TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells

Rowe, Helen M.; Kapopoulou, Adamandia; Corsinotti, Andrea; Fasching, Liana; Macfarlan, Todd S.; Tarabay, Yara; Viville, Stephane; Jakobsson, Johan; Pfaff, Samuel L.; Trono, Didier

Published in:
Genome Research

DOI:
10.1101/gr.147678.112

Published: 2013-01-01

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells

Helen M. Rowe,1 Adamandia Kapopoulou,1,2 Andrea Corsinotti,1 Liana Fasching,3 Todd S. Macfarlan,4 Yara Tarabay,5 Stéphane Viville,5 Johan Jakobsson,3 Samuel L. Pfaff,4 and Didier Trono1,6

1School of Life Sciences and Frontiers in Genetics Program, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland; 2Swiss Bioinformatics Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland; 3Wallenberg Neuroscience Center, Lund University, BMC A11, 221 84 Lund, Sweden; 4Gene Expression Laboratory and the Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, California 92037, USA; 5Institute of Genetics and Molecular and Cellular Biology (IGBMC), University of Strasbourg, BP10142, Illkirch Cedex, France

TRIM28 is critical for the silencing of endogenous retroviruses (ERVs) in embryonic stem (ES) cells. Here, we reveal that an essential impact of this process is the protection of cellular gene expression in early embryos from perturbation by cis-acting activators contained within these retroelements. In TRIM28-depleted ES cells, repressive chromatin marks at ERVs are replaced by histone modifications typical of active enhancers, stimulating transcription of nearby cellular genes, notably those harboring bivalent promoters. Correspondingly, ERV-derived sequences can repress or enhance expression from an adjacent promoter in transgenic embryos depending on their TRIM28 sensitivity in ES cells. TRIM28-mediated control of ERVs is therefore crucial not just to prevent retrotransposition, but more broadly to safeguard the transcriptional dynamics of early embryos.

[Supplemental material is available for this article.]
 prevent retrotransposition, but more broadly to safeguard the timely activation of genes during early development.

Results

Transcriptional deregulation in Trim28 knock-out ES cells

Using a previously described tamoxifen-inducible Cre/lox system (Rowe et al. 2010), we first compared mRNA-sequencing (mRNA-seq) data from control and Trim28-deleted murine ES cells (Fig. 1A,B). Transcripts from ~20,000 genes were detected in control cells. Four days after Cre induction, based on a twofold cutoff and a significant difference of $P \leq 0.05$, around 5700 of them were up-regulated (29%), including 1850 transcripts that were more than twofold up-regulated, while around 720 were down-regulated (4%) and 13,600 unchanged (67%). From now on, we refer to these gene groups as “Up,” “Down,” and “Stable,” respectively. In contrast, in mouse embryonic fibroblasts (MEFs), transcriptional deregulation was only modest upon Trim28 deletion (Fig. 1A). This correlates the difference between the dramatic phenotype of Trim28-deleted ES cells, which die or differentiate after a few days and overexpress ERVs, and MEFs, which can be stably maintained and do not up-regulate ERVs (Rowe et al. 2010). Of note, genes affected by Trim28 deletion (both Up and Down) in ES cells were lowly expressed at baseline compared with genes unaffected by removal of this regulator (according to a Wilcoxon rank-sum test that was used to calculate significance here and for all boxplots) (Supplemental Fig. S1A). We decided to focus on up-regulated genes since they represented the larger category and Gene Ontology analysis indicated these genes to be involved in developmental pathways (see Supplemental Fig. S1B; Supplemental Table 1), including through expression at the embryonic two-cell stage as recently described (Macfarlan et al. 2012).

Chromatin state at genes affected by Trim28 deletion

Surprisingly, confrontation of these results with Trim28 ChIP-seq data performed in the same cells revealed that <1% of up-regulated gene promoters were direct targets of Trim28 (Supplemental Table 2). This suggested that Up genes could be indirectly affected by Trim28 deletion and/or were normally subjected to Trim28-controlled nearby cis-acting influences. We thus compared the chromatin status of Up, Down, and Stable genes more broadly using available ChIP-seq data (Mikkelsen et al. 2007). We focused on H3K4me3, a Trithorax group– or TrxG-deposited mark typically associated with active transcription, H3K9me3, frequently a signature of TRIM28/SETDB1 acting through expression at the embryonic two-cell stage (Bernstein et al. 2006; Gan et al. 2007; Guenther and Young 2010). As previously observed (Mikkelsen et al. 2007), H3K4me3 and H3K27me3 were significantly enriched at gene promoters, while H3K9me3 was generally depleted from these functional domains (Supplemental Fig. S1C). Genes deregulated upon TRIM28 deletion, whether up or down, were significantly closer to H3K9me3-enriched regions than

![Figure 1. Trim28 deletion in ES cells leads to up-regulation of genes close to ERVs, including many bivalent genes. (A) mRNA-seq in Trim28 wild-type (WT) and knock-out (KO) embryonic stem (ES) cells (left panel) or Trim28 WT and KO MEFs (right panel). Transcripts (assembly mm9) are plotted in black with the ratio on the y-axis and expression level on the x-axis. (Sqrt) Square root. (Horizontal lines) Levels of gene deregulation (e.g., only 1% of genes lie above the 99% line). The genes Zfp575, Prnp, and Serinc3 (referred to later) are highlighted, as well as Trim28. (B) Data from ES cells in A were used to group transcripts depending on whether they were greater than twofold up-regulated (Up), greater than twofold down-regulated (Down), or less than twofold affected (Stable). Up and Down genes were significantly changed based on a DESeq test (Anders and Huber 2010) (adjusted $P$-values $\leq 0.05$). (C) The distance to the nearest peak (of either H3K9me3 on the left panel, 19,128 peaks, or dual H3K27me3, H3K4me3 peaks on the right panel, 12,390 peaks) from Up, Down, and Stable gene groups. (Left $P$-values) Up versus Down, not significant (NS), $P = 0.48$; Up versus Stable, $P = 7.7 \times 10^{-10}$; Down versus Stable, $P = 0.0010$. (Right $P$-values) Up versus Down, $P = 9.9 \times 10^{-11}$; Up versus Stable, $P = 4.1 \times 10^{-04}$; Down versus Stable, $P = 4.1 \times 10^{-12}$. (D) Bivalent genes (as defined above by the presence of dual H3K27me3, H3K4me3 peaks) are enriched for up-regulated genes compared with all genes. (E) ERV locations ($N = 82,382$) were downloaded from the UCSC Genome Browser to include the categories ERV, ERV1, ERV, and ERVL as defined by Repbase with a size cutoff of 500-bp minimum and used to plot the distance to the nearest ERV from Up, Down, and Stable gene groups (left). A Mann-Whitney Wilcoxon test was used to calculate significance: Up genes were significantly closer than the other two gene groups; (***) $P = 0.001$. (Right) All genes were divided into groups based on their distance to the nearest ERV and their ratio between Trim28 WT and KO ES cells plotted on the y-axis. ($P$-values) The groups 10–20 versus 20–40 and 20–40 versus 40–100 are different: $P = 0.0048$ and $P = 0.01$, respectively. (F) Model showing that Up genes are close to H3K9me3 marks and ERVs and are often bivalent.
unaffected genes (Fig. 1C, left). More revealingly, Up genes almost completely coincided with H3K27me3 peaks (Supplemental Fig. S1D). In ES cells, the H3K27me3 repressive mark is found together with its activating counterpart H3K4me3 at so-called bivalent promoters, which are rapidly induced upon differentiation (Bernstein et al. 2006). We thus compared the relative distribution of these two marks over the three gene groups. Genes unaffected by TRIM28 removal were the closest to H3K4me3-alone peaks and the farthest away from H3K27me3-alone peaks (Supplemental Fig. S1E), consistent with their average higher levels of expression than Up or Down genes. In contrast and most strikingly, Up genes almost completely overlapped bivalent H3K4me3/H3K27me3 peaks (Fig. 1C, right), indicating that the promoters of many of the genes induced upon Trim28 deletion are poised for transcription. Reciprocally, up-regulated genes (2444) were enriched among bivalent genes (4999) (Mikkelsen et al. 2007), compared with all genes (Fig. 1D, Fisher’s exact test: P-value ≤ 1 × 10−16).

**Genes up-regulated upon Trim28 deletion are located close to ERVs**

Since few gene promoters were direct targets of TRIM28 (see above), we hypothesized that up-regulation of many genes could reflect the deregulation of TRIM28-controlled cis-acting elements situated in their nearby vicinity. In that respect, TRIM28, together with H3K9me3, is found enriched at ERV sequences in ES cells but not MEFs (Matsui et al. 2010; Rowe et al. 2010). Because ERVs are known to contain transcription-regulating sequences, we asked whether they were spatially associated with genes induced upon Trim28 deletion. Indeed, matching the genomic locations of ERVs (82,382 sites) with the three gene groups differentially affected by TRIM28 removal revealed that Up genes were on average significantly closer to these elements than Down or Stable genes (Fig. 1E, left). We also verified that it is not the case that all bivalent genes are enriched in ERVs but rather that bivalent Up genes (2444) are on average closer to ERVs than bivalent stable genes (2314, P = 0.001470) (Supplemental Fig. S2A). Interestingly, Up genes also clustered with long interspersed nuclear elements (LINE1s) but lay further from short interspersed nuclear elements (SINEs) than Down and Stable genes (Supplemental Fig. S2B–D), consistent with the previous observation that LINEs but not SINEs are modestly up-regulated in Trim28-deleted ES cells (Rowe et al. 2010). Reciprocally, the closer genes were to an ERV or particularly to an ERV of the subclass IAPs, the higher their average up-regulation upon TRIM28 removal, with genes also affected (although to a lesser extent) at distances of 100 kb (Fig. 1E, right; data not shown). Of note, this phenomenon of nearby cis-acting regulation is consistent with the previously documented modulation of the Agouti gene by an IAP located some 100 kb away, leading to variable coat colors in mice (Duhl et al. 1994; Michaud et al. 1994). In sum, these data indicate that many Up genes harbor bivalent promoters and lie close to H3K9me3 and ERVs (Fig. 1F).

Trim28 deletion triggers a switch from repressive to active chromatin marks at ERVs

Mapping the genomic location of specific TRIM28-regulated ERVs based on a TRIM28 ChIP-seq is problematic because of the sharpness of the corresponding peaks, which only rarely extend beyond the borders of these multicopy elements. We thus turned to a comparison of H3K9me3 peaks in wild-type and Trim28-deleted ES cells, since this histone modification can spread a few kilobases into the junction of ERV proviruses with their flanking regions (Karimi et al. 2011; Rebollo et al. 2011). We found around 19,000 H3K9me3 peaks, that is, about half of those detected in control ES cells, to be TRIM28 dependent as indicated by their absence in knock-out cells (Fig. 2A, left). In agreement with their noted proximity to ERVs (see Fig. 1E), Up genes lay closer to TRIM28-dependent H3K9me3 peaks than Down and Stable genes (Fig. 2A, right). Likewise, in an element-centric analysis, we used the TRIM28-dependent H3K9me3 peaks to determine the nearest gene, generating a list significantly enriched for up-regulated genes (giving 2220 Up genes, Fisher’s exact test, P > 2.2 × 10−10) (Supplemental Fig. S3A; Supplemental Table 3), in line with the gene-centric analysis above. Of note, upon further examination of the high number of H3K9me3 peaks “newly present” in Trim28 knock-out cells, we found them to be in the same locations as the WT peaks but just slightly displaced and smaller in height and diameter rather than representing new peaks (Fig. 2A). These peaks thus most likely represent remnants of TRIM28-specific peaks, which is not surprising considering that our analyses

Figure 2. Trim28 deletion triggers a switch from repressive to active chromatin marks at ERVs. (A) Venn diagram of H3K9me3 ChIP-seq peaks in WT versus KO ES cells (left). 19,057 peaks are present in WT but lost in KO cells and so are defined as TRIM28-dependent peaks, which cluster closer to Up genes than Down (P = 0.001418) and Stable (P = 2.2 × 10−10) genes (right). (B) TRIM28-dependent H3K9me3 peaks (see above) were assessed for correlation with ChIP-seq data sets. Positive correlations are shown on the left graph and anti-correlations on the right. All data displayed after global normalization of ChIP-seq counts. (C) ChIP results for repressive (left panel) and active (right panel) marks present at global IAPs (using IAP 5′-UTR primers). Bars show the mean and SD of three to four ChIPs per antibody with immunoprecipitate values normalized to total inputs (IP/TI) relative to Gapdh. Negative controls of no antibody were used in all experiments giving no enrichments, while the Pou5F1 enhancer served as a positive control with high enrichments for both H3K27ac and H3K4me1 of 1.1 and 7.5, respectively. Results were also reproduced in an independent ES cell line (Row et al.). Paired t-tests were used to compare WT and TRIM28-depleted samples for each antibody: H3K9me3, P = 0.014; TRIM28, P = 0.027; SETDb1, P = 0.036; H4K20me3, P = 0.0308; H3ac, P = 0.0337; H3K27ac, P = 0.0001; H3K4me1, P = 0.011.
were performed only 4 d after inducing Trim28 excision to avoid lethality.

Interestingly, we observed that the Trim28-dependent H3K9me3 peaks not only correlated with repressive histone marks, Trim28, SETDB1 peaks (the latter data set obtained from Bilodeau et al. 2009), and with ERVs, but anti-correlated with H3K4me1 and H3K27ac, marks typically found together on active enhancers (Creyghton et al. 2010; Rada-Iglesias et al. 2010; Shen et al. 2012), while displaying no particular association with H3K4me3 or H3K27me3 (Fig. 2B; data not shown). In line with this, Up genes themselves also lay far from enhancer marks (Supplemental Fig. S3B). We therefore hypothesized that ERVs may gain these marks upon Trim28 deletion, thereby enhancing expression of neighboring genes. To test this idea, we focused on IAPs since we identified a motif highly represented in our H3K9me3 ChIP-seq peaks (in 64% of peaks) normally present in IAP consensus sequences (Supplemental Fig. S3C,D). Supporting this model, ChIP-qPCR with primers designed to amplify the majority of IAPs revealed that, indeed, in Trim28 knock-out ES cells, these elements not only lost Trim28, SETDB1, and the repressive marks H3K9me3 and H4K20me3, but also gained active marks, including H3K27ac and H3K4me1 (Fig. 2C). This observation fits with the recent detection of H3K9me3 at poised enhancers (Zentner et al. 2011), and indicates that loss of this mark upon Trim28 deletion may be sufficient to activate such regulatory elements, notably those located within IAPs and likely other ERVs. The derepression of cryptic enhancers within ERVs thus appears to be one prominent mechanism in the transcriptional deregulation triggered by Trim28 deletion in ES cells.

**Figure 3.** Expression and cytosine methylation at the Zfp575 gene and adjacent IAP. (A) Map (drawn to scale) of the Zfp575 gene that just overlaps a full-length IAP (named IAP575 and of the IAPe2 type) with both gene and IAP in the same orientation (sizes of each are stated). (LTR) Long terminal repeat; (PB5) primer binding site; (Gag) group-specific antigen; (Pro) protease; (Pol) polymerase. (B) Trim28 knock-out and knockdown (comparing control, shEmpty and KD, shTRIM28) cell lines were assessed for their expression of Zfp575 (left panel) using two different primer sets, or TRIM28 or IAPs as controls (right panel). Unpaired t-tests were used to compare controls with TRIM28-depleted samples for all ES and EC cell lines: Zfp575, P = 0.0015; IAP, P = 0.0344; Trim28, P = 0.0008. Since Zfp575 is normally expressed specifically in brain, we also verified it to be expressed in primary neurospheres and brain (data not shown). (C) Quantitative pyrosequencing was used to measure DNA methylation levels at the Zfp575 promoter versus the flanking S’-LTR IAP575 promoter (left panel). Control primers were specific for the Pou5f1 promoter or global LINE1s or global IAP LTRs (IAPs). Bars represent means over multiple Cpg positions with error bars showing the SD across all Cpgs. (Right panel) Samples were compared (across six Cpg positions) for their methylation levels at the IAP575 promoter. Primordial germ cells were also used to show that IAP575 is demethylated in germ cells to a level not much lower than in Trim28-deleted ES cells (e.g., to an average of 69% instead of 76%) (data not shown). Two-tailed paired t-tests display all significant differences: Trim28 WT versus KO ES, P = 0.0088; Ehmt2 WT versus KO, P = 0.0001.

**Activation of specific ERV-based enhancers upon loss of Trim28 leads to activation of nearby genes**

To explore the molecular mechanism of this process further, we examined transcription and chromatin state at specific ERV–Up gene pairs. We first focused on an element that was 90% identical to IAP sequences previously found to be Trim28 regulated (Rowe et al. 2010) and named this ERV IAP575 because of its position 3’ to the bivalent gene Zfp575 (Mikkelsen et al. 2007; Bilodeau et al. 2009) in the sense orientation (Fig. 3A). Zfp575 was markedly up-regulated in Trim28-depleted ES cells but not MEFs, consistent with our mRNA-seq data, paralleling the modulation of IAPs in TRIM28-depleted ES cells but not MEFs, consistent with our mRNA-seq data, paralleling the modulation of IAPs in TRIM28-depleted ES cells but not MEFs, consistent with our mRNA-seq data, paralleling the modulation of IAPs in TRIM28-depleted ES cells but not MEFs, consistent with our mRNA-seq data, paralleling the modulation of IAPs in TRIM28-depleted ES cells but not MEFs, consistent with our mRNA-seq data, paralleling the modulation of IAPs in TRIM28-depleted ES cells but not MEFs, consistent with our mRNA-seq data, paralleling the modulation of IAPs in TRIM28-depleted ES cells but not MEFs, consistent with our mRNA-seq data, paralleling the modulation of IAPs in TRIM28-depleted ES cells but not MEFs, consistent with our mRNA-seq data, paralleling the modulation of IAPs in TRIM28-depleted ES cells but not MEFs, consistent with our mRNA-seq data, paralleling the modulation of IAPs in TRIM28-depleted ES cells but not MEFs, consistent with our mRNA-seq data, paralleling the modulation of IAPs in TRIM28-depleted ES cells but not MEFs, consistent with our mRNA-seq data, paralleling the modulation of IAPs in TRIM28-depleted ES cells but not MEFs, consistent with our mRNA-seq data, paralleling the modulation of IAPs in TRIM28-depleted ES cells but not MEFs, consistent with our mRNA-seq data, paralleling the modulation of IAPs in TRIM28-depleted ES cells but not MEFs, consistent with our mRNA-seq data, paralleling the modification of IAPs in these targets (Figs. 3B, 1A). Similar to its Pou5f1 counterpart, the Zfp575 promoter was unmethylated in ES cells. In contrast, the IAP575 LTR displayed high rates of Cpg methylation, as did the IAP family as a whole, and to a lesser extent LINEs (Fig. 3C, left). The failure of DNA methylation to extend from the IAP575 LTR to the promoter of the adjacent Zfp575 gene fits with recent observations that (1) DNA methylation only spreads a few kilobases from Trim28 binding sites (Quenneville et al. 2012; Rowe et al. 2013), and (2) ERV methylation rarely affects flanking regions (Rebollo et al. 2011). Interestingly, while methylation of the IAP575 LTR was unaltered by Trim28 deletion in MEFs, it significantly decreased in their ES cell counterparts, albeit not as dramatically as in ES cells deleted for Ehmt2 (G9a), a histone methyltransferase involved in the maintenance of DNA methylation (Fig. 3C, right; Dong et al. 2008; Tachibana et al. 2008). Perhaps explaining this latter difference, Trim28 loss is lethal after a few days in ES cells (Rowe et al. 2010), while EHMT2-depleted cells can be stably maintained for many passages, allowing for extensive loss of cytosine methylation through multiple rounds of DNA replication. However, since this only modest decrease in DNA methylation was observed in parallel to the striking up-regulation of genes, it is possible that it contributes to this phenotype.

We then mapped histone marks across the Zfp575/IAP575 locus (Fig. 4). Trim28, SETDB1, H3K9me3, and H4K20me3 were
markedly enriched at IAP575, yet did not spread back to the zfp575 promoter. Upon Trim28 deletion, these repressive histone modifications collectively decreased, to be replaced by the active marks H3K4me1, H3K27ac, and H3Ac over the whole locus, albeit in the most pronounced fashion over its IAP575 part (Fig. 4B–D). We then further validated the up-regulation of several other ERV–Up gene
pairs and verified that at these loci, TRIM28-dependent H3K9me3 is substituted by the active mark H3K27ac, as documented by ChIP-seq (Supplemental Figs. S4–S6), in support of our model.

**ERV sequences that escape TRIM28-mediated repression can act as activators during embryogenesis**

These results indicate that some ERVs carry intrinsic enhancer sequences that are silenced at the ES cell stage via TRIM28-induced repression. To probe this model further, we tested previously identified TRIM28-sensitive and TRIM28-resistant IAP sequences (Rowe et al. 2010) for their ability to modulate a nearby cellular promoter during embryonic development. To this end, we placed these elements in the antisense direction upstream of a phosphoglycerate kinase (PGK) promoter because at baseline this promoter drives only weak expression of GFP in embryos. We then used these lentiviral vectors for transgenesis via transduction of fertilized murine oocytes. Examination of the resulting embryos at E13 revealed that, while a TRIM28-sensitive IAP-derived sequence (IAP4) was able to limit expression from the PGK promoter contained in the lentiviral provirus, its TRIM28-resistant counterpart (IAP1, ~87% identical) (see Rowe et al. 2010), in contrast, enhanced GFP expression (Fig. 5). Thus, TRIM28 susceptibility can condition the cis-acting transcriptional impact of specific ERV sequences in vivo during embryonic development.

![Figure 5. ERV sequences that escape TRIM28-mediated repression can act as activators during embryogenesis.](image)

Figure 5. ERV sequences that escape TRIM28-mediated repression can act as activators during embryogenesis. Lentiviral transgenesis was performed with an empty PGK-GFP vector (PGK-GFP control, upper panels), or with the same vector including either an IAP4 (TRIM28-sensitive IAP-PGK-GFP, middle panels) or an IAP1 (TRIM28-resistant IAP-PGK-GFP, lower panels) sequence cloned antisense upstream of the PGK promoter. At E13, embryos were scored for GFP expression and vector copy numbers. For the PGK-GFP control, 13/29 embryos were green. For the TRIM28-sensitive IAP-PGK-GFP, 4/19 embryos were green (all with copy numbers above 16), and 4/19 pale green (including numbers 3 and 4 in this figure). For the TRIM28-resistant IAP-PGK-GFP, 12/17 embryos were green (including one with a copy number above 10), and 2/17 pale green (with copy numbers of 0.95 and 0.89). Embryos with similar copy numbers per vector group are shown in each column with increasing copy numbers by row. Vectors were injected twice with similar results. In one experiment, MEFs were derived from embryos to verify that microscopy differences were reproduced by flow cytometry (data not shown).

**Discussion**

The present work unveils a fundamental aspect of transcriptional regulation during the early embryogenesis of higher vertebrates. At the heart of this system lies, on one side, retroelements that have colonized eukaryotic genomes from the earliest times, and on the other side, the tetrapod-specific KRAB-ZFP gene family (Urrutia 2003; Huntley et al. 2006; Emerson and Thomas 2009; Wolf and Goff 2009; Thomas and Schneider 2011), which acts as the targeting machinery for TRIM28. We previously demonstrated that TRIM28 is responsible for the silencing of ERVs in ES cells and early embryos (Rowe et al. 2010). Here, we reveal that an important role of this process is to protect the transcriptional dynamics of early embryos from perturbation by cis-acting activators contained in these mobile elements.

For this, we deleted Trim28 in ES cells and monitored chromatin signatures at deregulated genes and ERVs. We found that half of the ~5700 transcriptional units up-regulated upon Trim28 deletion in ES cells bore, at baseline, the bivalent histone marks H3K4me3 and H3K27me3 characteristic of genes poised for transcription (Bernstein et al. 2006). Moreover, we noted that, remarkably, these genes were on average located closer to ERVs than genes down-regulated or unaffected following TRIM28 removal. We then further observed that, while in wild-type ES cells, ERVs bound TRIM28 and SETDB1 and accordingly were enriched in H3K9me3 and H4K20me3, they lost these repressive marks upon Trim28 deletion and instead acquired chromatin modifications typically associated with active enhancers such as H3K4me1 and H3K27ac, a phenomenon that was documented both at global IAPs and at the level of specific ERV-up-regulated gene loci. Finally, we could demonstrate that ERV-derived sequences could either repress or activate an adjacent cellular promoter in transgenic mouse embryos, depending on whether they were recognized or not by a TRIM28-containing complex in ES cells.

The model emerging from our study (Fig. 6) is one whereby, in ES cells, the recruitment of TRIM28 and its partners, including SETDB1, at ERV-contained enhancers leads to the maintenance of H3K9me3, H4K20me3, and DNA methylation, which prevents the untimely activation of nearby genes, in particular, those harboring bivalent promoters. Indeed, DNA methylation is known to anticorrelate with active marks (Okitsu and Hsieh 2007; Ooi et al. 2007; Weber et al. 2007; Stadler et al. 2011), and SETDB1 has previously been shown to maintain H3K9 trimethylation and, secondarily, the H3K4me3/H4K20me3 at ERVs (Matsui et al. 2010). Inactivation of this machinery leads not only to the loss of silent histone marks and to a mild decrease in cytosine methylation but also to the acquisition of active enhancer marks at these loci, which tilts nearby genes, notably those poised for transcription, toward expression. Noteworthy, the NuRD complex, also recruited by TRIM28, is known to mediate deacetylation of H3K27 through its HDAC1 and HDAC2 subunits (Reynolds et al. 2011), which would explain the genome-wide anti-correlation observed between H3K27ac and TRIM28 target sites at baseline. Likewise, LSD1, which shares at least some targets with TRIM28 and NuRD (MacFarlan et al. 2011, 2012), is able to demethylate and therefore demethylate the active mark H3K4me1 (Whyte et al. 2012). Accordingly, disruption of either SETDB1 or LSD1 leads to effects on cellular transcripts (Biloudeau et al. 2009; Yuan et al. 2009; Karimi et al. 2011; MacFarlan et al. 2011, 2012). In the case of SETDB1 deletion, this includes the induction of chimeric transcripts initiating from derepressed ERVs, which we also see evidence for here, since some of the same transcripts are induced (Karimi et al. 2011;
this study). Here we demonstrate that in the absence of TRIM28, retrotransposon-based enhancers become active.

The heterogeneity of the TRIM28-recruiting ERV loci uncovered here, with sequences intrinsic to IAP, MERVL, and ERVK families, suggests that a large number of different KRAB-ZFPs engage in directing TRIM28 to ERVs in ES cells. Additionally, TRIM28 can also interact with KRAB-O proteins that lack zinc fingers but bridge DNA through other factors such as SRY (Peng et al. 2009). Remarkably, TRIM28 and some KRAB-ZFPs are also detected in adult tissues, albeit along exquisitely cell- and stage-specific fashions, where they have become coopted to influence tissue-specific gene regulation (Jakobsson et al. 2008; Bojkowska et al. 2012; Chikuma et al. 2012; Krebs et al. 2012; Santoni de Sio et al. 2012a,b). Whether some ERV-derived enhancers serve as docking sites for this repressor system in these adult tissues warrants exploration. There is evidence that some ERV sequences function as authentic regulators, including enhancers, in certain cells, not only during development but also in adult tissues (Pi et al. 2004; Bourque et al. 2008; Kunarro et al. 2010; Teng et al. 2011; Mey et al. 2012; Schmidt et al. 2012). Our data indicate that these rare coopted elements represent only exceptions within a large group, most members of which are repressed through TRIM28. This may explain why most KRAB-ZFP genes are expressed in both mouse and human ES cells, while at least in this latter species, most if not all endogenous retroviruses have accumulated mutations that would anyway preclude their retrotransposition. The need to preserve the transcription dynamics of ES cells, rather than to protect the genome from further spread of these elements, is likely what constitutes the strongest selective pressure on the KRAB/TRIM28 system in higher species.

**Cell culture**

ES cells were cultured in standard conditions as described (Rowe et al. 2013). The ES cell lines used were two Trim28loxP/loxP lines called ES3 and ES6 and their derived Trim28-conditional knock-out cell lines that are transduced with a tamoxifen (4-OHT)-inducible Cre vector (Rowe et al. 2010). For analysis of expression and chromatin marks, knock-out cells were collected 4 d after treatment with 4-OHT (used overnight at 1 μM, Sigma-Aldrich: H7904) due to the lethality of Trim28 knock-out for longer time periods. Rex1GFP ES cells (Wray et al. 2011) were additionally used where stated (kind gift from A.G. Smith, University of Cambridge, UK) or Elmi2 parental or stable knock-out ES cells (Dong et al. 2008; Tachibana et al. 2008) (a kind gift from Yoichi Shinkai, RIKEN Institute, Japan). Trim28-knockdown was induced with shRNA vectors (see above), and cells selected with puromycin 2 d post-transduction and collected 4 d post-puromycin selection, a time point giving similar expression changes to 4 d post-knock-out. Knockdown efficiency was verified by qRT-PCR. Trim28loxP/loxP 4-OHT-inducible MEFs were used to delete Trim28, while Trim28 knockdowns were also performed in MEFS and P9 EC cells where stated.

**Flow cytometry**

Vector titers and GFP repression were measured by FACS, as well as the differentiation status of ES cells as monitored by staining with an SSEA-1 PE-conjugated antibody or isotype control (BD Pharmingen: 560142 and 555584).

**RNA extraction and quantification**

Total RNA was extracted with TRIzol (Invitrogen: 15596-018), purified using a PureLink RNA kit (Ambion: 12183018A), treated with DNase (Ambion: AM1907) and 500 ng reverse-transcribed using random primers and SuperScript II (Invitrogen: 18064-022). Primers (see Supplemental Table 4) were designed for an Applied Biosystems 7900HT machine using Primer Express (Applied Biosystems) and used for SYBR Green qPCR. Primer specificity was confirmed by dissociation curves and samples were normalized to Gapdh, although Actin gave similar results.

**mRNA sequencing**

Total RNA (10 μg) from Trim28 WT and KO ES cells and MEFs was subject to mRNA selection, fragmentation, cDNA synthesis, and library preparation for Illumina high-throughput sequencing, after checking RNA quality on a Bioanalyzer. Single read sequencing was performed on a Genome Analyzer Ix machine with 40 cycles generating ~33 million reads per sample. Additionally, mRNA sequencing was performed on Trim28 control (shEmpty) and knock-down (shTrim28) Rex1 ES cells with 50 cycles on an Illumina HiSeq 2000 machine generating around 200 million reads per sample and confirming our knock-out ES cell results.

**Chromatin immunoprecipitation (ChiP)**

ES cell samples were washed twice (in PBS + 2% FCS), counted to normalize by cell number, cross-linked (10 min rotation in 1% formaldehyde), quenched with glycine (at 125 mM on ice), washed three times (PBS), and pelleted at 10^7 cells per Eppendorf. Pellets were lysed, resuspended in 1 mL of sonication buffer on ice (10 mM Tris at pH 8, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% NaDOC, 0.25% NLS, and protease inhibitors), transferred to glass 12 x 24-mm tubes (Covaris: S2005), and sonicated (Covaris settings: 20% duty cycle, intensity 5, 200 cyles/burst, 30 min).

**Lentiviral vectors**

For in vivo experiments, the transfer vector pRRLSIN.cPPT.PGK-GFP.WPRE (available from Addgene) was used with either IAP1 or IAP4 sequences (Rowe et al. 2010) included upstream of the PGK (phosphoglycerate kinase-1) promoter in the antisense orientation (Rowe et al. 2013). For Trim28 knockdown experiments, shRNA lentiviral plasmids (against mouse Trim28 or the empty vector control) were ordered from Sigma-Aldrich (pLKO.1-puro). All vectors were produced by transient transfection of 293T cells with the transfer vector, packaging, and VSVG envelope plasmids (Barde et al. 2010) and titrated on 3T3 fibroblasts.
Sonication was then assessed by reverse cross-linking overnight in the presence of proteinase K and RNase, followed by DNA extraction and quantification on a Bioanalyzer (Agilent 2100 machine). Fragment sizes were equivalent between wild-type and knock-out samples, which were done in parallel (with mean fragment sizes of ~200 bp for Experiment 1 and ~400 bp for Experiments 2 and 3). Samples were also checked for the absence of single-stranded DNA by Exonuclease 1 treatment. Immunoprecipitations were performed in duplicates or triplicates with Dynabeads (100.03D) using 1 x 10^6 to 2 x 10^6 cells, 80 μL of pre-blocked beads, and 5 μg of antibody (or no antibody as a control) per sample in IP buffer (167 mM NaCl, 16.7 mM Tris at pH 8.1, 1.2 mM EDTA, 0.5 mM EGTA, 1.1% Triton X-100, and protease inhibitors) overnight. After washing and reverse cross-linking (also overnight) and DNA extraction, results were quantified by SYBR Green qPCR (for primers, see Supplemental Table 4). The antibodies used were TRIM28 (Tronolah, rabbit polyclonal SY 3267-68, 30–50 μL per sample), H3K9me3 (Abcam: ab8898), SETDB1 (Santa Cruz, 50 μL per sample), H4K20me3 (Millipore: 07-463), H3ac (Millipore: 06-599), H3K27ac (Abcam: ab4729), and H3K4me1 (Abcam: ab8895).

ChIP sequencing

Total input (TI) and corresponding immunoprecipitated (IP) ChIP libraries were prepared using 10 ng of material with gel selection of 200-bp- to 300-bp-sized fragments. Libraries were ligated with Illumina adaptors and paired-end sequenced (or single-end for H3K27ac) on an Illumina HiSeq 2000 machine with 50–100 cycles and two samples multiplexed in one lane, generating ~100 million sequences per sample. TI samples gave background enrichment patterns distinct from IPs.

Quantitative bisulfite pyrosequencing

Genomic DNA was converted (200 ng/sample) and used for PCR and pyrosequencing as previously described (Rowe et al. 2013). We thank A. Reymond (CIG, UNIL, Lausanne) for kind use of the pyrosequencer. Results were analyzed using Pyro Q-CpG Software.

Lentiviral transgenesis

Lentiviral vectors for transgenesis were prepared using Episfer medium (Invitrogen: 10732022), the particle concentration obtained by p24 ELISA (PerkinElmer: NEK050B001KT), and the infectious titer determined on HCT116 cells by GFP flow cytometry. Ratios for the three vectors were between 1/319 and 1/428 of infectious to physical particles with titers between 2 and 2.4 x 10^9 infectious units/mL. Transgenesis was performed by perivitelline injection of vectors into fertilized oocytes that were transferred to foster mothers (strain B6D2F1/J) and then recovered at embryonic day 13 (E13). Photographs were taken using the same saturation, gain, and exposure settings and image settings for all embryos.

Bioinformatics analyses and statistics

mRNA-seq analysis

Reads were mapped to the mouse genome mm9 using the short read aligner program Bowtie (Langmead et al. 2009) with reads (three mismatches allowed) excluded that mapped more than five times. The SAMtools and bedtools suites (Li et al. 2009; Quinlan and Hall 2010) were used to generate files to be visualized on the UCSC Genome Browser (http://genome.ucsc.edu/) (Kent et al. 2002).

MA plots

MA plots were generated from rpkm values (number of reads normalized by gene length and total reads) using the maplot Python package (https://github.com/delafont/maplot).

Boxplots

Boxplots showing bootstrapped values (generated using R: http://www.R-project.org) were used in gene-centric analyses to determine if up-regulated (Up) genes were closer to the indicated histone marks/ERVs compared with two control gene groups (down-regulated, “Down” or unaffected, “Stable” genes). Statistical significance was calculated using the Wilcoxon rank-sum test.

H3K9me3 ChIP-seq analysis

Paired-end reads were mapped to the mouse genome (three mismatches allowed) mm9 using the short read aligner program Bowtie (Langmead et al. 2009). Several analyses were performed, showing the same global results where reads were either excluded if mapping more than one time, five times, or 20 times to the genome. Peaks were called from the data where reads were mapped with a cutoff of 20 to allow more coverage of repeats, although individual peaks of interest were validated using the analysis where a cutoff of one was used (in this case, only exact matches were allowed). Enriched regions were defined using the ChIP-Part analysis module from the ChIP-seq analysis suite (http://ccg.vital-it.ch/chipseq/). H3K27ac ChIP-seq data were confirmed to correlate (by 53%) with previous H3K27ac ChIP-seq in ES cells (Creyghton et al. 2010) and verified to be normally present at active genes and gained at specific ERV loci (see Supplemental Figs. S5, S6). TRIM28 ChIP-seq peaks were defined using MACS (default threshold P-value < 1 x 10^-5) and normalized to the total input generating 3099 peaks. Direct binding sites to promoters of up-regulated genes were identified using a cutoff of ±2 kb from the TSS giving 49 genes, 13 of which were excluded due to the binding being through an ERV.

Public ChIP-seq data

Raw or already mapped reads were downloaded from publicly available ChIP-seq data (GEO IDs: GSE12241, GSE18371, and GSE24165) and peaks called using MACS. ChIP-correlation analyses were performed with bed files, using the online tool ChIP-Cor (http://ccg.vital-it.ch/chipseq/chip_cor.php). Histograms were analyzed using raw counts and count densities, and those showing a correlation were displayed after global normalization, where ChIP-seq counts are normalized by the total number of counts and the window width to allow visualization of multiple data sets on the same plot.

Motif identification

The MotifRegressor and motifsComparator softwares were used to identify DNA sequence binding motifs (Conlon et al. 2003; Carat et al. 2010).

Other statistical analyses

GraphPad Prism version 4.00 (http://www.graphpad.com) was used for other statistical analyses, where control and knock-out groups were compared with paired or unpaired t-tests (as noted) that were one-tailed except where stated as two-tailed.

Data access

All next-generation sequencing data have been submitted to the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) and are accessible with the accession no. GSE41903.
Acknowledgments

We thank P.V. Maillard and J. Marquis for advice, S. Verp and S. Öffner for technical assistance through the EPFL Lentiviral Transgenesis platform, B. Khubieh and J. Rougemont for bioinformatics help (EPFL Biostatistics and Bioinformatics core facility), Y. Shinkai and D. Schübeler for the Ehm2 knockout and parental ES cells, and K. Harshman and A. Reymond (Center for Integrative Genomics, University of Lausanne) for high-throughput sequencing or use of a pyrosequencer, respectively. All computing for high-throughput sequencing was done on the Vital-it cluster. This work was supported through grants from the Swiss National Science Foundation and the European Research Council to D.T. and an NIH grant to S.L.P.

Author contributions: H.M.R. conceived the study, designed and performed the experiments, analyzed the data, and wrote the manuscript. A.K. performed bioinformatic analyses. A.C., L.E., T.S.M., and Y.T. designed experiments. J.J., S.V., and S.L.P. designed experiments. D.T. conceived the study, designed experiments, and wrote the manuscript.

References


TRIM28 silences ERV-based enhancers in ES cells


Received August 10, 2012; accepted in revised form December 6, 2012.