HHS Public Access

Author manuscript

Cell. Author manuscript; available in PMC 2015 December 21.

Published in final edited form as:

Cell. 2015 September 24; 163(1): 230–245. doi:10.1016/j.cell.2015.08.037.

Systematic Identification of Factors for Provirus Silencing in Embryonic Stem Cells

Bin Xia Yang[1,22], **Chadi A. EL Farran**[1,2,22], **Hong Chao Guo**[3,22], **Tao Yu**[1,2,22], **Hai Tong Fang**[1], **Hao Fei Wang**[1,2], **Sharon Schlesinger**[4,5], **Yu Fen Samantha Seah**[1], **Germaine Yen Lin Goh**[6], **Suat Peng Neo**[7], **Yinghui Li**[9], **Matthew C. Lorincz**[10], **Vinay Tergaonkar**[9,21], **Tit-Meng Lim**[2], **Lingyi Chen**[3], **Jayantha Gunaratne**[7,8], **James J. Collins**[11,12,13,14], **Stephen P. Goff**[4,5,15], **George Q. Daley**[14,16,17,18,19], **Hu Li**[20], **Frederic A. Bard**[6,21], and **Yuin-Han Loh**[1,2,*]

[1]Epigenetics and Cell Fates Laboratory, A*STAR Institute of Molecular and Cell Biology, 61 Biopolis Drive Proteos, Singapore 138673, Singapore

[2]Department of Biological Sciences, National University of Singapore, Singapore 117543, Singapore

[3]College of Life Sciences, Nankai University, Tianjin 300071, China

[4]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA

[5]Department of Microbiology and Immunology, Columbia University, New York, NY 10032, USA

[6]Membrane Traffic Laboratory, A*STAR Institute of Molecular and Cell Biology, 61 Biopolis Drive Proteos, Singapore 138673, Singapore

[7]Quantitative Proteomics Group, A*STAR Institute of Molecular and Cell Biology, 61 Biopolis Drive Proteos, Singapore 138673, Singapore

[8]Department of Anatomy, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597, Singapore

[9]Division of Cancer Genetics and Therapeutics, Laboratory of NF-kB Signaling, A*STAR Institute of Molecular and Cell Biology, 61 Biopolis Drive Proteos, Singapore 138673, Singapore

[10]Department of Medical Genetics, Life Sciences Institute, University of British Columbia, Vancouver, BC V6T 1Z3, Canada

[*]Correspondence: yhloh@imcb.a-star.edu.sg.

[22]Co-first author

ACCESSION NUMBERS

The accession number for all sequencing samples reported is GEO: GSE70865.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.08.037.

AUTHOR CONTRIBUTIONS

B.X.Y. and C.A.E.F. designed and performed research, analyzed data, and wrote the paper. H.C.G., T.Y., H.T.F., H.F.W., Y.F.S.S., Y.L.G.G., S.P.N., and Y.-H.L. designed and conducted research. S.S., M.C.L., V.T., T.-M.L., L.Y.C., J.G., J.J.C., S.P.G., G.Q.D., H.L., and F.A.B. analyzed data. Y.-H.L. designed research, analyzed data, and wrote the paper.

[11]Department of Biological Engineering, Synthetic Biology Center, Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[12]Harvard-MIT Program in Health Sciences and Technology, Broad Institute of MIT and Harvard, Cambridge, MA 02139, USA

[13]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA

[14]Howard Hughes Medical Institute, Boston, MA 02115, USA

[15]Howard Hughes Medical Institute, New York, NY 10032, USA

[16]Stem Cell Transplantation Program, Division of Pediatric Hematology/Oncology, Boston Children's Hospital and Dana-Farber Cancer Institute, Boston, MA 02115, USA

[17]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA

[18]Harvard Stem Cell Institute, Boston, MA 02115, USA

[19]Manton Center for Orphan Disease Research, Boston, MA 02115, USA

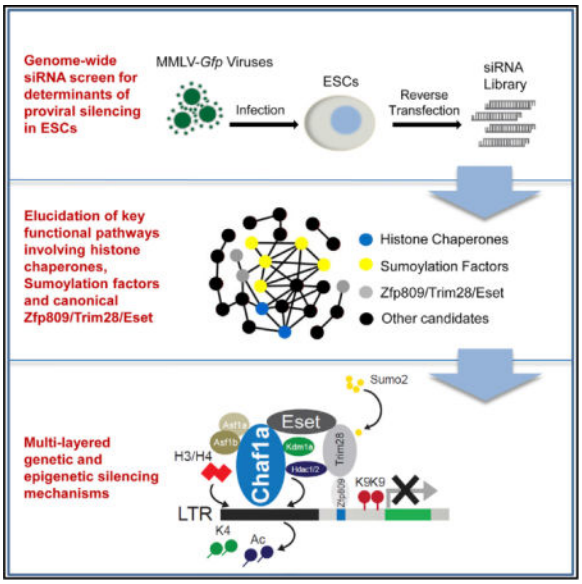
[20]Center for Individualized Medicine, Department of Molecular Pharmacology & Experimental Therapeutics, Mayo Clinic, Rochester, MN 55905, USA

[21]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119077, Singapore

SUMMARY

Embryonic stem cells (ESCs) repress the expression of exogenous proviruses and endogenous retroviruses (ERVs). Here, we systematically dissected the cellular factors involved in provirus repression in embryonic carcinomas (ECs) and ESCs by a genome-wide siRNA screen. Histone chaperones (Chaf1a/b), sumoylation factors (Sumo2/Ube2i/Sae1/Uba2/Senp6), and chromatin modifiers (Trim28/Eset/At-f7ip) are key determinants that establish provirus silencing. RNA-seq analysis uncovered the roles of Chaf1a/b and sumoylation modifiers in the repression of ERVs. ChIP-seq analysis demonstrates direct recruitment of Chaf1a and Sumo2 to ERVs. Chaf1a reinforces transcriptional repression via its interaction with members of the NuRD complex (Kdm1a, Hdac1/2) and Eset, while Sumo2 orchestrates the provirus repressive function of the canonical Zfp809/Trim28/Eset machinery by sumoylation of Trim28. Our study reports a genome-wide atlas of functional nodes that mediate proviral silencing in ESCs and illuminates the comprehensive, interconnected, and multi-layered genetic and epigenetic mechanisms by which ESCs repress retroviruses within the genome.

Graphical abstract



INTRODUCTION

The expression of proviruses and endogenous retroviruses (ERVs) is restricted in pluripotent stem cells (Feuer et al., 1989; Niwa et al., 1983; Teich et al., 1977). This silencing has likely evolved for the protection of germline cells from insertional mutagenesis (Gaudet et al., 2004; Walsh et al., 1998). The expression and DNA methylation profiles of the Moloney murine leukemia virus (MMLV) have been investigated in embryonic carcinoma cells (ECs) and embryonic stem cells (ESCs) (Niwa et al., 1983). DNA methylation is thought to repress the expression of viral genes in differentiated cells, while repression in pluripotent cells is mediated by both *cis*-acting de novo methylation of the integrated proviruses (Gaudet et al., 2004; Walsh et al., 1998) and *trans*-acting transcriptional repressors (Petersen et al., 1991; Stewart et al., 1982; Walsh et al., 1998; Wolf et al., 2008a; Wolf and Goff, 2007).

It has been reported that many ERVs affect cellular gene activity by acting as alternative promoters or enhancers (Peaston et al., 2004). For example, MERVL is transiently activated during the mouse two-cell (2C) stage, regulating the expression of 2C-specific genes (Macfarlan et al., 2012). ERVs may also function in the reprogramming of somatic cells into induced pluripotent stem cells (iPSCs). Specific ERVs are re-activated during the reprogramming process, while other classes of ERVs have to be silenced to attain complete reprogramming (Friedli et al., 2014; Wissing et al., 2012). Together, these studies suggest that proviral silencing is a characteristic of the pluripotent state, and the precise expression of ERVs have critical roles during embryogenesis and development.

Various studies have implicated diverse epigenetic mechanisms in the silencing of retroviruses and ERVs. Repression is thought to be dependent on a conserved sequence element termed the primer binding site (PBS). Factors such as Zfp809, Trim28, and Eset are responsible for mediating the H3K9me3 repressive silencing mechanism (Friedli et al., 2014; Rowe et al., 2010; Wolf and Goff, 2007, 2009; Wolf et al., 2008b). Eset was shown to be involved in the repression of retroviruses and subfamilies of ERVs, predominantly of

class I and II ERVs (Karimi et al., 2011; Matsui et al., 2010). More recently, viral-silencing factors such as the zinc finger protein Yin yang 1 (Yy1), Erb3 binding protein 1 (Ebp1), and the polycomb repressive complex 2 (PRC2) catalytic subunit Ezh2 (Schlesinger et al., 2013; Schlesinger and Goff, 2013; Wang et al., 2014) have been described. Other studies reporting the role of host factors governing ERVs in model organisms, such as *Saccharomyces cerevisiae* (Maxwell and Curcio, 2007) have also provided critical evolutionary insight into the dynamics of retroviral regulation.

Despite many efforts to identify the factors involved, many components of the epigenetic machinery required for stable silencing of proviruses and ERVs remains poorly characterized. To advance our understanding, we developed a powerful high-throughput screening approach based on a provirus MMLV-*Gfp* reporter (Schlesinger et al., 2013) and genome-wide small interfering RNA (siRNA) knockdown. Our screen identified 303 determinants of viral silencing in mouse ESCs with high confidence and provides a genome-wide functional interrogation of determinants mediating proviral silencing in pluripotent embryonic stem cells.

RESULTS

Unbiased Genome-wide siRNA Screen for Determinants of Proviral Silencing in Embryonic Carcinoma Cells

To define the factors involved in the silencing process, we developed a high-throughput screening approach based on a provirus MMLV-*Gfp* reporter and siRNA knockdown in F9 ECs (Figure 1A). F9 cells were infected with the MMLV-*Gfp* virus and then reverse transfected with siRNA in 384-well plates. Expression of *Gfp* on day 4 post-infection indicated retrovirus activation.

We first confirmed the sensitivity of the reporter assay via knockdown of canonical repressive genes *Trim28* and *Eset*. Consistently, imaging, and fluorescence-activated cell sorting (FACS) analysis showed that knockdown of both factors dramatically relieved the repression of retroviral *Gfp* (Figures S1A and S1B). We next carried out a pilot screen on the kinome siRNA library in F9 cells, using non-targeting (siNT) *Trim28* and *Eset* siRNAs as controls. The kinome library screen was analyzed by Z-prime score (Figures S1C–S1F). From the screen, we identified both known (*Trim28* and *Cdk9*) and undetermined factors (*Chuk*, *Epha4*, *Csnk1e*, *Sgpp1*, and *Npp4a*) responsible for retrovirus silencing (Figure S1G). *Cdk9* was previously reported to interact with HIV-1 Tat protein and regulate HIV-1 transcription (Kao et al., 1987).

Next, we carried out a whole genome siRNA screen targeting 20,000 genes in F9 cells (Figure 1A). Candidates that caused excessive cell death upon siRNA knockdown were excluded using a stringent nuclei number cut-off threshold. Based on the normalized Gfp signal cut-off value, which short-listed factors that had values larger than 2 SDs from the mean of the negative controls (Figure 1B), 650 factors were short-listed (Table S1). Among the hits are factors previously implicated in retroviral silencing process, such as *Eset*, *Zfp809*, *Yy1*, and *Trim28*. In addition, new candidates identified include *Ube2i*, *Pcna*, *Hist1h3c*, *Mphosph8*, *Adcy6*, *Sh3bp1*, and *Thyn1* (Figure 1C).

To validate the genome-wide siRNA screen, we performed secondary siRNA screens utilizing the MMLV-*Gfp* reporter and an independent MMLV-*mCherry* reporter. We observed strong correlation between the two reporters (Figure 1D). To minimize possible non-specific effects from the pooled siRNA, we designed two pairs of short hairpin RNAs (shRNAs) for 31 candidate genes and three non-candidate genes. shRNA validation was performed in F9 cells, followed by FACS analysis of *Gfp* expression. shRNA knockdown efficiencies were confirmed by qPCR (Figure S1H) and western blot analysis for selected genes (Figure S1I). Notably, we observed robust Gfp reactivation for the majority of top hits (Figure 1E). From the results of secondary siRNA and shRNA screens, we focused on the top 303 hits that were highly corroborative with the primary screen and are considered high confidence candidates.

Network Analysis of the Candidates Reveals Multiple Interacting Pathways Involved in Proviral Silencing

We performed Gene Ontology (GO), KEGG, and Interpro analysis (Huang et al., 2009) on the top 303 hits and elucidated 148 statistically enriched biological processes and pathways, including chromatin modification and organization, protein sumoylation and phosphorylation, regulation of transcription, DNA replication, DNA repair, and methylation (Figure S2A; Table S2). Protein-protein interaction analysis of the high confidence hits demonstrates tight and dense interaction between the candidate proteins (Figure 2A). In addition, cellular component analysis revealed that the candidates were widely distributed in different sub-cellular fractions (Figures 2A and S2B). These suggest that proviral silencing is controlled by multilayered machineries involving components of different cellular pathways and with varied cellular localization.

Candidate Genes Are Potent Repressors of Provirus Expression in Embryonic Stem Cells

We analyzed the expression profiles of the candidate genes in over 100 cell lines using the cTen database (Shoemaker et al., 2012). The majority of candidate genes are highly expressed in embryonic stem cell lines and are low in other tissue-specific cell lines (Figure S2C). The expression of selected candidates was further tested in the mouse ESC lines E14 and D3, mouse EC lines F9 and P19, as well as in differentiated mouse embryo fibroblasts (MEFs). Consistent with cTen enrichment scores, qPCR analyses showed embryonal and stem cell-specific expression of the candidates (Figure S2D).

To further interrogate the function of our candidate hits, we performed network analysis of the hits based on their tiered ranking. We observed greater interactions among our top 50 candidates, although the lower ranked hits also exhibited specific interactions indicative of their biological significances (Figure 2B). Among the top 20 hits are the histone chaperones (Chaf1a/b), sumoylation modification genes (Ube2i, Sumo2, Uba2, Sae1, and Senp6), and chromatin-bound factors (Eset, Atf7ip, Zfp809, Trim28). To test the functional specificity of these strong candidates in mESCs, we conducted siRNA and shRNA knockdowns in two mESC lines E14 and D3 and in two differentiated cell types, 3T3 and MEFs. The results of the *Gfp* reporter rescue assay from mESC lines corroborate well with the primary screen done in F9 cells (Figures 2C, 2D, and S2E). In contrast, MMLV-driven expression of *Gfp* or

mCherry was high in 3T3 and MEFs at the outset and knockdown of candidate genes did not result in perturbations of the reporter signal in these cell lines (Figures 2E and S2F).

To further assess the ESC specificity of our candidates, we differentiated E14 and D3 cells via embryoid body (EB) formation and neural differentiation (Ying et al., 2003). The differentiated cells lost their ESC-specific morphologies and pluripotency markers and expressed high levels of differentiation genes (Figures 2F and S2G). Consistent with a previous report, the MMLV virus remain silenced in differentiated ESCs (Niwa et al., 1983). None of the candidate gene knockdowns in the differentiated cells could rescue MMLV-*Gfp* reporter expression (Figures 2G and S2H), suggesting that alternative or additional silencing pathways are active in these cells. Relative copy number of integrated reporters in E14 and the differentiated cells was indistinguishable, ruling out the possibility of reduced viral integration in the latter (Figures S2I and S2J). In addition, knockdown of the top hits did not reduce provirus integration efficiency in E14 cells (Figure S2K). Of note, we observed no significant change in Gfp signal driven by an integrated non-LTR reporter (PiggyBac-CAG-*Gfp*) upon knockdown of the top hits (Figures S2L and S2M). This strongly suggests that the mode of proviral regulation by the factors is transcriptional or epigenetic.

Chaf1a/b and Sumoylation Modification Complex Play Critical Roles in Regulating ERVs

To evaluate the roles of Chaf1a/b and the sumoylation factors in ERV regulation, we measured ERV expression by qPCR upon depletion of the candidates. Consistent with a previous study, *Trim28* knockdown elicited reactivation of IAP elements in ESCs (Figure S3A) (Rowe et al., 2010). Intriguingly, we found up-regulation of class I (GLN), class II (MMERVK10c), and class III (MERVL) elements following depletion of the factors from the Caf1 complex, sumoylation complex, and *Atf7ip* (Figure S3A). Notably, Northern blot assays confirmed increased transcription of MERVL, but not of IAP and MusD elements in *Chaf1a/b* depleted E14 cells (Figure S3C). Meanwhile, knockdown of selected weaker candidates also showed consistent de-repression of MERVL but not of the other ERVs (Figure S3B).

To further delineate the regulatory roles of the candidates on ERVs, we performed genome-wide RNA sequencing (RNA-seq) of *Chaf1a/b*-, *Sumo2*-, *Sae1*-, *Ube2i*-, *Ube2*-, *Senp6*-, *Trim28*-, *Eset*-, and *Atf7ip*-depleted cells. Transcriptomic analyses revealed significant de-repression of several families of ERVs upon depletion of each factor (Figure 3A; Table S3). In contrast to their effects on global gene expression (Figure S3D), the majority of the ERV targets are upregulated upon shRNA knockdown (Figure 3B). Together, these suggest an ERV-specific repressive function of the candidates.

Next, we evaluated the ERV classes regulated by the candidates. Chaf1a/b depletion resulted in the de-repression of large numbers of Class III ERVs, while the sumoylation and canonical factors regulated more Class II ERVs (Figure 3C). Cluster analysis detected strong correlation of ERV regulation within the Chaf1a/b, sumoylation factors, and the chromatin binding factors Trim28, Atf7ip, and Eset (Figure 3D), whereas the analysis of global gene expression displayed a different pattern (Figure S3E).

Remarkably, Trim28 shares significant similarity with both the Chaf1a/b and sumoylation factors in their ERV regulation (Figure 3D), suggesting overlapping mechanisms. ERVs controlled by Atf7ip overlapped extensively with the ones regulated by Trim28 or Eset (Figure S3F), indicating that Atf7ip may be integral to the canonical Krab-Zfp/Trim28/Eset machinery. Atf7ip was shown to be a co-factor of Eset that helps in facilitating the conversion of H3K9me2 to H3K9me3 (De Graeve et al., 2000; Wang et al., 2003). Furthermore, the ERVs regulated by Chaf1a overlaps significantly with the ones regulated by Chaf1b (Figures 3D, 3E and 3G), but differ significantly from those controlled by Sumo2 (Figures 3F and 3H). One key feature of the cluster of sumoylation genes is the strong correlation between the factors in the specific control of their ERV targets as shown by the tight pairwise correlation (Figure 3E). This suggests a coordinated mechanism involving multiple members of the same sumoylation pathway. Interestingly, most ERVs regulated by Sae1 and Ube2i are part of the larger number of ERVs governed by Sumo2, suggesting a central role for Sumo2 in this sumoylation process (Figure 3G). It is noteworthy that many ERVs regulated by Sumo2 are similarly governed by Trim28 (Figures 3F and 3H).

To validate the RNA-seq data, we performed qPCR on each class of ERVs (Figure S3G). Consistently, RLTR6_Mm/ERV1 was specifically regulated by the sumoylation factors, while ET-nERV3-int/ERVK was regulated by Atf7ip, Eset, and Chaf1a, but not by the sumoylation factors. MT2_Mm/ERVL was sharply upregulated upon the depletion of Chaf1a/b, while expression was less perturbed with depletion of factors from the other two clusters. Finally, LTR16D was upregulated upon depletion of genes from all the clusters.

Chaf1a and Sumo2 Are Directly Recruited to ERVs

We wanted to determine whether Chaf1a and Sumo2 are enriched on genomic ERVs. First, we introduced 3xHA tags at the 3′end of the endogenous *Chaf1a* locus in F9 cells using CRISPR/Cas technology (Figure S4A). The Chaf1a-3xHA cell line was characterized by shRNA knockdown, which led to the specific reduction of Chaf1a-3xHA as measured by western blot and immunostaining (Figure S4B). In addition, a Zfp809-3xHA overexpression D3 cell line was also established and similarly characterized (Figure S4C). The reliability of the Sumo2 antibodies used for chromatin immunoprecipitation (ChIP) was confirmed with knockdown of *Sumo2* followed by western blot analysis (Figure S4D). To survey the global binding profiles of Chaf1a, Sumo2, Trim28, and Zfp809 on genomic ERV loci, we performed ChIP sequencing (ChIP-seq). The quality of the ChIP DNA was determined by qPCR and motif analysis. Zfp809-3xHA ChIP-qPCR yielded high enrichment at proline PBS site (Figures S4E and S4F), and Trim28 ChIP-qPCR showed strong binding at a previously reported target gene *Ptpn18* (Figure S4G) (Hu et al., 2009).

ChIP-seq analysis revealed that both Chaf1a and Sumo2 are recruited to loci of members of several classes of ERVs (Figures 4A and 4B; Table S4). We next asked if the bound ERV loci are enriched for any histone modifications. We compared the Chaf1a, Sumo2, Trim28, and Zfp809 ChIP-seq data with publicly available datasets of histone marks and Eset ChIP-seq. Although the majority of ERVs bound by Chaf1a are enriched with H3K9me3 (Figure 4C), the H3K9me3 is of lower intensity compared to that of Trim28, Zfp809, and Sumo2 bound ERVs (Figure 4C). Intriguingly, considerable proportions (15%) of Chaf1a bound

ERVs are also enriched for the active H3K4me3 modification (Figure 4C). Furthermore, Chaf1a-bound ERVs exhibit higher levels of H3K4me2 and H3K9Ac (Figure S4H). This raises the possibility that additional accessory proteins may be required for Chaf1a to exert the silencing effects. Notably, Sumo2-targeted ERV loci are associated with elevated H3K9me3 levels and reduced levels of H3K4me3 modification. This binding pattern strongly resembles that of Zfp809 and Trim28 (Figure 4C). In contrast, the non-ERV loci bound by Chaf1a were enriched with abundant H3K4me3 marks and had no trace of H3K9me3 modifications. On the other hand, Sumo2/Trim28/Zfp809-bound loci exhibit detectable but low levels of H3K9me3 (Figure 4D). Collectively, this indicates differing modes of regulation by which individual factors control ERVs and non-ERV targets (Figures 4C, 4D, and S3C).

To determine the action of Chaf1a and Sumo2, we represented ERV loci bound by these factors in Venn diagrams. We found that Trim28 binds 56% of Chaf1a-bound sites, while 57% of Chaf1a ERVs are also targets of Sumo2 (Figure 4E). Moreover, only 31% of Chaf1a ERV loci are enriched for Zfp809 (Figure S4I). In contrast, 77% of Trim28 targets and 73% of Eset-bound ERVs are accompanied by enrichment of Sumo2 (Figures 4E and S4I). When we extend the analysis to three factors, we observed that more than 80% of Chaf1a/Trim28 and Chaf1a/Eset common targets have Sumo2 binding (Figure 4F). These observations strongly suggest a possible role of Sumo2 in Trim28/Eset ERV regulation. The co-regulation of Chaf1a and Sumo2 with the canonical Zfp809/Trim28/Eset machinery seems to be ERV-specific as very little overlap was observed between the factors on non-ERV loci (Figure S4J). Collectively, in terms of ERV regulation, Chaf1a binding is clustered away from the Sumo/Zfp809/Trim28/Eset axis (Figures 4G, 4H, and S4K). This is remarkably similar to the pattern observed from the RNA-seq data (Figure 3D). Overall, our ChIP data provides the first biochemical demonstration that a histone chaperone and a sumoylation modification protein can exert direct regulation of genomic ERVs.

Sumo2 Orchestrates the Viral Silencing Activities of Trim28 through Its Sumoylation Modification

Our genome-wide siRNA screen identified Sumo2, and not Sumo1 or Sumo3, to have a distinct role in proviral silencing (Figures S5A–S5C). The global RNA-seq and ChIP-seq data further suggest that Sumo2 may repress proviruses and ERVs through modulation of the Trim28/Eset machinery (Figure 5A). To test this possibility, we first performed Sumo2 ChIP-qPCR and identified its binding on the proviral LTR. Importantly, when *Trim28* was knocked-down, the level of Sumo2 binding on both proviral elements and most of the ERVs tested was drastically reduced (Figures 5B and 5C). In contrast, enrichment of Sumo2 was not affected by *Chaf1a* knockdown (Figures S5D and S5E). Furthermore, the removal of Sumo2 abolished the binding of Trim28 at the LTR (Figure 5D).

To interrogate whether Sumo2 directly targets Trim28 for sumoylation, we studied well characterized 3xFlag-Sumo2 E14 cells generated using CRISPR/Cas technology (Figures S5F–S5I). Notably, we identified Trim28 in the pull-down of sumoylated proteins (Figure 5E).

Venn diagram analysis of ChIP-seq data indicates that ~90% of Sumo2/Trim28-bound ERV sites are marked with H3K9me3 modifications (Figure 5F). Trim28 is known to mediate the recruitment of Eset, which in turn deposits the repressive H3K9me3 mark at the proviral LTR (Matsui et al., 2010). Consistently, *Sumo2* knockdown resulted in concomitant reduction in H3K9me3 marks and elevation of H3K4me3 modifications at the proviral elements and ERVs to levels that are comparable with that seen upon Trim28 knockdown (Figures 5G–5I).

Chaf1a Has Differential Regulatory Roles on Class I, II, and III ERVs

Venn diagram analysis on ERVs bound by Chaf1a, Trim28, and those associated with the H3K9me3 modification revealed that about 64% of ERVs co-bound by Chaf1a and Trim28 are enriched with H3K9me3 (Figure 6A). In comparison, only 23% of Chaf1a/Trim28 bound non-ERV loci are marked with H3K9me3 (Figure S6A). This concurs with the notion that Chaf1a and Trim28 exert ERV-specific repressive functions. In particular, there are significant numbers of ERVs co-bound by Chaf1a and Trim28, or exclusively bound by Chaf1a that are not marked with H3K9me3, suggesting that Chaf1a may adopt alternative repressive mechanisms on these ERVs. To this end, we classified the ERVs into four categories, namely, those bound by Chaf1a+Trim28+H3K9me3, Chaf1a+Trim28, Chaf1a only, and Trim28 only (Table S5). Interestingly, the Chaf1a only category has the highest percentage of class III ERVs (Figure 6B), while the Chaf1a+Trim28+H3K9me3 category primarily belong to class I and class II ERVs (Figure 6B). Consequentially, the dot plots (Figures 6C and S6B) correlating ERV upregulation and the enrichment of histone marks further highlighted the low levels of H3K9me3 on Chaf1a-regulated class III ERVs.

Specific class III ERVs are highly expressed in early embryonic development and downregulated at the morula and blastula stages. Histone demethylase Kdm1a (Macfarlan et al., 2012) and H3K9 dimethyl transferase G9a are the key epigenetic regulators of these ERVs (Leung et al., 2011; Maksakova et al., 2013). It was found that Kdm1a and histone deacetylase Hdac1/2 cooperatively contribute to transcriptional silencing (Shi et al., 2004). Hdacs have been shown to repress MERVL in concert with Kdm1a in pluripotent stem cells (Macfarlan et al., 2011; Reichmann et al., 2012). Interestingly, Kdm1a is one of the candidate hits in our siRNA screen (Table S1). To further dissect the mode of ERV regulation within each of the four categories, we integrated our Chaf1a and Trim28 ChIP-seq data with datasets for epigenetic factors, such as Kdm1a and Hdac1/2. Surprisingly, the ERVs from the Chaf1a only category display the highest enrichment of Kdm1a and Hdac1/2 in comparison to the other categories (Figures 6D and 6E). In contrast, the ERVs bound by Chaf1a+Trim28+H3K9me3 exhibit low levels of Kdm1a and Hdac1/2 binding (Figures 6D–6F and S6D). Consistently, the Chaf1a only category is characterized by significantly higher levels of H3K4me2, H3K9Ac, and H3K27Ac marks, which are the substrates of Kdm1a and Hdacs, respectively (Figures 6D and S6C). We further performed ERV expression analysis using a published mESC *Kdm1a* knockdown RNA-seq dataset (Agarwal et al., 2015). *Kdm1a* knockdown resulted in mostly class I and III ERV upregulation, in a manner similar to *Chaf1a* knockdown (Figure S6E). In terms of ERVs regulated, Kdm1a/Chaf1a knockdown has 80% more overlap than Kdm1a/Trim28 knockdown (Figures S6F and S6G).

Overall, our data indicates that Chaf1a regulates class I, II, and III ERVs through vastly different mechanisms, which may depend on the co-regulators.

Chaf1a Represses Proviruses through Epigenetic Co-factors

Chaf1a is the core component of the chromatin assembling factor complex (Caf1) that also includes Rbbp4. Interestingly, only Chaf1a/b exhibited a proviral silencing function, while the knockdown of *Rbbp4* had no effect (Figures S7A and S7B). Moreover, our siRNA screen did not uncover other histone chaperones necessary for retroviral silencing, further highlighting the specificity of Chaf1a/b in this process (Figure S7B). To further delineate the function of Chaf1a, we performed a pull-down of Flag-tagged Chaf1a followed by stable isotope labeling using amino acids (SILAC)-based quantitative mass spectrometry (MS) analysis (Figure 7A). The complete list of Chaf1a-interacting proteins includes several known and unknown factors (Figure 7A; Table S6). Chaf1a has previously been shown to interact with chromatin modifying factors (Quivy et al., 2004; Sarraf and Stancheva, 2004). Indeed, we identified several epigenetic modifiers that appeared in both the Chaf1a MS and genome-wide siRNA screen list, such as Kdm1a, Smarcc1, and Eset. Using co-immunoprecipitation (coIP), we confirmed the interaction of Chaf1a with histone methyltransferase Eset, histone de-methylase Kdm1a, deacetylase Hdac2, and histone chaperones Chaf1b (Figures 7B–7D, S7C, and S7D).

To investigate the direct effects of Chaf1a at provirus loci, we used the Chaf1a-3xHA CRISPR F9 cell line for ChIP-qPCR analysis. We observed direct localization of Chaf1a to the proviral LTR elements (Figure 7E), which was further confirmed by Chaf1a-V5 ChIP (Figure S7E). To address the relationship between Chaf1a and Trim28, we performed ChIP on Trim28 upon *Chaf1a* knockdown. The binding of Trim28 was significantly abolished by the knockdown of *Trim28* itself, whereas Chaf1a-knockdown elicited no effect (Figure S7F). This suggests that Trim28 recruitment to the provirus is independent of Chaf1a. Moreover, we did not detect any change in Chaf1a enrichment upon *Sumo2* depletion (Figure S7G).

To understand the mechanisms by which Chaf1a silences the newly introduced proviruses, we performed ChIP on the Chaf1a interacting histone modifiers Kdm1a and Hdac2. To our surprise, both Kdm1a and Hdac2 were enriched at the proviral LTR (Figures 7F and 7G). In addition, consistent with the siRNA screen, shRNA knockdown of *Kdm1a* was able to rescue the expression of MMLV-*Gfp* reporter (Figure S7H). Treatment of E14 cells using the Hdac inhibitor TSA also relieved silencing of the MMLV-*Gfp* reporter (Figure S7I). Next, we tested the dynamic changes of the histone marks on the provirus LTR and ERVs upon the depletion of *Chaf1a*. The enrichment of H3K9me3 on provirus LTR was slightly reduced (Figure 7H), while the active H3K4me3 and total H3Ac marks were significantly increased (Figures 7I, 7J, and S7J–S7L). Together, our data shows that the repressive function of Chaf1a on proviruses is reinforced by the presence of its interacting partners, Kdm1a, Hdac2, and Eset.

To test whether Chaf1a can directly bind the viral DNA, we performed electrophoretic mobility shift assays (EMSA). We did not observe a specific EMSA band for the Chaf1a protein, indicating that Chaf1a does not bind directly to the viral DNA (Figures S7M–S7O). The Caf1 complex is thought to assemble histones H3/H4 during DNA replication and repair

(Gaillard et al., 1996; Kaufman et al., 1995). Other studies have indicated that histone chaperones Asf1a/b work synergistically with Caf1 (Tyler et al., 1999). Our proteomics data also identified Asf1a/b as components of the Chaf1a interactome (Figure 7A), and the interaction between Chaf1a and Asf1a/b was confirmed by coIP (Figure S7P). To further test the function of histone assembly on proviral silencing, we performed single and combinatorial shRNA knockdown of *Asf1a/b*. Surprisingly, combinatorial depletion of *Asf1a/b* induced strong *Gfp* reactivation to a level comparable to that observed following *Chaf1a* depletion (Figure S7Q), indicating functional redundancy between Asf1a and Asf1b. This data substantiates a possible role of histone assembly in the silencing of proviral elements and ERVs.

DISCUSSION

Mammalian genomes are cluttered with endogenous viral elements, vestiges of the long history of coevolution with retrotransposons that have shaped the genome. Complex mechanisms have evolved to manage these elements, restricting their expression and reactivation. Silencing of retroviruses also played a fortuitous role in the development of somatic cell reprogramming by transcription factors, as extinction of the reprogramming transgenes that occurs when fibroblasts revert to a pluripotent state is essential for the induced pluripotent stem cells to avoid oncogenic transformation and manifest their multi-lineage differentiation potential (Takahashi and Yamanaka, 2006). Our work provides insights into the role of the histone chaperone Chaf1a and sumoylation factor Sumo2 in the silencing of exogenous proviruses and ERVs. It supports a model where Chaf1a promote the deposition of histone H3/H4, thus marking the integrated proviral DNA for silencing, helping to localize the Chaf1a protein to the viral LTR region (Figure 7K). The binding and transcriptional repression of the proviral chromatin by Chaf1a is further reinforced via the enzymatic activities of Chaf1a-interacting proteins Eset, Kdm1a, and Hdac1/2, which modify proviral chromatin with the repressive histone mark H3K9me3 and reduce the acquisition of activating H3K4me3 and H3Ac marks (Figure 7K). In parallel, Sumo2 is required to play critical roles in the canonical Zfp809/Trim28/Eset complex via post-translational sumoylation of Trim28. Sumoylation enhances the recruitment of Trim28 to the proviral DNA, which in turn results in the modification of proviral chromatin with repressive histone H3K9me3 marks (Figure 7K). Our unbiased screen for factors involved in proviral silencing has thus revealed a complex set of genetic and epigenetic mechanisms by which exogenous proviruses and ERVs are transcriptionally silenced in pluripotent stem cells.

Cross-Talk between the Sumoylation Pathway and the Canonical Complex

Among the Sumo2-related candidates, Senp6 deconjugates Sumo2 from targeted proteins (Mukhopadhyay and Dasso, 2007), while the other factors are involved in covalent attachment of Sumo2 to the targeted proteins (Desterro et al., 1999; Geiss-Friedlander and Melchior, 2007; Gong et al., 1999; Hay, 2005; Johnson, 2004; Zhao, 2007). As such, it is tempting to speculate that the modification of key determinants by sumoylation or de-sumoylation may affect their capacity to silent the proviruses and ERVs. The cross-talk between chromatin modifying complex subunits (such as Trim28, Atf7ip, and Eset) and

sumoylation factors can be inferred from the overlap of target ERVs observed, as well as their close protein-protein interactions. Importantly, our study clarifies the mechanism by which Sumo2 targets the proviral elements and ERVs—through the sumoylation of Trim28. Furthermore, Sumo modification on other epigenetic factors may potentially help mediate heterochromatin formation. It will be of great interest to determine the proteome-wide set of sumoylated proteins in ESCs.

Regulation of Different Classes of ERVs

Our RNA-seq analysis indicates that Chaf1a/b and sumoylation factors regulate different families of ERVs. Localization of Chaf1a and Sumo2 at ERV loci was confirmed by ChIP-seq analysis. It is noteworthy that the pattern of the ERVs regulated by Chaf1a is distinct from that of the sumoylation machinery or chromatin-modifying factors (Trim28, Eset, and Zfp809). Interestingly, Chaf1a regulates a significant number of ERVs from class III that are not marked with H3K9me3, but instead are enriched for H3K4me2 and H3K27Ac. Moreover, Chaf1a works with the enzymatic epigenetic modifiers, including Kdm1a and Hdac2 at these class III ERVs. In addition, Chaf1a also cooperates with Trim28 to repress the ERVs by reinforcing high levels of the H3K9me3 on class I and II ERVs. Thus, our study highlights how a chaperone like Chaf1a regulates different classes of ERVs through distinct interacting co-factors.

Suppressive Function of Histone Chaperone Chaf1a/b on Newly Integrated Proviruses

Caf1 has been reported to have diverse functions, including epigenetic regulation, DNA damage repair, and DNA replication (Green and Almouzni, 2003; Kaufman et al., 1995; Poleshko et al., 2010; Shibahara and Stillman, 1999). More recently, Chaf1a was shown to be critical for maintaining the heterochromatin state through its interaction with HP1, MBD1, and Eset (Murzina et al., 1999; Reese et al., 2003; Sarraf and Stancheva, 2004). In fact, protein structure analysis of Chaf1a indicates a PXVXL pentapeptide motif at the N terminus, which allows Chaf1a to specifically interact with the HP1 chromo shadow domain (Thiru et al., 2004). Stable association of Chaf1a with HP1 proteins may lead to its retention in heterochromatin (Murzina et al., 1999). HP1 proteins are “readers” of repressive H3K9me3 marks and interact extensively with Eset. Intriguingly, our proteomics identified Eset, HP1α, HP1β, and HP1γ among the Chaf1a interactome. Remarkably, only the knockdown of *Chaf1a/b* was capable of rescuing the viral reporter, but not the knockdown of Rbbp4 (Figures S7A and S7B). Previous studies suggest that Rbbp4 complexes with Chaf1a/b in G1 phase. Notably, the epigenetic modification brought about by Chaf1a through HP1 or Caf1/Mbd1/Eset is S-phase-specific (Quivy et al., 2004; Sarraf and Stancheva, 2004).

How does a histone chaperone like Chaf1a localize to the proviral LTR and ERVs? Previous work has localized histone chaperones such as Hira and Daxx to the genomic sites where histones are deposited (Banaszynski et al., 2013; Elsässer et al., 2012). A recent publication also described the role of histone variants H3.3 in regulating ERVs (Elsässer et al., 2015). Indeed, our Chaf1a ChIP-seq shows the enrichment of Chaf1a at the genomic sites of downstream ERV targets. When we knockdown the upstream histone chaperones of Chaf1a (Asf1a/b), we observed the abolishment of the viral silencing effect of Chaf1a. Thus, we

speculate that its nucleosome assembly function may play a role in localizing Chaf1a to the integrated proviruses.

In conclusion, our work reveals the genome-wide compendium of players that mediate proviral silencing in mouse ESCs. Multiple pathways and multi-layered machineries are employed by pluripotent embryonic stem cells to maintain the silencing of proviruses and ERVs. Further studies aimed at dissecting the intricate mechanisms by which the various factors act will help fill the remaining gap in our understanding of proviral repression.

EXPERIMENTAL PROCEDURES

Genome-wide siRNA Screen

F9 cells were seeded at 6×10^5/well in 6-well tissue culture plates. Twelve hours later, MMLV virus was added into the wells with 8 mg/ml polybrene (107689, Sigma). Eight hours later, F9 cells were trypsinized into single cells and seeded onto individual well of 384-well plates (REF 781091, Greiner) that were pre-printed with Mouse siGENOME SMARTpool library (G-015000, Thermo Scientific Dharmacon) and contain DharmaFECT 1 (Thermo Scientific). Four days later, cells were fixed with 4% paraformaldehyde and cell nuclei were stained with Hoechst 33342 (Invitrogen). Images were acquired using the ImageXpress Ultra Confocal High Content Screening System (Molecular Devices). *Gfp* signal was quantified by the MetaXpress software (Molecular Devices). Both the siRNA screens were carried out in duplicates. The average of the duplicate Gfp signal was calculated by normalizing to both positive and negative controls using ScreenSifter software (Kumar et al., 2013). A cut-off threshold was set at value >2 SD from mean of negative controls, above which siRNA of 650 candidate genes significantly increase *Gfp* expression level. Based on the secondary screening, 303 high-confidence hits with Gfp signal (CtrlNorm value = (X − Avg(xcn))/(Avg(xcp) − Avg(xcn)) cut off above 0.45 were selected.

RNA-Sequencing

Total RNA was extracted as described in the Supplemental Experimental Procedures. DNA contamination was removed using a QIAGEN RNeasy Kit. The RNA samples were subject to mRNA selection, fragmentation, cDNA synthesis, and library preparation using a TruSeq RNA Sample Prep Kit (RS-122-2001, Illumina). Library quality was analyzed on a Bioanalyzer. High-throughput sequencing was performed on the Genome Analyzer IIx (Illumina).

ChIP and ChIP-Seq Assay

Chromatin was prepared according to the methods provided in the Supplemental Experimental Procedures. Chromatin extracts were immunoprecipitated using H3K4me3 (Abcam), H3Ac (Abcam), H3K9me3 (Abcam), Eset (Abcam), Trim28 (Bethyl), Sumo2 (Abcam), and HA (Santa Cruz) antibodies. Input and immunoprecipitation samples were analyzed by qPCR. All primers used are listed in Table S7. ChIP-seq libraries were prepared according to manufacturer's instructions (Illumina). High-throughput sequencing was performed on a Genome Analyzer IIx (Illumina).

Bioinformatics Analysis

See detailed information in the Supplemental Experimental Procedures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Yang Lin for helpful discussion. We thank Wong Chee Wai for technical assistance. H.L. is supported by Mayo Clinic Center for Individualized Medicine. S.P.G. is supported by NIH grant R01 CA30488. Y.-H.L. is supported by the A*Star Investigatorship research award. We are grateful to the Biomedical Research Council, Agency for Science, Technology and Research, Singapore for research funding.

References

Agarwal S, Macfarlan TS, Sartor MA, Iwase S. Sequencing of first-strand cDNA library reveals full-length transcriptomes. Nat Commun. 2015; 6:6002. [PubMed: 25607527]

Banaszynski LA, Wen D, Dewell S, Whitcomb SJ, Lin M, Diaz N, Elsässer SJ, Chapgier A, Goldberg AD, Canaani E, et al. Hira-dependent histone H3.3 deposition facilitates PRC2 recruitment at developmental loci in ES cells. Cell. 2013; 155:107–120. [PubMed: 24074864]

De Graeve F, Bahr A, Chatton B, Kedinger C. A murine ATFa-associated factor with transcriptional repressing activity. Oncogene. 2000; 19:1807–1819. [PubMed: 10777215]

Desterro JM, Rodriguez MS, Kemp GD, Hay RT. Identification of the enzyme required for activation of the small ubiquitin-like protein SUMO-1. J Biol Chem. 1999; 274:10618–10624. [PubMed: 10187858]

Elsässer SJ, Huang H, Lewis PW, Chin JW, Allis CD, Patel DJ. DAXX envelops a histone H3.3-H4 dimer for H3.3-specific recognition. Nature. 2012; 491:560–565. [PubMed: 23075851]

Elsässer SJ, Noh KM, Diaz N, Allis CD, Banaszynski LA. Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. Nature. 2015; 522:240–244. [PubMed: 25938714]

Feuer G, Taketo M, Hanecak RC, Fan H. Two blocks in Moloney murine leukemia virus expression in undifferentiated F9 embryonal carcinoma cells as determined by transient expression assays. J Virol. 1989; 63:2317–2324. [PubMed: 2704078]

Friedli M, Turelli P, Kapopoulou A, Rauwel B, Castro-Díaz N, Rowe HM, Ecco G, Unzu C, Planet E, Lombardo A, et al. Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. Genome Res. 2014; 24:1251–1259. [PubMed: 24879558]

Gaillard PH, Martini EM, Kaufman PD, Stillman B, Moustacchi E, Almouzni G. Chromatin assembly coupled to DNA repair: a new role for chromatin assembly factor I. Cell. 1996; 86:887–896. [PubMed: 8808624]

Gaudet F, Rideout WM 3rd, Meissner A, Dausman J, Leonhardt H, Jaenisch R. Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing. Mol Cell Biol. 2004; 24:1640–1648. [PubMed: 14749379]

Geiss-Friedlander R, Melchior F. Concepts in sumoylation: a decade on. Nat Rev Mol Cell Biol. 2007; 8:947–956. [PubMed: 18000527]

Gong L, Li B, Millas S, Yeh ET. Molecular cloning and characterization of human AOS1 and UBA2, components of the sentrin-activating enzyme complex. FEBS Lett. 1999; 448:185–189. [PubMed: 10217437]

Green CM, Almouzni G. Local action of the chromatin assembly factor CAF-1 at sites of nucleotide excision repair in vivo. EMBO J. 2003; 22:5163–5174. [PubMed: 14517254]

Hay RT. SUMO: a history of modification. Mol Cell. 2005; 18:1–12. [PubMed: 15808504]

Hu G, Kim J, Xu Q, Leng Y, Orkin SH, Elledge SJ. A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes Dev. 2009; 23:837–848. [PubMed: 19339689]

Huang, da W.; Sherman, BT.; Lempicki, RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009; 37:1–13. [PubMed: 19033363]

Johnson ES. Protein modification by SUMO. Annu Rev Biochem. 2004; 73:355–382. [PubMed: 15189146]

Kao SY, Calman AF, Luciw PA, Peterlin BM. Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. Nature. 1987; 330:489–493. [PubMed: 2825027]

Karimi MM, Goyal P, Maksakova IA, Bilenky M, Leung D, Tang JX, Shinkai Y, Mager DL, Jones S, Hirst M, Lorincz MC. DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. Cell Stem Cell. 2011; 8:676–687. [PubMed: 21624812]

Kaufman PD, Kobayashi R, Kessler N, Stillman B. The p150 and p60 subunits of chromatin assembly factor I: a molecular link between newly synthesized histones and DNA replication. Cell. 1995; 81:1105–1114. [PubMed: 7600578]

Kumar P, Goh G, Wongphayak S, Moreau D, Bard F. Screen-Sifter: analysis and visualization of RNAi screening data. BMC Bioinformatics. 2013; 14:290. [PubMed: 24088301]

Leung DC, Dong KB, Maksakova IA, Goyal P, Appanah R, Lee S, Tachibana M, Shinkai Y, Lehnertz B, Mager DL, et al. Lysine methyltransferase G9a is required for de novo DNA methylation and the establishment, but not the maintenance, of proviral silencing. Proc Natl Acad Sci USA. 2011; 108:5718–5723. [PubMed: 21427230]

Macfarlan TS, Gifford WD, Agarwal S, Driscoll S, Lettieri K, Wang J, Andrews SE, Franco L, Rosenfeld MG, Ren B, Pfaff SL. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. Genes Dev. 2011; 25:594–607. [PubMed: 21357675]

Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature. 2012; 487:57–63. [PubMed: 22722858]

Maksakova IA, Thompson PJ, Goyal P, Jones SJ, Singh PB, Karimi MM, Lorincz MC. Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. Epigenetics Chromatin. 2013; 6:15. [PubMed: 23735015]

Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC, Shinkai Y. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature. 2010; 464:927–931. [PubMed: 20164836]

Maxwell PH, Curcio MJ. Host factors that control long terminal repeat retrotransposons in Saccharomyces cerevisiae: implications for regulation of mammalian retroviruses. Eukaryot Cell. 2007; 6:1069–1080. [PubMed: 17496126]

Mukhopadhyay D, Dasso M. Modification in reverse: the SUMO proteases. Trends Biochem Sci. 2007; 32:286–295. [PubMed: 17499995]

Murzina N, Verreault A, Laue E, Stillman B. Heterochromatin dynamics in mouse cells: interaction between chromatin assembly factor 1 and HP1 proteins. Mol Cell. 1999; 4:529–540. [PubMed: 10549285]

Niwa O, Yokota Y, Ishida H, Sugahara T. Independent mechanisms involved in suppression of the Moloney leukemia virus genome during differentiation of murine teratocarcinoma cells. Cell. 1983; 32:1105–1113. [PubMed: 6188535]

Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, Knowles BB. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell. 2004; 7:597–606. [PubMed: 15469847]

Petersen R, Kempler G, Barklis E. A stem cell-specific silencer in the primer-binding site of a retrovirus. Mol Cell Biol. 1991; 11:1214–1221. [PubMed: 1996087]

Poleshko A, Einarson MB, Shalginskikh N, Zhang R, Adams PD, Skalka AM, Katz RA. Identification of a functional network of human epigenetic silencing factors. J Biol Chem. 2010; 285:422–433. [PubMed: 19880521]

Quivy JP, Roche D, Kirschner D, Tagami H, Nakatani Y, Almouzni G. A CAF-1 dependent pool of HP1 during heterochromatin duplication. EMBO J. 2004; 23:3516–3526. [PubMed: 15306854]

Reese BE, Bachman KE, Baylin SB, Rountree MR. The methyl-CpG binding protein MBD1 interacts with the p150 subunit of chromatin assembly factor 1. Mol Cell Biol. 2003; 23:3226–3236. [PubMed: 12697822]

Reichmann J, Crichton JH, Madej MJ, Taggart M, Gautier P, Garcia-Perez JL, Meehan RR, Adams IR. Microarray analysis of LTR retrotransposon silencing identifies Hdac1 as a regulator of retrotransposon expression in mouse embryonic stem cells. PLoS Comput Biol. 2012; 8:e1002486. [PubMed: 22570599]

Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, Maillard PV, Layard-Liesching H, Verp S, Marquis J, et al. KAP1 controls endogenous retroviruses in embryonic stem cells. Nature. 2010; 463:237–240. [PubMed: 20075919]

Sarraf SA, Stancheva I. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Mol Cell. 2004; 15:595–605. [PubMed: 15327775]

Schlesinger S, Goff SP. Silencing of proviruses in embryonic cells: efficiency, stability and chromatin modifications. EMBO Rep. 2013; 14:73–79. [PubMed: 23154467]

Schlesinger S, Lee AH, Wang GZ, Green L, Goff SP. Proviral silencing in embryonic cells is regulated by Yin Yang 1. Cell Rep. 2013; 4:50–58. [PubMed: 23810560]

Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA, Shi Y. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell. 2004; 119:941–953. [PubMed: 15620353]

Shibahara K, Stillman B. Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. Cell. 1999; 96:575–585. [PubMed: 10052459]

Shoemaker JE, Lopes TJ, Ghosh S, Matsuoka Y, Kawaoka Y, Kitano H. CTen: a web-based platform for identifying enriched cell types from heterogeneous microarray data. BMC Genomics. 2012; 13:460. [PubMed: 22953731]

Stewart CL, Stuhlmann H, Jähner D, Jaenisch R. De novo methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. Proc Natl Acad Sci USA. 1982; 79:4098–4102. [PubMed: 6955793]

Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006; 126:663–676. [PubMed: 16904174]

Teich NM, Weiss RA, Martin GR, Lowy DR. Virus infection of murine teratocarcinoma stem cell lines. Cell. 1977; 12:973–982. [PubMed: 202395]

Thiru A, Nietlispach D, Mott HR, Okuwaki M, Lyon D, Nielsen PR, Hirshberg M, Verreault A, Murzina NV, Laue ED. Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. EMBO J. 2004; 23:489–499. [PubMed: 14765118]

Tyler JK, Adams CR, Chen SR, Kobayashi R, Kamakaka RT, Kadonaga JT. The RCAF complex mediates chromatin assembly during DNA replication and repair. Nature. 1999; 402:555–560. [PubMed: 10591219]

Walsh CP, Chaillet JR, Bestor TH. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet. 1998; 20:116–117. [PubMed: 9771701]

Wang H, An W, Cao R, Xia L, Erdjument-Bromage H, Chatton B, Tempst P, Roeder RG, Zhang Y. mAM facilitates conversion by ESET of dimethyl to trimethyl lysine 9 of histone H3 to cause transcriptional repression. Mol Cell. 2003; 12:475–487. [PubMed: 14536086]

Wang GZ, Wolf D, Goff SP. EBP1, a novel host factor involved in primer binding site-dependent restriction of moloney murine leukemia virus in embryonic cells. J Virol. 2014; 88:1825–1829. [PubMed: 24227866]

Wissing S, Muñoz-Lopez M, Macia A, Yang Z, Montano M, Collins W, Garcia-Perez JL, Moran JV, Greene WC. Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility. Hum Mol Genet. 2012; 21:208–218. [PubMed: 21989055]

Wolf D, Goff SP. TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. Cell. 2007; 131:46–57. [PubMed: 17923087]

Wolf D, Goff SP. Embryonic stem cells use ZFP809 to silence retroviral DNAs. Nature. 2009; 458:1201–1204. [PubMed: 19270682]

Wolf D, Cammas F, Losson R, Goff SP. Primer binding site-dependent restriction of murine leukemia virus requires HP1 binding by TRIM28. J Virol. 2008a; 82:4675–4679. [PubMed: 18287239]

Wolf D, Hug K, Goff SP. TRIM28 mediates primer binding site-targeted silencing of Lys1,2 tRNA-utilizing retroviruses in embryonic cells. Proc Natl Acad Sci USA. 2008b; 105:12521–12526. [PubMed: 18713861]

Ying QL, Stavridis M, Griffiths D, Li M, Smith A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat Biotechnol. 2003; 21:183–186. [PubMed: 12524553]

Zhao J. Sumoylation regulates diverse biological processes. Cell Mol Life Sci. 2007; 64:3017–3033. [PubMed: 17763827]

Highlights

- Genome-wide siRNA screen identifies key determinants for proviral silencing in ESCs
- Histone chaperones, sumoylation factors, and chromatin modifiers can repress ERVs
- Sumo2 orchestrates viral silencing through sumoylation modification of Trim28
- Chaf1a regulates provirus and ERVs via its interaction with Eset, Kdm1a, and Hdac1/2

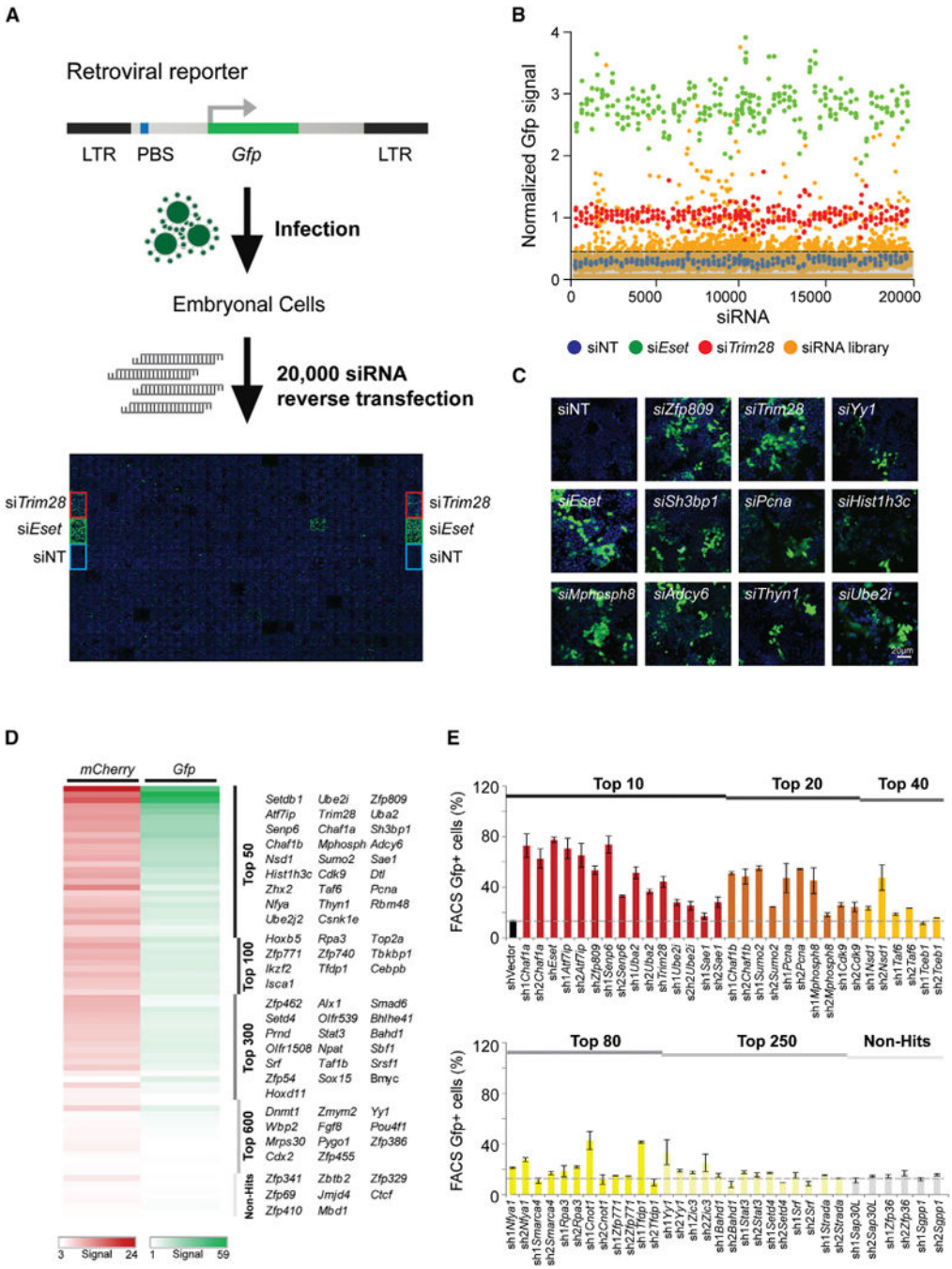


Figure 1. Genome-wide siRNA Screen for Regulators of Proviral Silencing in Mouse F9 ECs

(A) Schematic of the proviral MMLV-*Gfp* reporter assay. The map of the proviral reporter is shown (upper panel). LTR (black) indicates the long terminal repeats, while PBS (blue) represents the primer binding site. F9 cells were infected with the reporter virus and subjected to reverse transfection with the siRNA library in 384-well plates. A representative image for Gfp fluorescence (green) and nuclear Hoechst 33342 staining (blue) in a 384-well plate is shown. In each 384-well plate, non-targeting siRNA control (siNT) and positive control siRNA against *Trim28* and *Eset* (si*Trim28* and si*Eset*) were added.

(B) Dot plot for genome-wide siRNA screen. A cut-off threshold was set at 0.37 (dotted line). Candidate genes above the threshold showed significant Gfp reactivation.
(C) Representative images of *Gfp* rescue for selected hits from the genome-wide screen. Gfp (green) and Hoechst 33342 staining of the nucleus (blue) are shown.
(D) Secondary siRNA screen for 74 genes. Results for reactivation of proviral *Gfp* or *mCherry* reporters are shown as heatmaps. Intensity of green or red color represents the level of reactivation of *Gfp* and *mCherry* reporters respectively. See Supplemental Experimental Procedures for details on the gene selection criteria and experimental design.
(E) Validation of candidate genes using shRNA knockdown. *Gfp* signal was detected by FACS. The percentage of *Gfp* activation is shown on the y axis. Values are mean ± SEM from independent replicate experiments.
See also Figure S1 and Table S1.

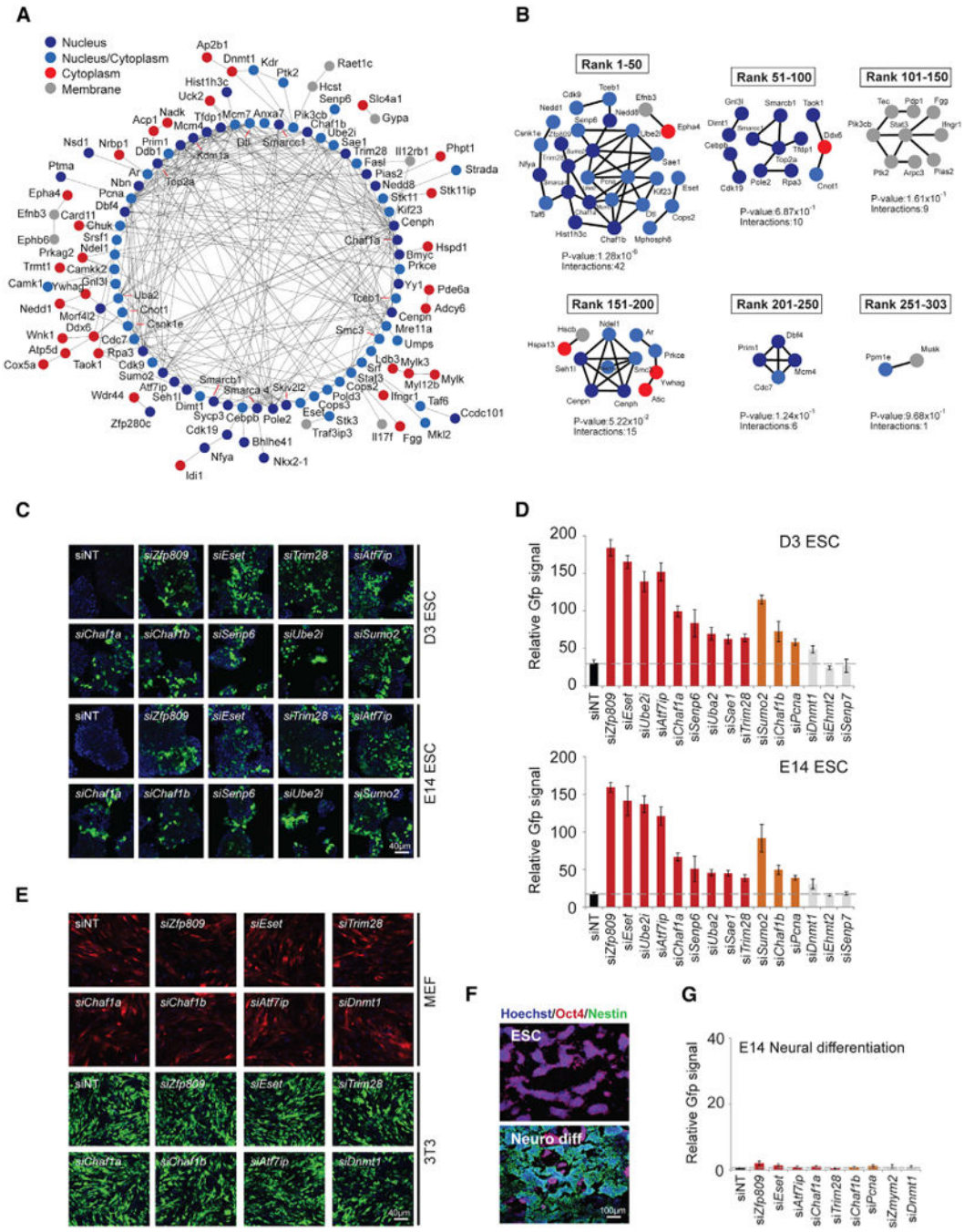


Figure 2. Bioinformatics Analyses for the Genome-wide siRNA Screen and the ESC Specificity of the Candidate Genes

(A) Interactome analysis. Cellular localization of the hits is indicated.

(B) Interactions observed in hits of different ranking tiers. Localization of hits is indicated as in (A). P values and number of interactions are indicated.

(C and D) Validation of MMLV-*Gfp* rescue by siRNA knockdown of the top candidates in D3 and E14 ESCs. Non-targeting siRNA (siNT) and siRNA targeting non-hits (*Dmnt1*, *Ehmt2*, *Senp7*) were selected as controls. (C) Representative images of *Gfp* rescue by siRNA

knockdown of the indicated hits. *Gfp* (green) and Hoechst 33342 nucleus staining (blue) are shown. (D) Bar chart graphs for *Gfp* activation. Relative *Gfp* signal is shown on the y axis. Values are mean ± SEM from independent replicate experiments.

(E) Representative images of MMLV-*mCherry* and MMLV-*Gfp* rescue by siRNA knockdown of selected top hits in MEF and 3T3 cells. *mCherry* (red), *Gfp* (green) and Hoechst 33342 nucleus staining (blue) are shown.

(F) Representative images for Oct4 and Nestin staining on E14 cells (upper panel) and E14 ESCs derived differentiated neural cells (lower panel).

(G) MMLV-*Gfp* rescue in E14-derived neural cells by siRNA knockdown of selected top hits. Relative *Gfp* signal is shown on the y axis. Values are mean ± SEM from independent replicate experiments.

See also Figure S2 and Table S2.

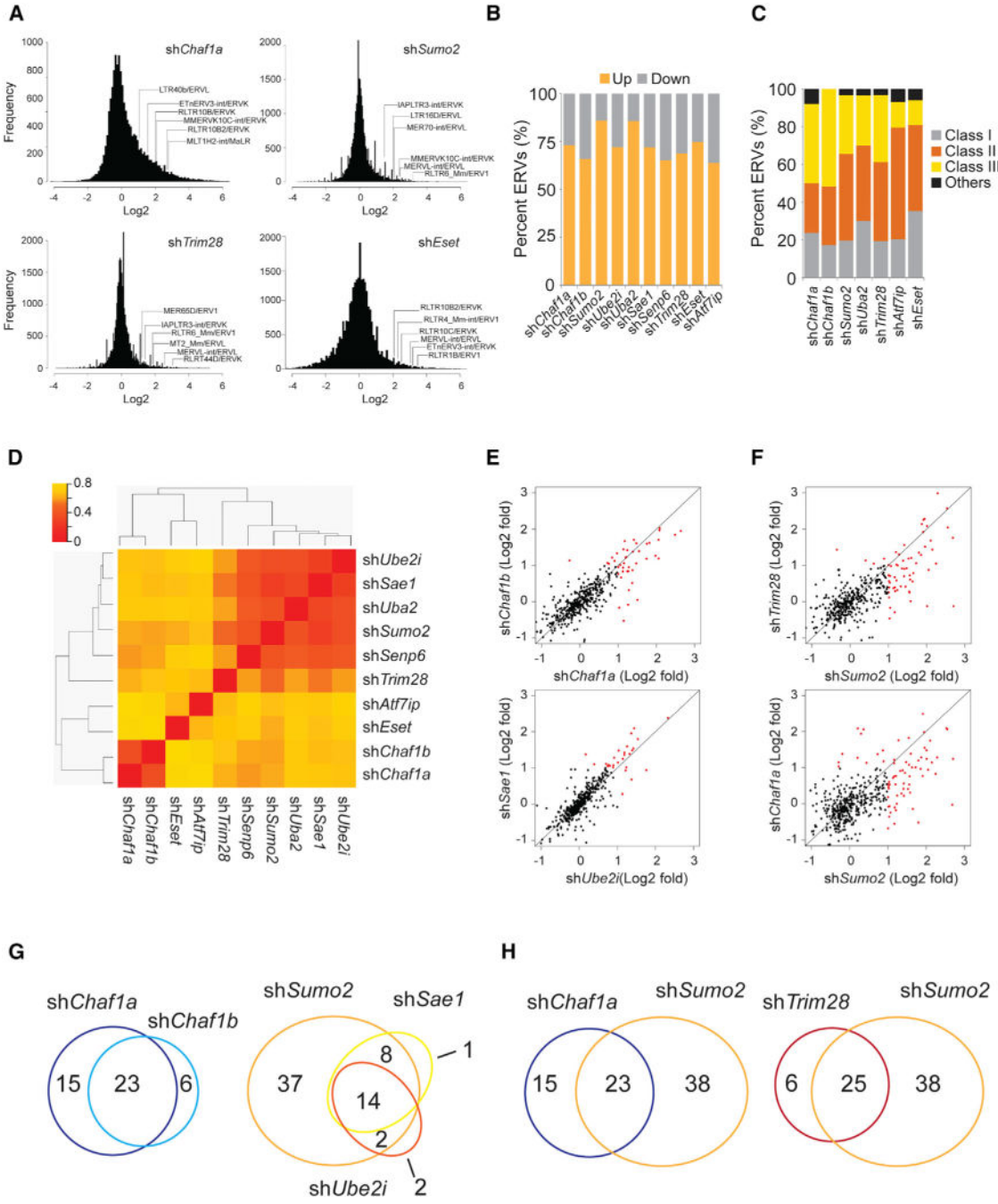


Figure 3. Histone Modifiers and Sumoylation Factors Regulate ERVs in Mouse Embryonic Stem Cells

(A) Frequency histogram of gene expression from RNA-seq data after *Chaf1a*, *Sumo2*, *Trim28*, or *Eset* depletion in E14 cells. Log2 fold change of expression levels is shown on the x axis. The number of genes at a given expression level is shown on the y axis.

(B) Percentage stacked columns indicating the up or downregulation of ERVs upon the depletion of the indicated factors.

(C) Percentage stacked columns indicating the classes of upregulated ERVs upon the depletion of the indicated factors.

(D) Clustering analysis of the indicated RNA-Seq libraries based on differential ERV expression. Heatmap color intensity signifies the correlation strength between 0 (red-high similarity) to 0.8 (yellow-high difference).
(E and F) Genome-wide de-regulation of ERVs in E14 cells after depletion of the indicated genes. RNA-seq data for RNAi samples and the shVector control were used to calculate the Log2 fold change values. Red dots indicate the elements with significantly increased expression.
(G and H) Venn diagrams demonstrating the number of commonly and differentially upregulated ERVs among the depletion of indicated factors.
See also Figure S3 and Table S3.

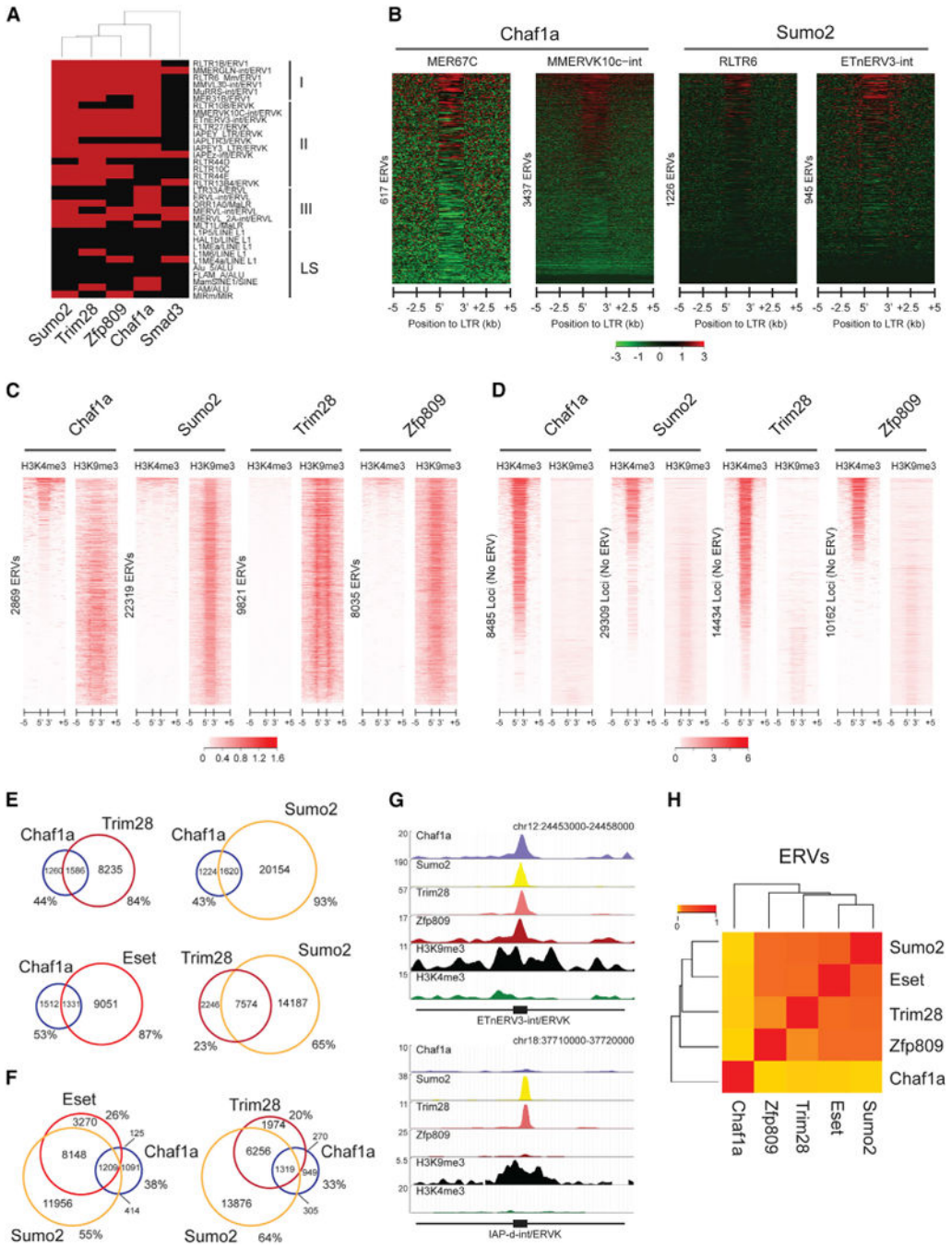


Figure 4. Direct Recruitment of Chaf1a and Sumo2 to Genomic ERVs

(A) Heatmap indicating the recruitment of Sumo2, Trim28, Zfp809, and Chaf1a on the indicated ERVs of different classes (I–III) and Line/Sine elements (LS). ChIP-seq was performed for the indicated factors, Smad3 is used as a control. Red indicates binding whereas black indicates the absence of binding.

(B) Heatmaps of Chaf1a enrichment at the genomic regions flanking MER67C and MMERVK10c-int (left panels) and Sumo2 enrichment at the genomic regions flanking RLTR6 and ETnERV3-int (right panels).

(C) Heatmaps of histone modifications at the genomic regions of the ERV loci bound by Chaf1a, Sumo2, Trim28, and Zfp809. The heatmaps are clustered according to the enrichment profile of H3K4me3.

(D) Enrichment of several histone marks at the genomic regions of the non-ERV loci that are bound by indicated factors. The reads in the heatmaps are clustered according to the enrichment profile of H3K4me3.

(E and F) Venn diagrams demonstrating the number of commonly and uniquely-bound ERV loci among the indicated factors. Percentage values indicate uniquely bound sites.

(G) UCSC genome browser screenshots. Chaf1a, Sumo2, Trim28, and Zfp809 bind eTnERV3-int-ERVK, while IAP-d-int/ERVK is bound specifically by Sumo2 and Trim28. Both ERVs are enriched with H3K9me3.

(H) Clustering analysis of the ERVs bound by the indicated factors. The color intensity signifies strength of correlation. Red indicates strong correlation, whereas yellow indicates weak correlation.

See also Figure S4 and Table S4.

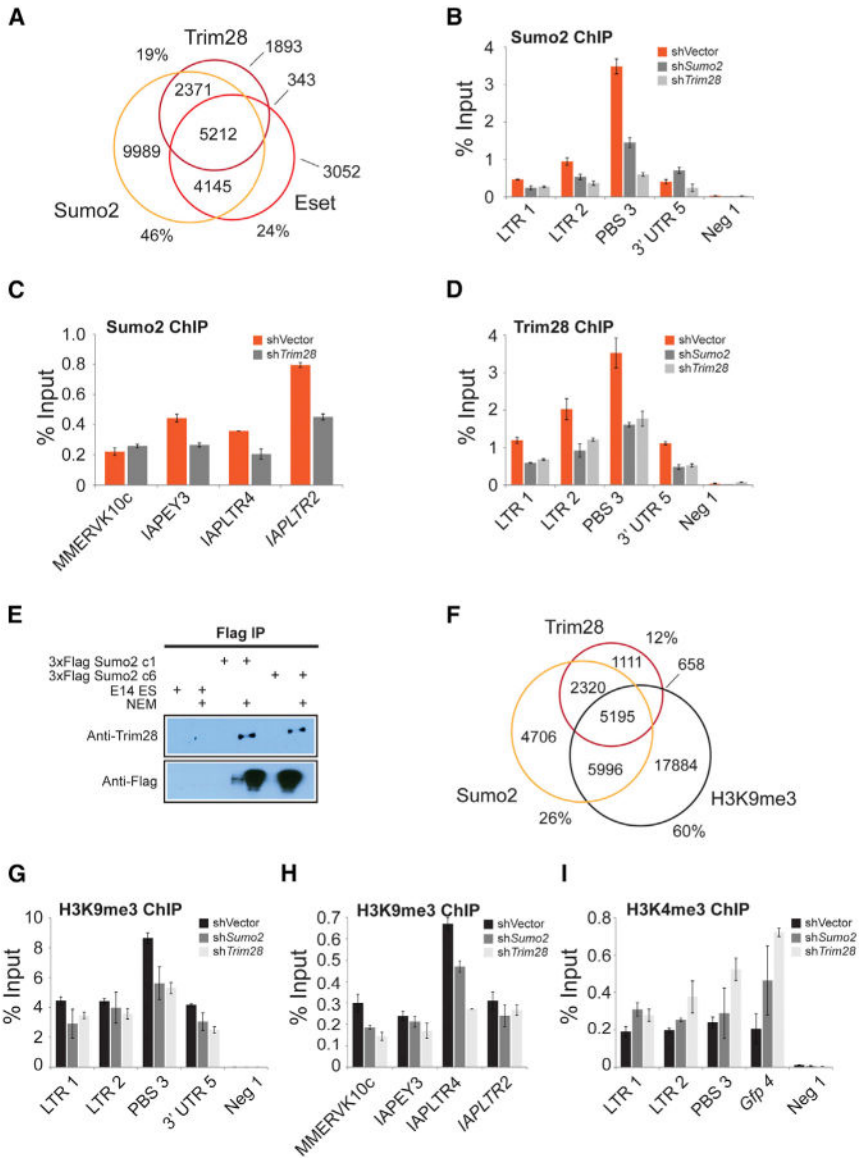


Figure 5. Sumo2 Regulates Proviruses by Post-translational Modification of Trim28

(A) Venn diagrams demonstrating the number of common and uniquely-bound ERV loci among the indicated factors. Sumo2 interacts extensively with the factors from the canonical pathway. Percentage values indicate uniquely bound sites.

(B–D) Sumo2 functions through Trim28 in proviral silencing. Sumo2 and Trim28 ChIP experiments were conducted on the samples with depletion of *Sumo2* or *Trim28*. The enrichment was measured by qPCR. Data is presented as mean ± SEM from independent replicate experiments.

(E) Trim28 is modified by Sumo2 in vivo. A 3xFlag tag was added to the 5′ end of *Sumo2* genomic region using CRISPR/Cas in E14 cells. Two homozygous lines were selected for the immunoprecipitation assays. NEM was added to protect the sumoylated proteins from desumoylation by SENPs in the cell lysates.

(F) Venn diagrams demonstrating the number of common and uniquely-bound ERV loci among indicated factors. The majority of the Trim28/H3K9me3 enriched ERVs are also bound by Sumo2. Percentage values indicate uniquely-bound sites.
(G and H) Knockdown of *Sumo2* and *Trim28* significantly reduced the H3K9me3 enrichment on proviral PBS and ERVs. H3K9me3 ChIP was performed on the samples with depletion of *Sumo2* or *Trim28*. Data is presented as mean ± SEM from independent replicate experiments.
(I) Knockdown of *Trim28* and *Sumo2* increased the active H3K4me3 mark on proviral elements. H3K4me3 ChIP was performed on samples with depleted *Sumo2* or *Trim28*. Data is presented as mean ± SEM from independent replicate experiments.
See also Figure S5.

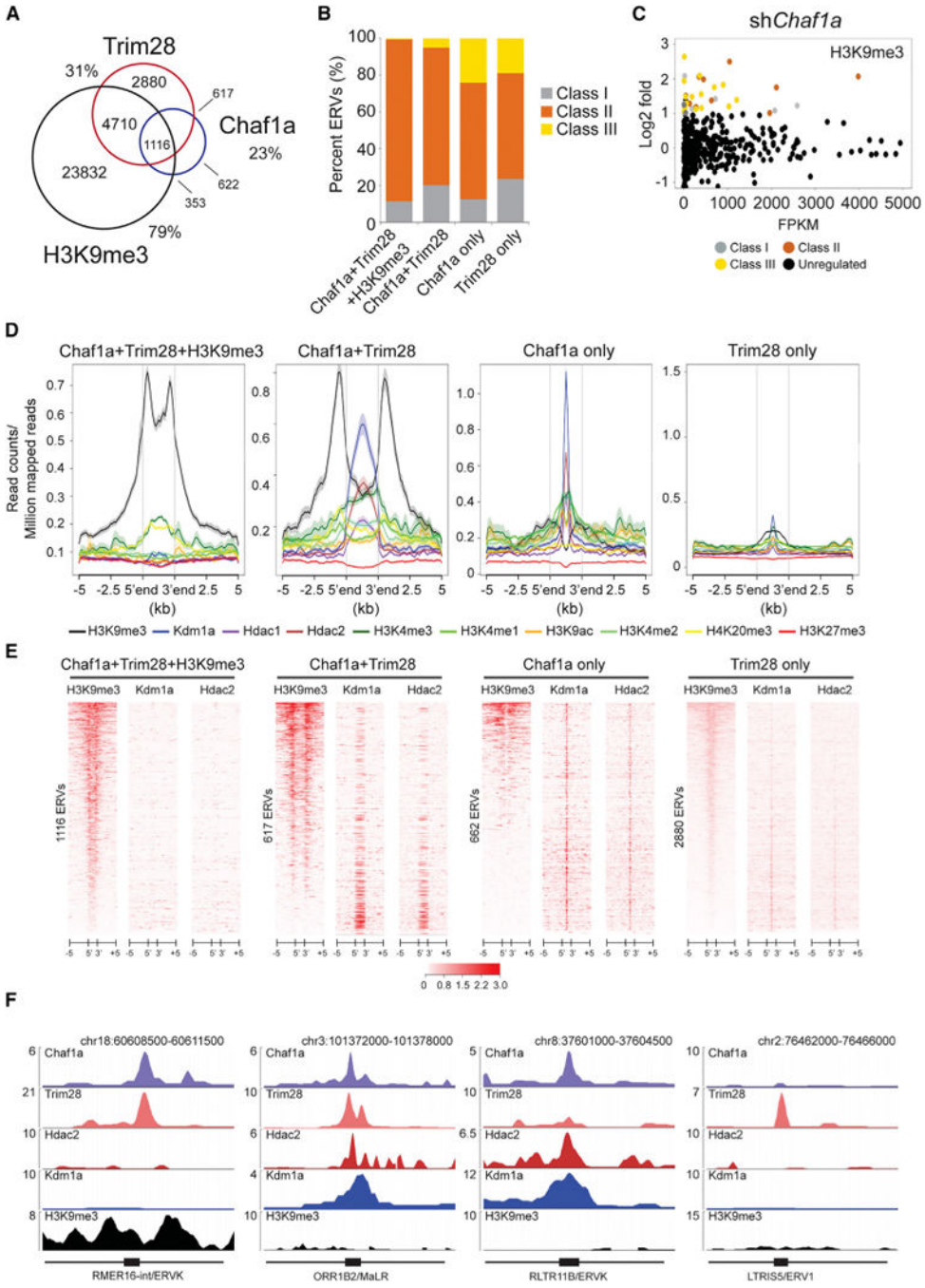


Figure 6. Differential Regulation of Class I, II, and III ERVs by Chaf1a

(A) Venn diagrams demonstrating the number of common and uniquely-bound ERV loci among Chaf1a, Trim28, and H3K9me3. Percentage values indicate uniquely-bound ERVs.

(B) Percentage stacked columns demonstrating the classes of ERVs bound by the indicated categories on the x axis.

(C) The correlation between the upregulation of the different classes of ERVs upon *Chaf1a* depletion and the enrichment of H3K9me3 mark. The data is plotted using sh*Chaf1a* RNA-seq and H3K9me3 ChIP-seq. Grey, orange, and yellow dots represent ERVs with

significantly increased expression in class I, II, and III, respectively. Black dots indicate the non-regulated ERVs.

(D) Average binding profiles of the individual categories shows that ERVs belonging to the Chaf1a only and Chaf1a+Trim28 categories are highly enriched with Kdm1a and Hdac2 in comparison to the other categories.

(E) Enrichment of H3K9me3, Kdm1a, and Hdac2 in the genomic regions of the indicated categories. The reads in the heatmaps are clustered according to the enrichment profile of H3K9me3.

(F) UCSC genome browser screenshots of representative repeat elements. RMER16-int bound by Chaf1a and Trim28 is highly enriched with H3K9me3. In contrast, ORR1B2 is bound by Chaf1a, Trim28, Hdac2, and Kdm1a with very low H3K9me3 enrichment. Chaf1a, Hdac2, and Kdm1a bind RLTR11B with the absence of Trim28 and H3K9me3, while LTRIS5 is bound exclusively by Trim28.

See also Figure S6 and Table S5.

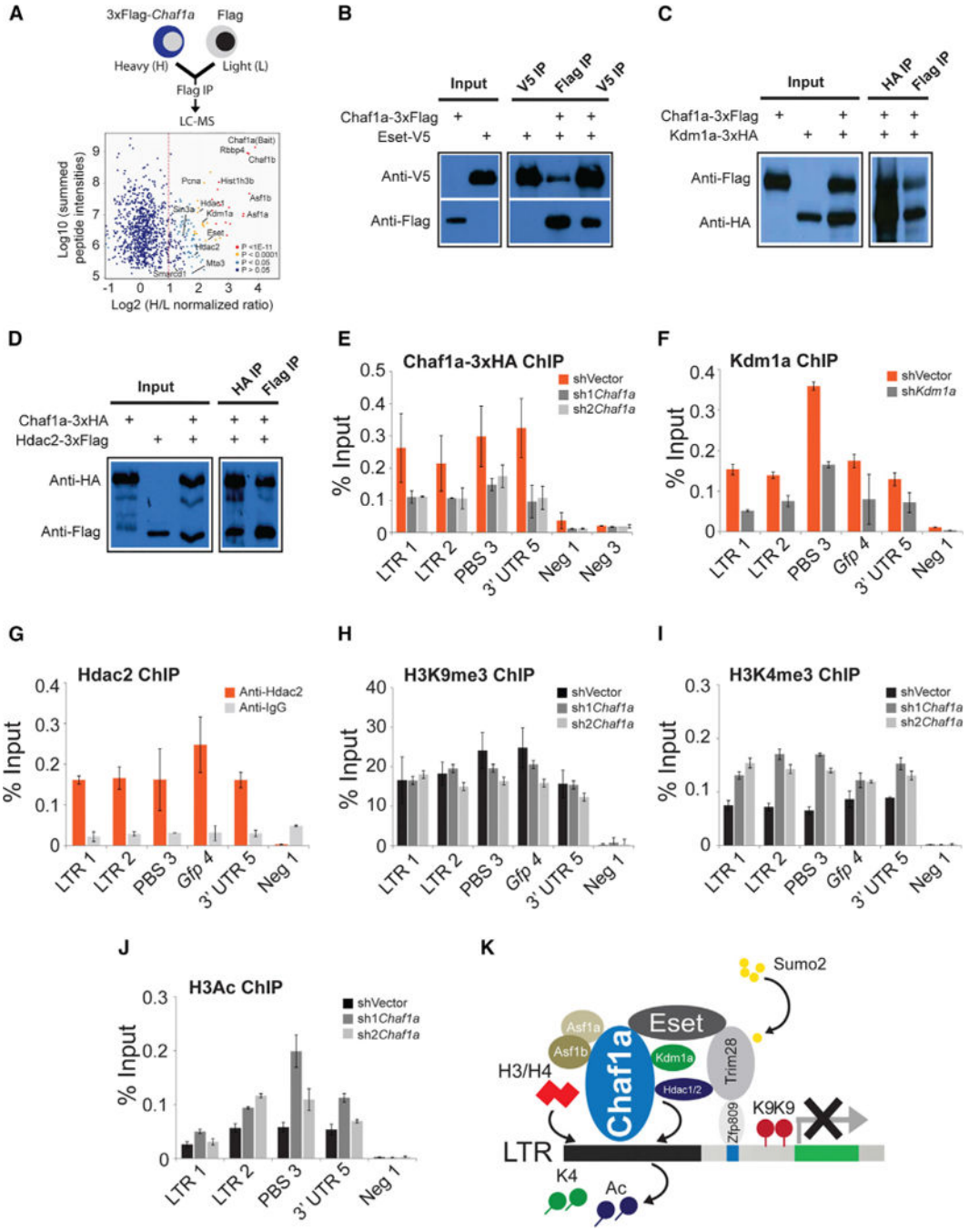


Figure 7. Chaf1a Are Enriched at Proviruses and Regulates Their Expression through Its Interacting Epigenetic Co-factors

(A) SILAC mass spectrometry (MS) analysis uncovers the Chaf1a interactome network. Upper panel: schematic representation of the SILAC MS work-flow as described in the supplemental procedures. Lower panel: differential protein identification in Flag-tagged Chaf1a immunoprecipitation. Several epigenetic and chromatin regulators are indicated. (B–D) Western blots confirm the interacting proteins identified by MS. Western blots showing co-immunoprecipitation (coIP) of Chaf1a with Eset, Kdm1a, and Hdac2.

(E) Chaf1a is enriched at the proviral elements. Chaf1a-3xHA ChIP was carried out in F9 Chaf1a-3xHA cell line using a HA antibody. The enrichment was analyzed by qPCR. Data are presented as mean ± SEM from independent replicate experiments.
(F and G) Localization of Kdm1a and Hdac2 on proviral DNA. ChIP was performed using antibodies against Kdm1a or Hdac2 and the enrichment was tested by qPCR.
(H–J) The perturbation of histone mark enrichment on proviral elements upon the depletion of *Chaf1a* in F9 cells. H3K9me3, H3K4me3, and H3Ac ChIP were performed on the samples upon depletion of *Chaf1a*. Data are presented as mean ± SEM from independent replicate experiments.
(K) Schematic model for the silencing mechanism of the proviruses in mESCs involving Chaf1a, Sumo2, and the canonical Zfp809/Trim28/Eset pathway. Chaf1a and its upstream histone chaperones Asf1a/b promote the deposition of histone H3/H4 to mark the integrated proviral DNA. Transcriptional repression of the proviral chromatin is reinforced by the enzymatic activities of Chaf1a-interacting proteins, including the members of the NuRD complex (Kdm1a, Hdac1/2) and Eset. This results in reduced acquisition of activating H3K4me3 and H3Ac marks. In parallel, Sumo2 sumoylates Trim28, which is necessary for recruiting Trim28 onto the proviral DNA, in turn resulting in the deposition of the repressive H3K9me3 mark.
See also Figure S7 and Table S6.