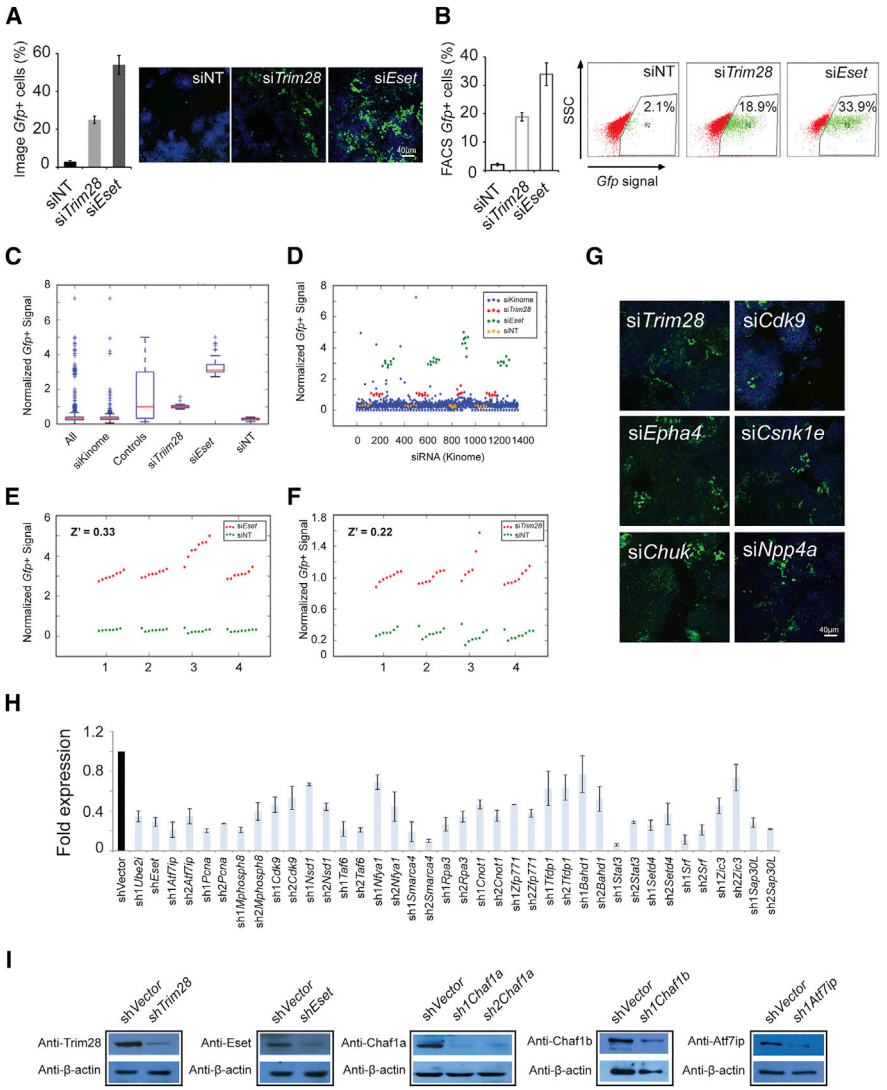


Figure S1. Optimization and Validation of siRNA Library Screening, Related to Figure 1

(A) siRNA knockdown of *Trim28* and *Eset* induced *Gfp* expression in F9 cells infected with the MMLV-*Gfp* reporter as detected by fluorescence imaging. *Gfp* signal was tested by IXU image system and quantified by the MetaXpress software. Left panel, the percentage of *Gfp* positive cells is shown on the y axis. Values are presented as mean ± s.e.m from independent replicate experiments. Right panel, representative images of *Gfp* fluorescence (green) and nuclear Hoechst 33342 staining (blue) are shown for treatments with siNT, si*Trim28* and si*Eset*.

(B) Reactivation of *Gfp* expression in F9 cells by *Trim28* and *Eset* RNAi as analyzed by FACS. The FACS data corroborated with the measurements from fluorescence imaging (A). Left panel, the percentage of *Gfp* positive cells is shown on y axis. Values are presented as mean ± s.e.m from independent replicate experiments. Right panel, representative dot plots of treatments with siNT, si*Trim28* and si*Eset*.

(C) Box plot demonstrating the normalized *Gfp* signal obtained from indicated cluster of wells. Normalized *Gfp* signal is shown on the y axis. The red horizontal line in each box corresponds to the sample mean. The kinome siRNA library screen was conducted in duplicates.

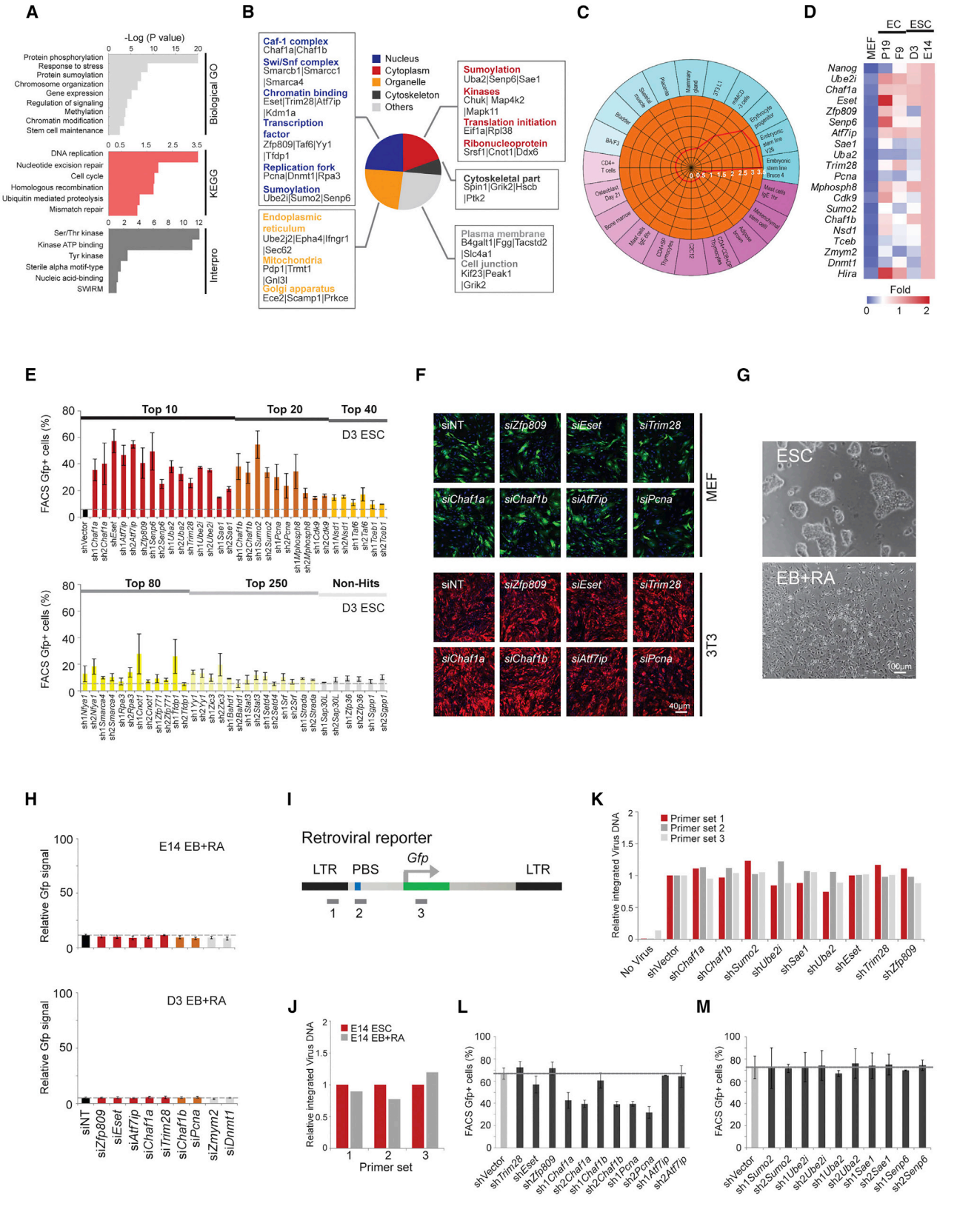
(D) Scatter plot for the kinome siRNA screen. Normalized *Gfp* signal is shown on the y axis. Non-targeting controls are shown as yellow dots. siRNA targeting *Eset* and *Trim28* were used as positive controls and are shown as green and red dots respectively. siRNA from the kinome library are shown as blue dots.

(E and F) Z-prime of *Gfp* activation by *Eset* (E) or *Trim28* (F) knockdown. Normalized *Gfp* signal is shown on the y axis. Kinome siRNA library plate serial number is shown on the x axis. Non-targeting siRNA are shown as green dots, siRNA targeting on *Eset* or *Trim28* are shown as red dots. Z-prime was calculated by ScreenSifter software using the formula: Z-prime = $1-3(\sigma_{p+}\sigma_n)/|\mu_p-\mu_n|$.

(G) Representative images of *Gfp* activation by knockdown of indicated hits from the kinome siRNA library screen. *Gfp* is shown as green, and Hoechst 33342 staining of the nucleus is shown as blue.

(H) Real-time qPCR analysis of shRNA knockdown efficiency in mouse ES E14 cells. The relative expression levels of each gene were normalized to gene expression levels in the shVector control. Values are mean ± s.e.m from independent replicate experiments.

(I) Western blot assay of shRNA knockdown efficiency in E14 cells. The cells were transfected with shRNA targeting *Trim28*, *Eset*, *Chaf1a*, *Chaf1b*, or *Atf7ip*. Changes in the levels of protein were detected by Western blot using antibodies specific to the indicated proteins.



(legend on next page)

Figure S2. GO Analysis and Specificity of the Candidate Factors in ESC Proviral Silencing, Related to Figure 2

(A) GO analysis of the 650 candidate genes identified from the genome-wide siRNA screen. The functional categories plotted are Biological Process (upper panel), KEGG pathway (medium panel), and Interpro (bottom panel). GO terms with *p*-value > 0.1 and with less than 3 genes are filtered out.

(B) Cellular component analysis. The significant hits (*Gfp* signal > 0.45 threshold) were subjected to cellular components analysis using BiNGO, to identify the complexes formed by these hits and their locations in the cell. Examples of these complexes are indicated.

(C) Cell type enrichment analysis. Significant hits from the siRNA screen were subjected to cTen software analysis to identify the cell types in which they are expressed. This analysis indicated a specific and high enrichment of these candidate hits in embryonic stem cells. The enrichment score represents the –log10 of the Benjamini-Hochberg adjusted *p*-values across all the cell types present in the database.

(D) Heat map demonstrating the expression profile of the selected candidate genes in embryonal versus differentiated cells. Total RNA was extracted from mouse embryonic stem cell line E14 and D3, mouse embryonal carcinoma cell line F9 and P19, and mouse embryonic fibroblast (MEF) cells. The relative expression of candidate genes was measured by real-time qPCR and normalized to gene expression levels from E14 cells. Values are mean ± s.e.m from independent replicate experiments.

(E) Validation of the genome-wide siRNA screen using shRNA knockdown of indicated genes in mouse embryonic stem cell D3. Two specific shRNA were designed for each candidate gene. *Gfp* signal was detected by FACS. The percentage of *Gfp* activation is shown on the y axis. Values are mean ± s.e.m from independent replicate experiments.

(F) Representative images of *Gfp* or *mCherry* signals for indicated hits knockdown in MEF and 3T3 cells. Cells were infected with the corresponding MMLV reporter virus, and conducted with siRNA knockdown. *Gfp* (green), *mCherry* (red) and nucleus Hoechst 33342 (blue) are shown.

(G and H) MMLV-*Gfp* rescue by siRNA knockdown of indicated genes in embryoid body (EB) cells differentiated from E14 or D3 ESCs. E14 or D3 cells were differentiated via EB formation+Retinoic acid (RA). (G) Microscopic images for the morphology of E14 cells (upper panel) and E14-derived EB treated with RA (bottom panel). (H) MMLV-*Gfp* rescue in differentiated EB cells. *Gfp* signal was tested by IXU image system. Relative *Gfp* signal is shown on the y axis. Values are mean ± s.e.m from independent replicate experiments.

(I) Schematic drawing of the locations of the primer pairs used for MMLV-*Gfp* integration test.

(J and K) MMLV-*Gfp* integration assay. (J) E14 cells and the EB cells differentiated from E14 were infected with MMLV-Gfp retrovirus. (K) E14 cells were infected with MMLV-Gfp virus followed by shRNA knockdown of indicated genes. Genomic DNA was purified and qPCR was performed using the primer pairs targeting different reagents of the MMLV-*Gfp* reporter. The relative integration level of the virus was normalized to endogenous single copy gene m36B4, and further normalized to the corresponding cells without viral infection. Values are mean ± s.e.m from independent replicate experiments.

(L and M) *Gfp* expression of PiggyBac-CAG-*Gfp* reporter in F9 cells upon the knockdown of indicated candidate genes. Transposase plasmid was co-transfected with PiggyBac-CAG-*Gfp* into F9 cells to help PiggyBac-CAG-*Gfp* integrate into genomic DNA. The indicated candidate genes were depleted with shRNA. *Gfp* was tested using FASC. Grey horizontal lines are added to indicate baseline.

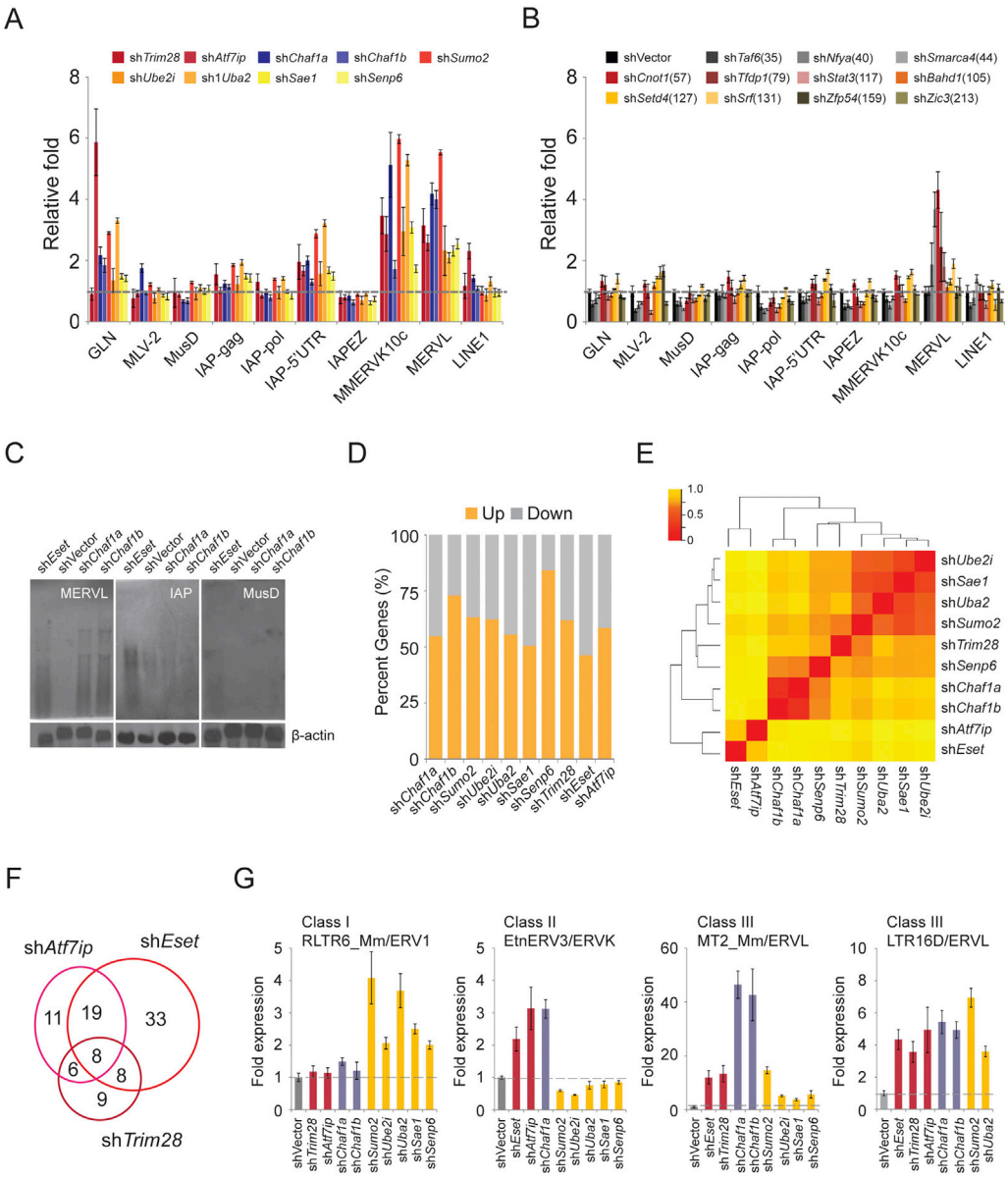


Figure S3. Caf-1 and Sumoylation Factors Play Roles in Regulating the Expression of ERVs, Related to Figure 3

(A and B) De-repression of ERVs in E14 ESCs with shRNA knockdown of indicated genes. The relative expression of several families of ERVs was determined by real-time qPCR. Levels of each ERV were normalized to the shVector control which is set as 1 (dotted line). Values are mean ± s.e.m from independent replicate experiments.

(C) Northern blot assay of ERV MERVL, IAP and MusD expression in *Chaf1a*, *Chaf1b* and *Eset* depleted E14 cells (upper panels). Total RNA was isolated from E14 cells after shRNA knockdown. β-actin probe was used as the loading control (lower panels).

(D) Percentage stacked columns indicating the regulatory effects upon the depletion of the indicated factors.

(E) Clustering analysis. The RNA-Seq libraries were clustered using R, based on differential gene expression. A hierarchical agglomerative was used to transform the different gene expression datasets into a similarity matrix. The correlation heat map was generated using gplots. The rows and columns represent different candidates. The color intensity signifies the correlation strength between 0 (red - high similarity) to 1.0 (yellow - high difference).

(F) Venn diagram showing the number of common and differential upregulated ERVs upon the knockdown of *Trim28, Eset* and *Atf7ip*. High correlation of regulated ERVs targets is observed.

(G) Validation of ERVs regulation identified from the RNA-seq data. Knockdown of the indicated genes were conducted in E14 cells. Relative ERV expression was analyzed by real-time qPCR using primers specific to class I LTR6_Mm/ERV1, class II EtnERV3-int/ERVK and class III MT2_Mm/ERVL and LTR16D/ERVL. Levels of each ERV were normalized to the shVector control which is set as 1 (dotted line). Values are mean ± s.e.m from independent replicate experiments.

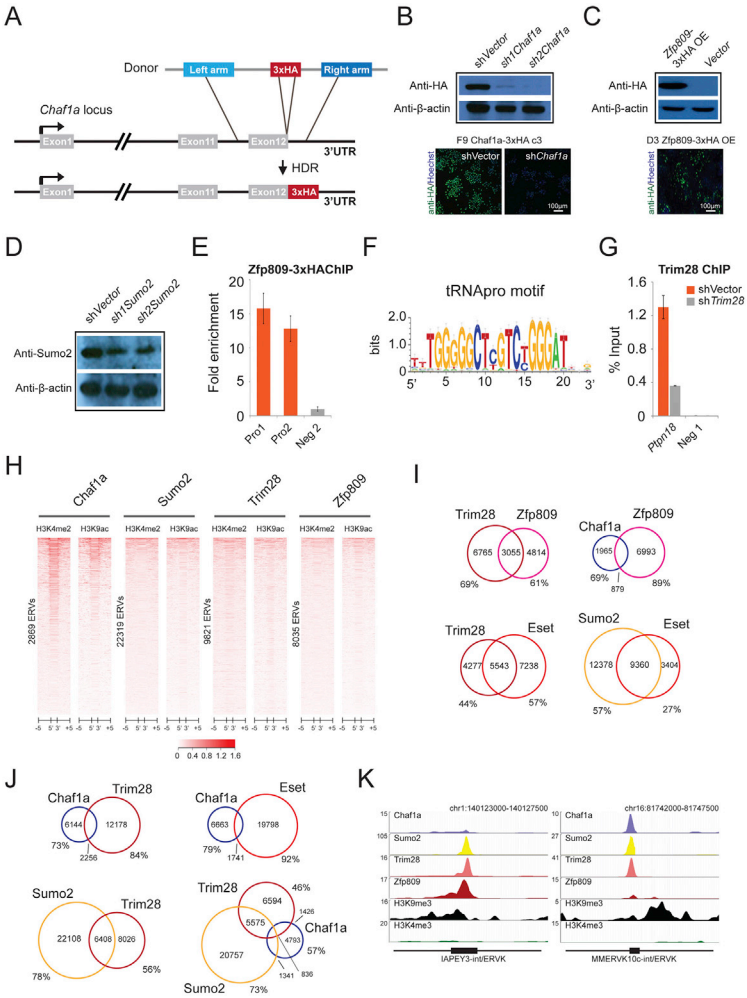


Figure S4. Genomic Binding of Chaf1a, Sumo2, Trim28, and Zfp809 on Endogenous ERVs, Related to Figure 4

(A) Schematic representation of the strategy for CRISPR-Cas engineering of 3xHA tag to the 3′ end of the *Chaf1a* loci. A site-specific double-strand break was induced with Cas9. 3xHA tag in the donor plasmid was added into the end of exon 12 by homology-directed repair (HDR).

(B) *Chaf1a* shRNA knockdown in CRISPR line F9 Chaf1a-3xHA clone 9. Changes in levels of protein were detected by Western blot (upper panel) and immunostaining (lower panel) using anti-HA antibodies. Anti-β-actin was used as a loading control for the Western blot.

(C) Overexpression of Zfp809-3xHA in D3 mESCs. The overexpressed Zfp809-3xHA protein was detected by Western blot (upper panel) and immunostaining (lower panel) using anti-HA antibodies.

(D) Specificity of the Sumo2 antibody. E14 cells were transfected with shRNA targeting *Sumo2*. Changes in the levels of Sumo2 protein were detected by Western blot using anti-Sumo2 antibody.

(E) Zfp809 binds to the proline PBS. Zfp809-3xHA ChIP was performed using HA antibody. The enrichment of ZFp809-3xHA on the proline PBS was measured by qPCR using two pairs of primers that are specific to proline PBS. Data are presented as mean ± s.e.m from independent replicate experiments.

(F) Identification of tRNA binding site proline PBS motif in Zfp809-3xHA ChIP-seq.

(G) *Ptpn18*, a known target of Trim28, was used to validate the Trim28 ChIP. The enrichment of Trim28 on *Ptpn18* was tested by qPCR. Data are presented as mean ± s.e.m from independent replicate experiments. Knockdown of *Trim28* caused a significant drop in Trim28 enrichment, indicating the robustness of the Trim28 ChIP.

(H) Heat maps of histone modifications at the ERV genomic regions bound by Chaf1a, Sumo2, Trim28 and Zfp809. The heat maps are clustered according to the enrichment profile of H3K4me2. Chaf1a-bound ERVs are characterized by detectable H3K4me2 and H3K9ac for significant numbers of ERVs in comparison to Sumo2, Trim28 and Zfp809 bound ERVs.

(I) Venn diagrams demonstrating the number of common and differentially bound ERVs of the indicated factors. Sumo2, Zfp809, Trim28 and Eset interact extensively whereas Chaf1a has less commonly bound ERVs with Zfp809.

(J) Venn diagrams indicating the number of common and uniquely-bound non-ERV loci among the indicated factors. The Venn diagrams demonstrate that interactions among the factors are significantly reduced at the non-ERV loci.

(K) UCSC genome browser screenshots of IAPEY3-int and MMERVK10C-int repeat elements. IAPEY3-int is bound by Sumo2, Trim28 and Zfp809, but not Chaf1a, while MMERVK10C-int_ERVK is bound by Chaf1a, Sumo2 and Trim28, but not Zfp809. Both the ERVs are enriched with H3K9me3.

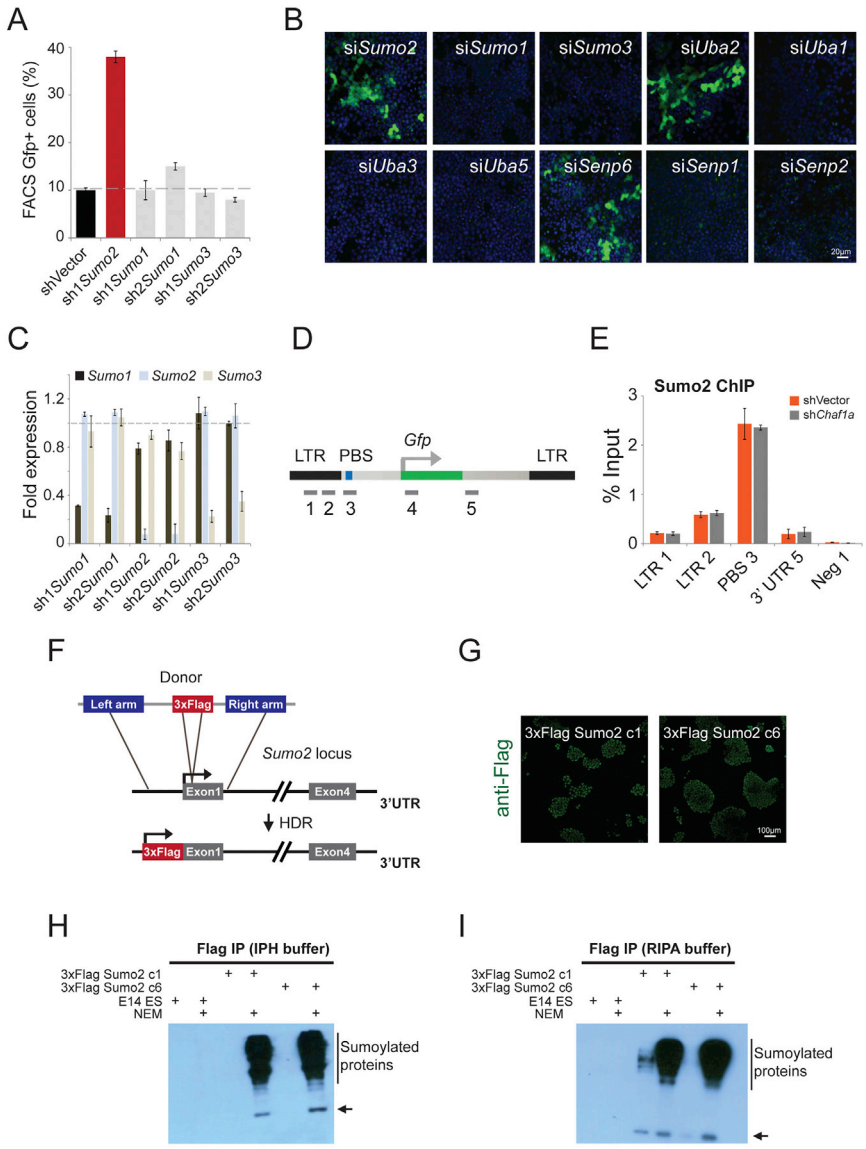


Figure S5. Specific Role of Sumo2 in Proviral Regulation, Related to Figure 5

(A) MMLV-*Gfp* rescue in the knockdown of different small ubiquitination-like modifiers (Sumo). shRNA knockdown targeting *Sumo1*, *Sumo2*, or *Sumo3* was performed in F9 cells infected with MMLV-*Gfp* retrovirus, *Gfp* expression was tested by FACS. Values are mean ± s.e.m from independent replicate experiments. Only the specific depletion of *Sumo2* resulted in the rescue of MMLV-*Gfp*.

(B) Representative images for MMLV-*Gfp* rescue by siRNA knockdown of different small ubiquitination-like modifiers from the genome-wide siRNA library screen.

(C) Real-time qPCR analysis of Sumo shRNA knockdown efficiency and specificity in mouse ES E14 cells. The relative expression levels of each gene were normalized to gene expression levels in the shVector control. Values are mean ± s.e.m from independent replicate experiments.

(D) Schematic drawing of the locations of primer pairs used for ChIP-qPCR.

(E) Knockdown of *Chaf1a* does not affect the binding of Sumo2 on the proviral elements. Sumo2 ChIP was performed on E14 cells after *Chaf1a* depletion. The enrichment of Sumo2 on the proviral elements was measured by qPCR. Values are mean ± s.e.m from independent replicate experiments.

(F) Schematic representation of the strategy for CRISPR-Cas engineering of 3xFlag tag to the exon1 of *Sumo2* loci in E14 cells.

(G) Characterization of the Sumo2 CRISPR cell lines by immunostaining. Two homozygous targeted lines were selected for immunostaining using anti-Flag antibody.

(H and I) Optimization of Sumo2 protein pull-down. Immunoprecipitation (IP) assays were carried out in two CRISPR cell lines using anti-Flag M2 beads in IPH buffer and RIPA buffer respectively. NEM was added to protect the sumoylated proteins from de-sumoylation by SENPs in the cell lysates. Arrows indicate Sumo2 protein.

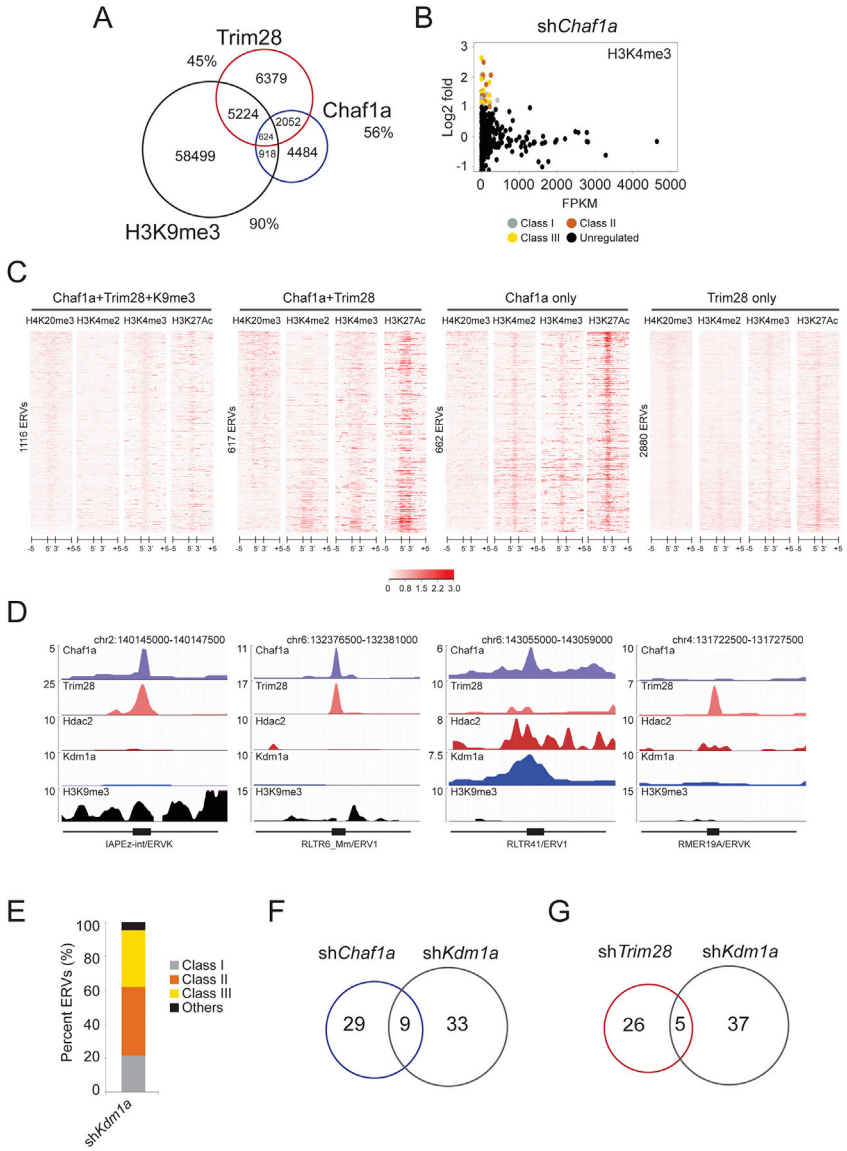


Figure S6. Regulation of Class I, II, and III ERVs by Chaf1a, Related to Figure 6

(A) Venn diagrams demonstrating the number of common and uniquely-bound non-ERV loci among the indicated factors. Chaf1a, Trim28, and H3K9me3 demonstrate little overlap and interactions at the non-ERV loci.

(B) Correlation between the upregulation of ERVs in E14 cells upon the depletion of *Chaf1a* and the enrichment of H3K4me3 on these regulated ERVs. RNA-Seq data for sh*Chaf1a* and the shVector control were used to calculate the Log2 fold changes values for all annotated ERVs, whereas the ChIP-Seq data for H3K4me3 was used to measure their enrichment on the ERVs in E14 cells. The data is plotted as shown using R. Grey, orange and yellow dots indicate the ERVs with significantly increased expression from class I, II and III ERVs respectively. Black dots indicate the non-regulated ERVs. All the Chaf1a-regulated ERVs have low levels of H3K4me3 enrichment.

(C) Enrichment of H4K20me3, H3K4me2, H3K4me3 and H3K27Ac at the genomic regions of the ERV loci bound in the indicated categories. The heat maps are clustered according to the enrichment profile of H3K20me3. ERVs bound by Chaf1a only show low but detectable enrichment of H3K4me2 and H3K27Ac.

(D) UCSC genome browser screenshots of the representative repeat elements. IAPEz-int is bound by both Chaf1a and Trim28 as well as enriched for H3K9me3. RLTR6_Mm is bound by both Chaf1a and Trim28 with low or undetectable Kdm1a, Hdac2 and H3K9me3 enrichment. RLTR41 is exclusively bound by Chaf1a, but not Trim28. It is enriched for Kdm1a and Hdac2, with undetectable H3K9me3. Lastly RME19A is bound exclusively by Trim28 where no Kdm1a, Hdac2 and H3K9me3 are detected.

(E) Percentage stacked columns indicating the classes of upregulated ERVs upon the depletion of Kdm1a. Similar to Chaf1a, Class III ERVs (Old ERVs) are largely regulated by Kdm1a.

(F and G) Venn diagrams demonstrating the number of commonly and differentially upregulated ERVs among the depletion of indicated factors. Stronger correlation is observed between *Chaf1a* and *Kdm1a* knockdowns (F) in comparison to *Kdm1a* and *Trim28* (G).

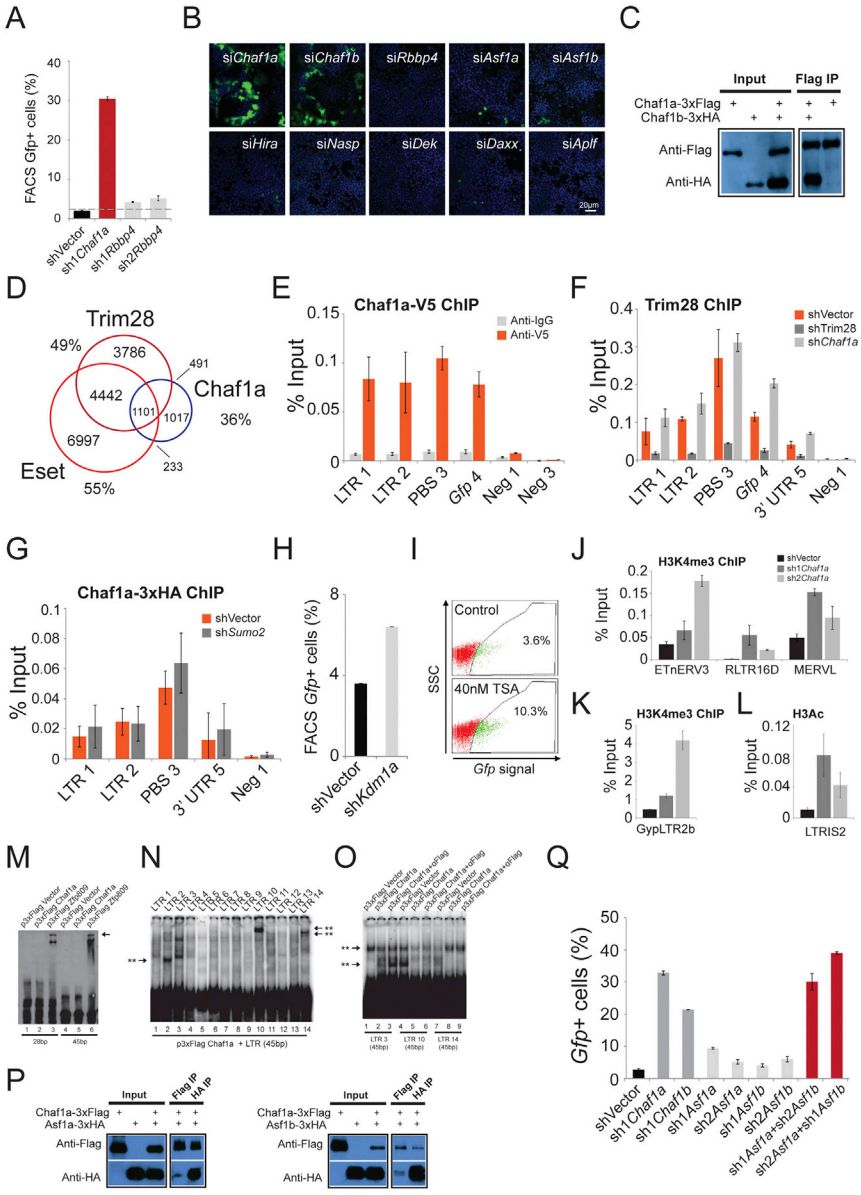


Figure S7. Direct Regulation of Infected Proviruses by Chaf1a, Related to Figure 7

(A) Specific silencing of proviruses in mESCs by Chaf1a, but not by the Caf-1 smallest subunit Rbbp4. *Chaf1a* or *Rbbp4* were depleted by shRNA in D3 ESCs, *Gfp* expression was analyzed by FACS. All the values shown are averages ± s.e.m from independent replicate experiments.

(B) Representative images for MMLV-*Gfp* rescue by different histone chaperone siRNA knockdowns from the genome-wide siRNA library screen. The result indicates the specific role of histone chaperone Chaf1a and Chaf1b in proviral silencing.

(C) Confirmation of Chaf1a and Chaf1b interaction by coIP. Flag-tagged Chaf1a and HA-tagged Chaf1b were co-transfected into 293T cells. CoIP was performed using Flag and HA antibody. The interaction was detected by Western blot.

(D) A Venn diagram showing the number of common and uniquely-bound ERV loci among the indicated factors. More than 2/3 of Trim28/Chaf1a ERVs were also enriched for Eset.

(E) Validation of Chaf1a binding on proviral DNA using anti-V5 antibody ChIP. V5 ChIP was performed on the sample with Chaf1a-V5 transient overexpression, and the enrichment was analyzed by qPCR.

(F) Knockdown of Chaf1a did not affect the binding of Trim28. Trim28 ChIP was performed with an antibody specific to Trim28. Data are presented as mean ± s.e.m from independent replicate experiments. Knockdown of *Trim28* itself resulted in the significant drop of Trim28 enrichment at proviruses.

(G) The binding of Chaf1a on the proviral DNA was not affected by the depletion of *Sumo2*. Chaf1a-3xHA ChIP was performed on the samples with *Sumo2* knockdown and the enrichment was analyzed by qPCR.

(H) MMLV-*Gfp* rescue by *Kdm1a* knockdown. *Kdm1a* was depleted by shRNA in D3 ESCs, *Gfp* expression was analyzed by FACS. All the values shown are averages ± s.e.m from independent replicate experiments.

(I) TSA treatment de-repressed the MMLV-*Gfp* in mESCs. D3 mESCs was treated with 40nM TSA for 72 hr, and the *Gfp* signal was analyzed by FACS.

(legend continued on next page)

(J and K) Knockdown of *Chaf1a* increased the active H3K4me3 mark on ERVs. The enrichment was analyzed by qPCR and values were normalized to the input samples. Data are presented as mean ± s.e.m from independent replicate experiments.
(L) Knockdown of *Chaf1a* increased the active H3Ac mark on ERVs. The enrichment was analyzed by qPCR and values were normalized to the input sample. Data are presented as mean ± s.e.m from independent replicate experiments.
(M–O) Chaf1a does not bind to the proviral DNA. (M) Electrophoretic mobility shift assay (EMSA) shows that Zfp809 but not Chaf1a binds specifically to proviral proline PBS probes labeled with biotin (both 28bp and 45bp probes). The P3xFlag vector was used as control. P3xFlag-*Chaf1a* denotes FLAG-purified Chaf1a protein. P3x*Zfp809* denotes FLAG-purified Zfp809 protein. Lane 1-3: p3xFlag vector, p3xFlag-*Zfp809* or p3xFlag-*Chaf1a* reaction with 1 fmol 28bp Pro PBS probe; Lane 4-6: pFlag vector, p3xFlag-*Zfp809* or p3xFlag-*Chaf1a* reaction with 1 fmol 45bp Pro PBS probe. The arrow denotes the specific binding of Zfp809. (N) The binding profile of FLAG-purified Chaf1a protein on ^{32}P labeled LTR fragments by EMSA. Lane 1-14: p3xFlag-*Chaf1a* reaction with fourteen 48bp probes covering the entire LTR. Two asterisks denote potential binding of Chaf1a. (O) Validation of the Chaf1a binding on the LTR probes by EMSA. P3xFlag vector, p3xFlag-*Chaf1a* or p3xFlag-*Chaf1a*+anti-Flag reaction with LTR 3 (Lane 1-3), LTR 10 (Lane 4-6) or LTR 14 (Lane 7-9).
(P) Chaf1a interacts with Asf1a/b. *Chaf1a-3xFlag* and *Asf1a-3xHA* or *Asf1b-3xHA* was co-transfected into 293T cells, and coIP was performed. The interaction between Chaf1a and Asf1a/b was detected by Western blot.
(Q) Asf1a and Asf1b play complementary roles in the regulation of exogenous MMLV-*Gfp*. Treatment with shRNA constructs targeting *Asf1a* and *Asf1b* was carried out singly or in combinations in F9 cells. *Gfp* expression was analyzed by FACS. All the values shown are mean ± s.e.m from independent replicate experiments.