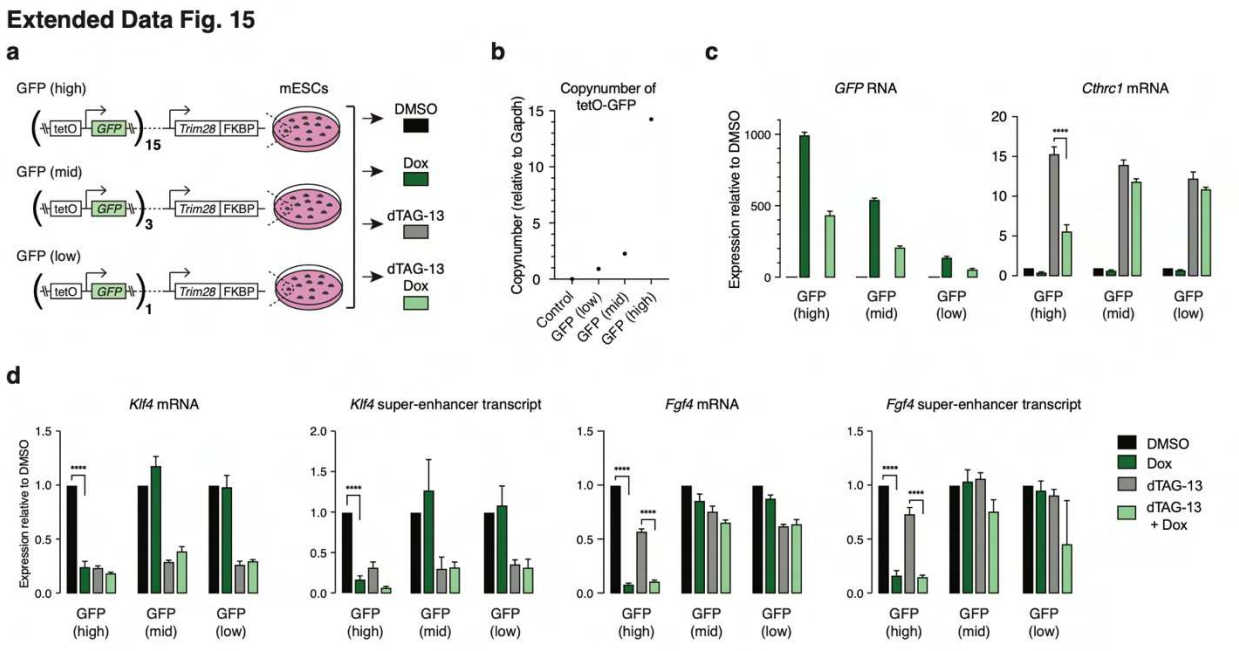
d. TRIM28-degradation does not substantially alter the protein levels of subunits of RNAPII and Mediator complex. Displayed are mass spectrometry-detected protein abundance for three individual replicate samples after 0h, 24h, 48h, 72h and 96h dTAG-13 treatment of mESCs. RNAPII subunits are highlighted in green. Mediator complex subunits are highlighted in red.

e. Elevated levels of *Klf4* super-enhancer transcript and *Klf4* nascent transcript in TRIM28-degraded iPSCs that ectopically express OSKM factors. The line plots show qRT-PCR data normalized to the 0h of dTAG-13 treatment. P values are from two-tailed t tests.

f. Elevated levels of *Fgf4* super-enhancer transcript and *Fgf4* nascent transcript in TRIM28-degraded iPSCs that ectopically express OSKM factors. The line plots show qRT-PCR data normalized to the 0h of dTAG-13 treatment. P values are from two-tailed t tests.

g. Elevated levels of *Oct4* super-enhancer transcript and *Oct4* nascent transcript in TRIM28-degraded iPSCs that ectopically express OSKM factors. The line plots show qRT-PCR data normalized to the 0h of dTAG-13 treatment. P values are from two-tailed t tests.

h. Elevated levels of *Mycn* super-enhancer transcript and *Mycn* nascent transcript in TRIM28-degraded iPSCs that ectopically express OSKM factors. The line plots show qRT-PCR data normalized to the 0h of dTAG-13 treatment. P values are from two-tailed t tests.



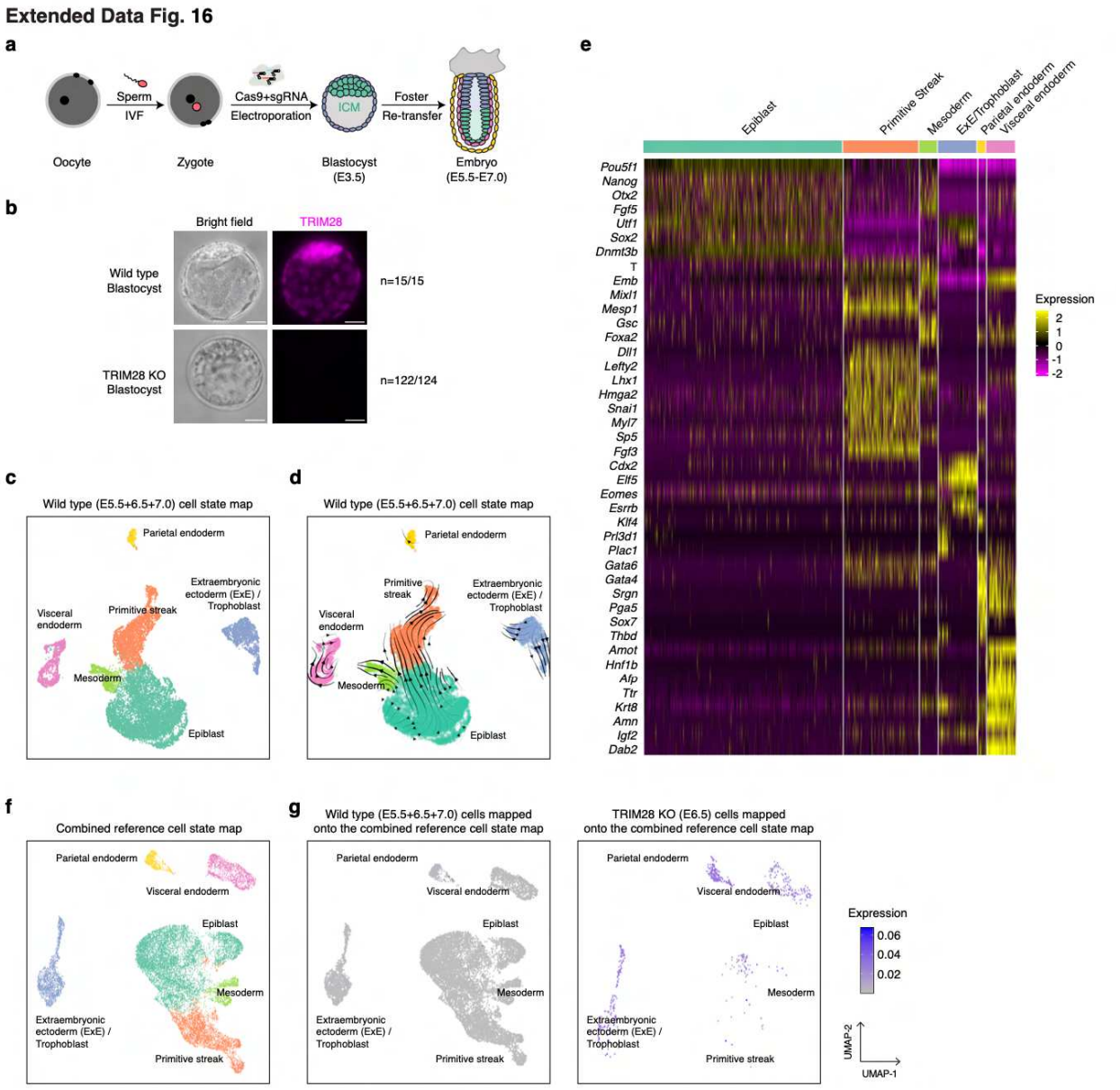
Extended Data Fig. 15. Simultaneous transcription of a high copy number of GFP transgenes compromises super-enhancer transcription

a. Schematic model of the experiment mimicking ERV transcription. mESC lines harboring 15, 3 or 1 copies of a Piggyback transposon encoding a Dox-inducible GFP transgene and degradation sensitive TRIM28-FKBP alleles were treated with Dox (to induce GFP expression) and/or dTAG-13 (to induce TRIM28 degradation).

b. qRT-PCR quantification of the copy number of GFP transgenes in the GPP (high), GFP (mid), GFP (low) and control (i.e. parental) mESC cell lines.

c. qRT-PCR quantification of GFP RNA and *Cthrc1* mRNA following the indicated treatments. Data are from biological triplicate experiments. P value is from a two-tailed t test. ****: $P<10^{-4}$

d. Simultaneous transcription of a high copy number of GFP transgenes leads to a reduction of transcription at the *Klf4* and *Fgf4* super-enhancers and their associated genes. The bar plots show qRT-PCR data as fold change normalized to the DMSO control treatment. P values are from two-tailed t tests. ****: $P<10^{-4}$, ***: $P<10^{-3}$.



Extended Data Fig. 16. Reference cell state maps in early mouse embryos

a. Scheme of the zygotic CRISPR/Cas9 – scRNA-Seq platform.

b. Immunofluorescence verification of TRIM28 KO. Representative images are shown, with the number of embryos where the immunofluorescence confirmed the genotype per the total number of embryos analyzed. Scale bars: 20µm.

c. UMAP of wild type early mouse embryos spanning E5.5-E7.0 developmental window. The wild type cells used in scRNA-Seq experiments from E5.5, E6.5 and E7.0 developmental stages were included.

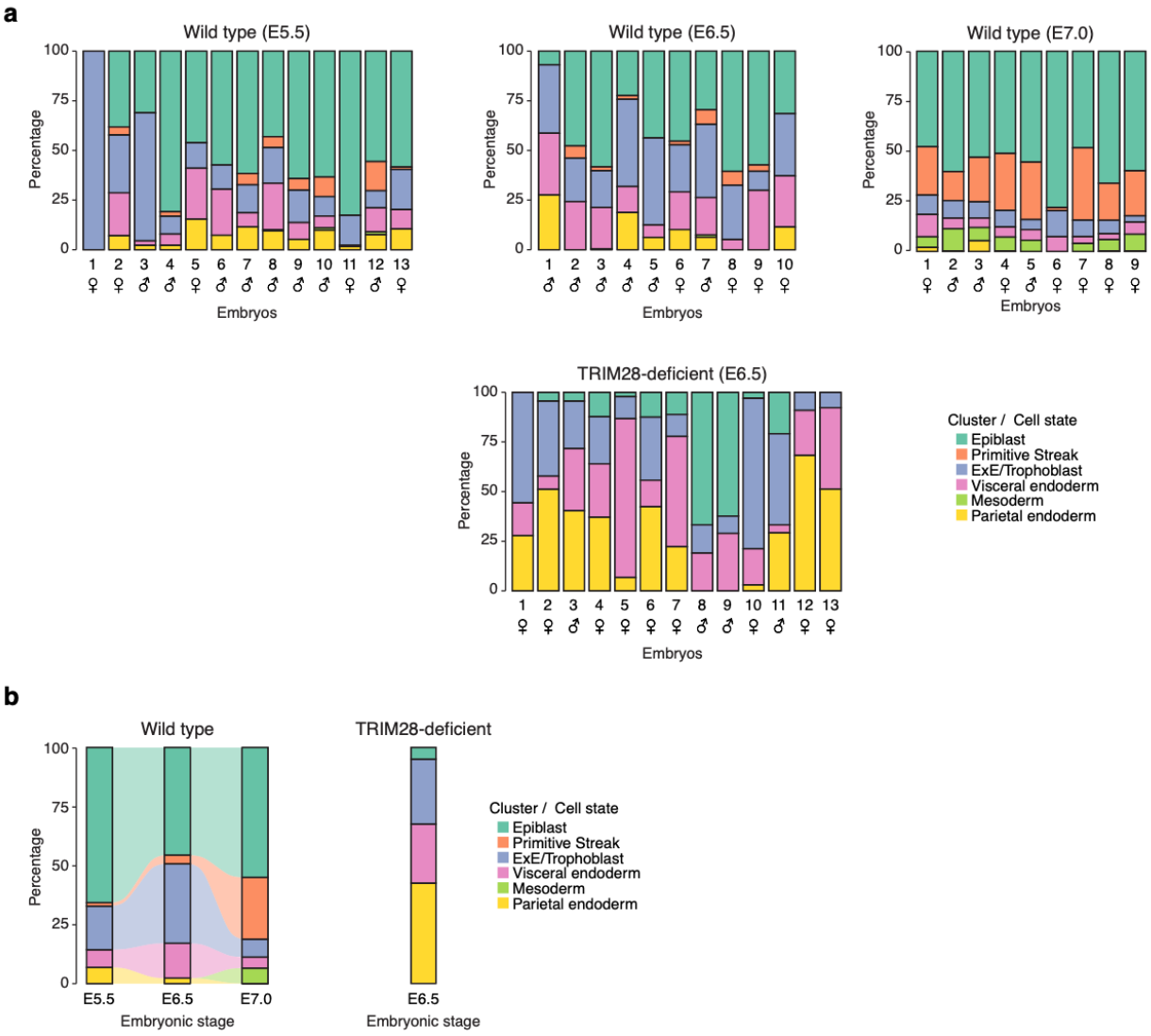
d. RNA velocity map of wild type early mouse embryos spanning E5.5-E7.0 developmental window.

e. Heatmap representation of the expression levels of marker genes of each cluster/cell state.

f. Combined cell state map. The wild type cells and TRIM28 KO cells used in scRNA-Seq experiments were included.

g. Elevated *IAP* expression in TRIM28 KO cells. (left) Wild type cells used in scRNA-Seq experiments from E5.5, E6.5 and E7.0 developmental stages are projected on the combined reference map, and are colored according to *IAP* expression. (right) TRIM28 KO E6.5 cells are projected on the combined reference map, and are colored according to *IAP* expression.

Extended Data Fig. 17



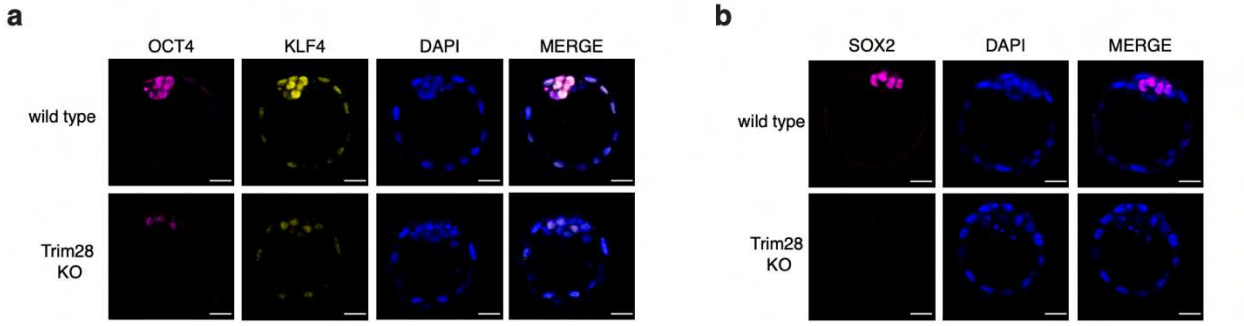
1535

1536 **Extended Data Fig. 17. Cell state proportions in individual embryos reveal depletion of**
1537 **epiblast cells in E6.5 TRIM28 KO mouse embryos**

1538 **a.** (top) Cell state proportions in each embryo from the E5.5 (left), E6.5 (middle), and E7.0
1539 (right) developmental stages analyzed in scRNA-Seq experiments. (bottom) Cell state
1540 proportions in E6.5 TRIM28 KO embryo analyzed in scRNA-Seq experiments.

1541 **b.** Combined cell state proportions per embryonic stage.

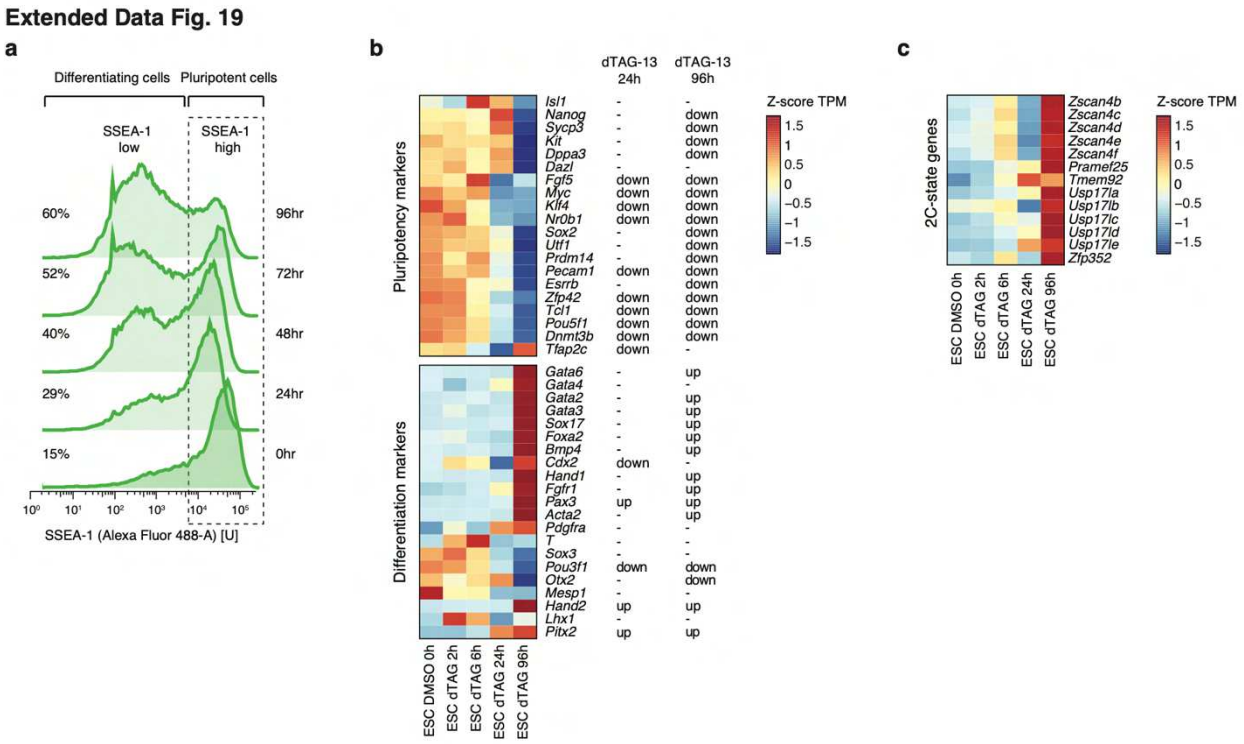
Extended Data Fig. 18



Extended Data Fig. 18. Loss of pluripotency markers in TRIM28 KO blastocysts

a. The cells in the inner part of TRIM28 KO blastocysts are populated by cells that express reduced levels of OCT4 and KLF4. Displayed are representative immunofluorescence images of OCT4 and KLF4 in E3.5 wild type and TRIM28 KO blastocysts. Nuclei are stained with DAPI. Scale bars: 20µm. Scale bars: 20µm.

b. The cells in the inner part of TRIM28 KO blastocysts are populated by cells that express reduced levels of SOX2. Displayed are representative immunofluorescence images of SOX2 and in E3.5 wild type and TRIM28 KO blastocysts. Nuclei are counterstained with DAPI. Scale bars: 20µm.



1552

1553 **Extended Data Fig. 19. Long term TRIM28 degradation induces differentiation markers in**
1554 **ESCs**

1555 **a.** 96h of TRIM28 degradation leads to an increased fraction of cells with reduced expression of
1556 the SSEA-1 pluripotency marker. Displayed are FACS profiles of cell populations treated for the
1557 indicated times with dTAG-13.

1558 **b.** Upregulation of differentiation marker genes upon dTAG-13 treatment in mESCs. Displayed
1559 is a heatmap representation of RNA-Seq read density (TPM) Z scores for the indicated genes.
1560 The annotation on the right indicates whether a gene is identified as differentially expressed at
1561 the indicated time points relative to the level at 0h.

1562 **c.** Upregulation of 2C marker genes upon dTAG-13 treatment in mESCs. Displayed is a heatmap
1563 representation of RNA-Seq read density (TPM) Z scores for the indicated genes. The annotation
1564 on the right indicates whether a gene is identified as differentially expressed at the indicated time
1565 points relative to the level at 0h.

Supplementary Table Legends

Table S1. List of reagents and datasets

a. RNA FISH probes

b. Primers used in the study

c. GEO accession IDs of public data used in the study

Table S2. Feature annotations

a. full length IAPs (coordinates in mm10 genome assembly coordinates)

b. full length MMERVKs (coordinates in mm10 genome assembly coordinates)

c. full length MMETns (coordinates in mm10 genome assembly coordinates)

d. Full length LINEs (coordinates in mm10 genome assembly coordinates)

e. All IAP LTRs (coordinates in mm10 genome assembly coordinates)

f. mESC enhancers (OSN constituent enhancers) (coordinates in mm10 genome assembly coordinates)

g. mESC super-enhancers (coordinates in mm10 genome assembly coordinates)

h. mESC typical enhancer (coordinates in mm10 genome assembly coordinates)

Table S3. Mass-spectrometry data

This table contains the mass spectrometry data used in Extended Data Fig. 2c and 14d.

Table S4. TT-SLAM-Seq fold change data (genes)

This table contains the fold change values used in Fig. 1g.

Table S5. TT-SLAM-Seq fold change data (enhancers)

This table contains the fold change values used in Fig. 1h.

Table S6. Marker genes in scRNA-Seq clusters

This table contains information on the marker genes used to identify clusters in the scRNA-Seq data.

References

1 Thompson, P. J., Macfarlan, T. S. & Lorincz, M. C. Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire. *Molecular cell* **62**, 766-776, doi:10.1016/j.molcel.2016.03.029 (2016).

2 Grewal, S. I. & Jia, S. Heterochromatin revisited. *Nature reviews. Genetics* **8**, 35-46, doi:10.1038/nrg2008 (2007).

3 Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nature reviews. Genetics* **10**, 691-703, doi:10.1038/nrg2640 (2009).

4 Payer, L. M. & Burns, K. H. Transposable elements in human genetic disease. *Nature reviews. Genetics* **20**, 760-772, doi:10.1038/s41576-019-0165-8 (2019).

5 Rowe, H. M. & Trono, D. Dynamic control of endogenous retroviruses during development. *Virology* **411**, 273-287, doi:10.1016/j.virol.2010.12.007 (2011).

6 Wells, J. N. & Feschotte, C. A Field Guide to Eukaryotic Transposable Elements. *Annual review of genetics* **54**, null, doi:10.1146/annurev-genet-040620-022145 (2020).

7 Rowe, H. M. *et al.* KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237-240, doi:10.1038/nature08674 (2010).

8 Cammas, F. *et al.* Mice lacking the transcriptional corepressor TIF1beta are defective in early postimplantation development. *Development* **127**, 2955-2963 (2000).

9 Matsui, T. *et al.* Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**, 927-931, doi:10.1038/nature08858 (2010).

10 Hu, G. *et al.* A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes & development* **23**, 837-848, doi:10.1101/gad.1769609 (2009).

11 Xu, W. *et al.* METTL3 regulates heterochromatin in mouse embryonic stem cells. *Nature* **591**, 317-321, doi:10.1038/s41586-021-03210-1 (2021).

12 Chelmicki, T. *et al.* m(6)A RNA methylation regulates the fate of endogenous retroviruses. *Nature* **591**, 312-316, doi:10.1038/s41586-020-03135-1 (2021).

13 Liu, J. *et al.* The RNA m(6)A reader YTHDC1 silences retrotransposons and guards ES cell identity. *Nature* **591**, 322-326, doi:10.1038/s41586-021-03313-9 (2021).

14 Mouse Genome Sequencing, C. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-562, doi:10.1038/nature01262 (2002).

15 Maksakova, I. A. *et al.* Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. *PLoS genetics* **2**, e2, doi:10.1371/journal.pgen.0020002 (2006).

16 Wolf, G. *et al.* KRAB-zinc finger protein gene expansion in response to active retrotransposons in the murine lineage. *eLife* **9**, doi:10.7554/eLife.56337 (2020).

17 Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116-120, doi:10.1038/35065132 (2001).

18 Wolf, G. *et al.* The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes & development* **29**, 538-554, doi:10.1101/gad.252767.114 (2015).

19 Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. & Rauscher, F. J., 3rd. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that

contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes & development* **16**, 919-932, doi:10.1101/gad.973302 (2002).
20 Yang, B. X. *et al.* Systematic identification of factors for provirus silencing in embryonic stem cells. *Cell* **163**, 230-245, doi:10.1016/j.cell.2015.08.037 (2015).
21 Elsasser, S. J., Noh, K. M., Diaz, N., Allis, C. D. & Banaszynski, L. A. Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature* **522**, 240-244, doi:10.1038/nature14345 (2015).
22 Bulut-Karslioglu, A. *et al.* Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Molecular cell* **55**, 277-290, doi:10.1016/j.molcel.2014.05.029 (2014).
23 Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57-63, doi:10.1038/nature11244 (2012).
24 Rowe, H. M. *et al.* TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome research* **23**, 452-461, doi:10.1101/gr.147678.112 (2013).
25 Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351**, 1083-1087, doi:10.1126/science.aad5497 (2016).
26 Nabet, B. *et al.* The dTAG system for immediate and target-specific protein degradation. *Nat Chem Biol* **14**, 431-441, doi:10.1038/s41589-018-0021-8 (2018).
27 Reichholf, B. *et al.* Time-Resolved Small RNA Sequencing Unravels the Molecular Principles of MicroRNA Homeostasis. *Molecular cell* **75**, 756-768 e757, doi:10.1016/j.molcel.2019.06.018 (2019).
28 Muhar, M. *et al.* SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. *Science* **360**, 800-805, doi:10.1126/science.aao2793 (2018).
29 Schwalb, B. *et al.* TT-seq maps the human transient transcriptome. *Science* **352**, 1225-1228, doi:10.1126/science.aad9841 (2016).
30 Boehning, M. *et al.* RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nature structural & molecular biology* **25**, 833-840, doi:10.1038/s41594-018-0112-y (2018).
31 Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science*, doi:10.1126/science.aar3958 (2018).
32 Cho, W. K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science*, doi:10.1126/science.aar4199 (2018).
33 Guo, Y. E. *et al.* Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543-548, doi:10.1038/s41586-019-1464-0 (2019).
34 Roden, C. & Gladfelter, A. S. RNA contributions to the form and function of biomolecular condensates. *Nature reviews. Molecular cell biology*, doi:10.1038/s41580-020-0264-6 (2020).
35 Henninger, J. E. *et al.* RNA-Mediated Feedback Control of Transcriptional Condensates. *Cell* **184**, 207-225 e224, doi:10.1016/j.cell.2020.11.030 (2021).
36 Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855 e1816, doi:10.1016/j.cell.2018.10.042 (2018).
37 Larson, A. G. *et al.* Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. *Nature* **547**, 236-240, doi:10.1038/nature22822 (2017).

38 Huo, X. *et al.* The Nuclear Matrix Protein SAFB Cooperates with Major Satellite RNAs to Stabilize Heterochromatin Architecture Partially through Phase Separation. *Molecular cell* **77**, 368-383 e367, doi:10.1016/j.molcel.2019.10.001 (2020).
39 Wernig, M. *et al.* A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nature biotechnology* **26**, 916-924, doi:10.1038/nbt1483 (2008).
40 Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).
41 Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583-588, doi:10.1038/nature14136 (2015).
42 Grosswendt, S. *et al.* Epigenetic regulator function through mouse gastrulation. *Nature* **584**, 102-108, doi:10.1038/s41586-020-2552-x (2020).
43 Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631-642, doi:10.1016/s0092-8674(03)00393-3 (2003).
44 Wang, J. *et al.* Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **516**, 405-409, doi:10.1038/nature13804 (2014).
45 Kunarso, G. *et al.* Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature genetics* **42**, 631-634, doi:10.1038/ng.600 (2010).
46 Quinodoz, S. A. *et al.* RNA promotes the formation of spatial compartments in the nucleus. *bioRxiv*, 2020.2008.2025.267435, doi:10.1101/2020.08.25.267435 (2020).
47 Deniz, O. *et al.* Endogenous retroviruses are a source of enhancers with oncogenic potential in acute myeloid leukaemia. *Nature communications* **11**, 3506, doi:10.1038/s41467-020-17206-4 (2020).
48 Geis, F. K. & Goff, S. P. Silencing and Transcriptional Regulation of Endogenous Retroviruses: An Overview. *Viruses* **12**, doi:10.3390/v12080884 (2020).
49 Li, W. *et al.* Human endogenous retrovirus-K contributes to motor neuron disease. *Sci Transl Med* **7**, 307ra153, doi:10.1126/scitranslmed.aac8201 (2015).
50 Karlsson, H. *et al.* Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 4634-4639, doi:10.1073/pnas.061021998 (2001).
51 Jonsson, M. E. *et al.* Activation of endogenous retroviruses during brain development causes an inflammatory response. *The EMBO journal* **40**, e106423, doi:10.15252/embj.2020106423 (2021).
52 Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature methods* **11**, 319-324, doi:10.1038/nmeth.2834 (2014).
53 Ni, Y. *et al.* Mutations in NDUFS1 Cause Metabolic Reprogramming and Disruption of the Electron Transfer. *Cells* **8**, doi:10.3390/cells8101149 (2019).
54 Martens, L. *et al.* PRIDE: the proteomics identifications database. *Proteomics* **5**, 3537-3545, doi:10.1002/pmic.200401303 (2005).
55 Orlando, D. A. *et al.* Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell reports* **9**, 1163-1170, doi:10.1016/j.celrep.2014.10.018 (2014).
56 Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680, doi:10.1016/j.cell.2014.11.021 (2014).

57 Basu, S. *et al.* Unblending of Transcriptional Condensates in Human Repeat Expansion Disease. *Cell* **181**, 1062-1079 e1030, doi:10.1016/j.cell.2020.04.018 (2020).
58 Nakagata, N. Cryopreservation of mouse spermatozoa and in vitro fertilization. *Methods in molecular biology* **693**, 57-73, doi:10.1007/978-1-60761-974-1_4 (2011).
59 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3, doi:10.14806/ej.17.1.200 (2011).
60 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
61 Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology* **33**, 290-295, doi:10.1038/nbt.3122 (2015).
62 Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research* **41**, D590-596, doi:10.1093/nar/gks1219 (2013).
63 Neumann, T. *et al.* Quantification of experimentally induced nucleotide conversions in high-throughput sequencing datasets. *BMC Bioinformatics* **20**, 258, doi:10.1186/s12859-019-2849-7 (2019).
64 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169, doi:10.1093/bioinformatics/btu638 (2015).
65 Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic acids research* **42**, W187-191, doi:10.1093/nar/gku365 (2014).
66 Ji, X. *et al.* Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 3841-3846, doi:10.1073/pnas.1502971112 (2015).
67 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
68 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).
69 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).
70 Kharchenko, P. V., Tolstorukov, M. Y. & Park, P. J. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature biotechnology* **26**, 1351-1359, doi:10.1038/nbt.1508 (2008).
71 Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947, doi:10.1016/j.cell.2013.09.053 (2013).
72 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
73 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).
74 Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature protocols* **4**, 1184-1191, doi:10.1038/nprot.2009.97 (2009).

75 Korotkevich, G., Sukhov, V. & Sergushichev, A. Fast gene set enrichment analysis. *bioRxiv*, 060012, doi:10.1101/060012 (2019).
76 Jurka, J. *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110**, 462-467, doi:10.1159/000084979 (2005).
77 Stocking, C. & Kozak, C. A. Murine endogenous retroviruses. *Cell Mol Life Sci* **65**, 3383-3398, doi:10.1007/s00018-008-8497-0 (2008).
78 Crichton, J. H., Dunican, D. S., Maclennan, M., Meehan, R. R. & Adams, I. R. Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline. *Cell Mol Life Sci* **71**, 1581-1605, doi:10.1007/s00018-013-1468-0 (2014).
79 Pockrandt, C., Alzamel, M., Iliopoulos, C. S. & Reinert, K. GenMap: ultra-fast computation of genome mappability. *Bioinformatics* **36**, 3687-3692, doi:10.1093/bioinformatics/btaa222 (2020).
80 Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841-1842, doi:10.1093/bioinformatics/btp328 (2009).
81 Gu, Z., Eils, R., Schlesner, M. & Ishaque, N. EnrichedHeatmap: an R/Bioconductor package for comprehensive visualization of genomic signal associations. *BMC genomics* **19**, 234, doi:10.1186/s12864-018-4625-x (2018).
82 Flyamer, I. M., Illingworth, R. S. & Bickmore, W. A. Coolpup.py: versatile pile-up analysis of Hi-C data. *Bioinformatics* **36**, 2980-2985, doi:10.1093/bioinformatics/btaa073 (2020).
83 McLeay, R. C. & Bailey, T. L. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11**, 165, doi:10.1186/1471-2105-11-165 (2010).
84 Chan, M. M. *et al.* Molecular recording of mammalian embryogenesis. *Nature* **570**, 77-82, doi:10.1038/s41586-019-1184-5 (2019).
85 Krueger, F. & Andrews, S. R. SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes. *F1000Res* **5**, 1479, doi:10.12688/f1000research.9037.2 (2016).
86 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821, doi:10.1016/j.cell.2019.05.031 (2019).
87 La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494-498, doi:10.1038/s41586-018-0414-6 (2018).
88 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome biology* **19**, 15, doi:10.1186/s13059-017-1382-0 (2018).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1NGA58214Hnisz210911.xlsx
- TableS2NGA58214Hnisz210911.xlsx
- TableS3NGA58214Hnisz210911.xlsx
- TableS4NGA58214Hnisz210911.xlsx
- TableS5NGA58214Hnisz210911.xlsx
- TableS6NGA58214Hnisz210911.xlsx