NP220 mediates silencing of unintegrated retroviral DNA

Yiping Zhu1,2,3,4, Gary Z. Wang2,3,4, Oya Cingöz2,3, and Stephen P. Goff2,3,*

The entry of foreign DNA into many mammalian cell types triggers the innate immune system, a complex set of responses to prevent infection by pathogens. One aspect of the response is the potent epigenetic silencing of incoming viral DNAs5, including the extrachromosomal DNAs that are formed immediately after infection by retroviruses. These unintegrated viral DNAs are very poorly transcribed in all cells, even in permissive cells, in contrast to the robust expression that is observed after viral integration2–3. The factors that are responsible for this low expression have not yet been identified. Here we performed a genome-wide CRISPR–Cas9 screen for genes that are required for silencing an integrase-deficient MLV–GFP reporter virus to explore the mechanisms responsible for repression of unintegrated viral DNAs in human cells. Our screen identified the DNA-binding protein NP220, the three proteins (MPP8, TASOR and PPHLN1) that comprise the HUSH complex—which silences proviruses in heterochromatin6 and retrotransposons7,8—the histone methyltransferase SETDB1, and other host factors that are required for silencing. Further tests by chromatin immunoprecipitation showed that NP220 is the key protein that recruits the HUSH complex, SETDB1 and the histone deacetylases HDAC1 and HDAC4 to silence the unintegrated retroviral DNA. Knockout of NP220 accelerates the replication of retroviruses. These experiments identify the molecular machinery that silences extrachromosomal retroviral DNA.

Retroviral infections begin with the reverse transcription of the RNA genome in the cytoplasm to form a linear double-stranded DNA that is soon delivered into the nucleus9. A portion of this linear DNA gives rise to two circular DNA forms that contain one or two tandem copies of the long terminal repeat (LTR) sequences at the ends of the linear DNA10,11. The linear DNA is inserted into the host genome to form the integrated provirus, which is actively transcribed and produces progeny virus. By contrast, the unintegrated DNAs are transcriptionally silent, do not replicate and disappear over time. To analyse the mechanism of silencing of unintegrated retroviral DNAs, we infected HeLa cells with integrase-deficient (IN(D184A)) or integrase-proficient (IN(WT)) MLV–Luc viruses and monitored expression of the luciferase reporter. Comparable levels of retroviral DNA were produced, but expression of the reporter was approximately 30-fold lower in cells infected with the IN(D184A) virus than those infected with the IN(WT) virus (Extended Data Fig. 1a, b). Treatment of the cells with the histone deacetylase (HDAC) inhibitor trichostatin A (TSA) markedly increased luciferase expression of unintegrated IN(D184A) MLV–Luc DNA without measureable effect on the integrated wild-type control virus (Extended Data Fig. 1c). Chromatin immunoprecipitation (ChIP) experiments showed that H3 histones on unintegrated viral DNA were largely deacetylated and carried repressive H3K9me3 marks but not H3K27me3 marks. By contrast, H3 histones on integrated retroviral DNA were heavily acetylated and barely methylated at H3K9 or H3K27 (Extended Data Fig. 1e–j).

To identify host factors responsible for the silencing of unintegrated retroviral DNA, we performed an unbiased genome-wide CRISPR–Cas9 screen, selecting for the knockout of host factor genes that relieve the silencing (Fig. 1a). HeLa–Cas9 cells transduced with a CRISPR single-guide RNA (sgRNA) library were infected with an integrase-deficient MLV–GFP reporter virus and GFP+ cells were isolated by sorting (Fig. 1b). After a second round of selection, targeted genes, the knockout of which was increased in the selected cells, were identified by sequencing of sgRNAs. Five genes stood out: NP220 (also known as ZNF638), all three subunits of the HUSH complex (MPP8 (also known as MPHOSPH8), TASOR (also known as FAM208A) and ZNF638) and FAM208A (also known as FAM199A1). Functional analysis using the CRISPR single-guide RNA library showed that knocking out these genes resulted in increased luciferase expression (Extended Data Fig. 1c, d).

Fig. 1 | CRISPR–Cas9 screen to identify host factors responsible for silencing of unintegrated retroviral DNA. a, Flowchart of the genome-wide CRISPR–Cas9 screening strategy. A pooled collection of HeLa knockout (KO) cells was infected with an integrase-deficient MLV–GFP virus and the 5% brightest GFP+ cells were isolated by sorting. These cells were subjected to a second round of infection and selection and DNAs were recovered for analysis of the resident sequences that encode the CRISPR single-guide RNAs (sgRNAs). b, GFP signals detected by fluorescence-activated cell sorting during round 1 and round 2 sorting. c, Candidate genes that are essential for the silencing were identified by analysing the fold change in abundance over control and the number of enriched sgRNAs per gene using HitSelect software. a–c, The CRISPR screen (a), GFP monitoring experiment (b) and sgRNA analysis (c) were performed only once. d, Validation of candidate genes from the screen. HeLa cells were transfected with the indicated siRNAs and then infected with an integrase-deficient (IN(D184A)) MLV–Luc virus. Luciferase activities were measured and activity in cells transfected with non-targeting (NT) siRNA was set to 1. Data are mean ± s.d.; n = 3 independent experiments.
The HUSH complex was previously identified to have a role in silencing binding domain and the zinc finger of NP220 are required for silencing. NP220 knockout cells (Fig. 2c). Therefore, both the DNA-binding domain is deleted; MPP8(W80A), MPP8 in which a mutation (W80A) renders the zinc-finger domain (DB) and a single C-terminal C2H2-type zinc finger (ZnF) motif (Fig. 2a). Knockdown or knockout of NP220 resulted in a marked decrease of wild-type NP220 or the NP220(ΔZnF) mutant restored binding, whereas expression of the NP220(ΔDB) mutant did not (Extended Data Fig. 3c). Notably, NP220 bound to DNA even when expression of any of the HUSH subunits or SETDB1 were very low (Fig. 3a) and neither was detected to be interacting with NP220 knockout cells, the levels of MPP8 and TASOR were increased (Fig. 3a) and immunoprecipitation of NP220 resulted in correspondingly increased co-immunoprecipitation of both MPP8 and TASOR (Fig. 3a).

ChIP assays were performed to monitor binding of these proteins to unintegrated viral DNA. NP220 was able to bind to both linear and circular unintegrated DNA (Fig. 3b and Extended Data Fig. 3e). NP220 knockout eliminated the ChIP signal (Fig. 3b) and re-expression of wild-type NP220 or the NP220(ΔZnF) mutant restored binding, because re-expression of NP220(W80A), which lacks methyl-binding activity, in the MPP8 knockout line showed only weak binding to viral DNA (Extended Data Fig. 3d). None of the HUSH subunits or SETDB1 bound to viral DNA in NP220 knockout cells (Fig. 3c–f), suggesting that NP220 has a key role in bringing the HUSH complex and SETDB1 to DNA.

Knockout of NP220, MPP8 or SETDB1 significantly decreased H3K9me3 modification on unintegrated retroviral DNA (Fig. 3g), indicating that NP220, the HUSH complex and SETDB1 are all required for H3K9 trimethylation on unintegrated retroviral DNA. Knockdown of HDAC1 or HDAC4 also increased the expression of unintegrated retroviral DNA (Fig. 4a, b) and increased the levels of acetylated histone H3 (Fig. 4c). Co-immunoprecipitation assays showed that endogenous NP220 bound to HDAC4, but not other HDACs (Fig. 4d and Extended Data Fig. 4a). The NP220–HDAC4 interaction was independent of DNA or RNA (Extended Data Fig. 4b). The N-terminal 471 amino acids of NP220 were sufficient to mediate the co-immunoprecipitation with MPP8 (Extended Data Fig. 3b). In MPP8 knockout cells, NP220 did not co-immunoprecipitate either MPP8 or TASOR (Fig. 3a), consistent with MPP8 serving as the direct partner of NP220. In TASOR knockout cells, levels of both TASOR and MPP8 were very low (Fig. 3a) and neither was detected to be interacting with NP220 (Fig. 3a). In PPHLN1 knockout cells, the levels of MPP8 and TASOR were increased (Fig. 3a) and immunoprecipitation of NP220 resulted in correspondingly increased co-immunoprecipitation of both MPP8 and TASOR (Fig. 3a).

ChIP assays were performed to monitor binding of these proteins to unintegrated viral DNA. NP220 was able to bind to both linear and circular unintegrated DNA (Fig. 3b and Extended Data Fig. 3e). NP220 knockout eliminated the ChIP signal (Fig. 3b) and re-expression of wild-type NP220 or the NP220(ΔZnF) mutant restored binding, whereas expression of the NP220(ΔDB) mutant did not (Extended Data Fig. 3c). Notably, NP220 bound to DNA even when expression of any of the HUSH subunits or SETDB1 was eliminated, suggesting that NP220 is the primary DNA-binding component (Fig. 3b). We also detected robust binding of each subunit of the HUSH complex (MPP8, TASOR and PPHLN1) to unintegrated viral DNA in wild-type cells, but not in MPP8 or SETDB1 knockout cells (Fig. 3c–e). Binding of the HUSH complex to unintegrated viral DNA required histone methylation, because re-expression of NP220(W80A), which lacks methyl-binding activity, in the MPP8 knockout line showed only weak binding to viral DNA (Extended Data Fig. 3d). None of the HUSH subunits or SETDB1 bound to viral DNA in NP220 knockout cells (Fig. 3c–f), suggesting that NP220 has a key role in bringing the HUSH complex and SETDB1 to DNA.

Knockout of NP220, MPP8 or SETDB1 significantly decreased H3K9me3 modification on unintegrated retroviral DNA (Fig. 3g), indicating that NP220, the HUSH complex and SETDB1 are all required for H3K9 trimethylation on unintegrated retroviral DNA. Knockdown of HDAC1 or HDAC4 also increased the expression of unintegrated retroviral DNA (Fig. 4a, b) and increased the levels of acetylated histone H3 (Fig. 4c). Co-immunoprecipitation assays showed that endogenous NP220 bound to HDAC4, but not other HDACs (Fig. 4d and Extended Data Fig. 4a). The NP220–HDAC4 interaction was independent of DNA or RNA (Extended Data Fig. 4b). NP220(W80A) did not co-immunoprecipitate with HDAC4 (Fig. 4e, f). The NP220–HDAC4 interaction was independent of DNA or RNA (Extended Data Fig. 4b). NP220(W80A) did not co-immunoprecipitate with HDAC4 (Fig. 4e, f). The C-terminal zinc finger motif is critical for this interaction. ChIP assays showed that both HDAC1 and HDAC4 were bound to unintegrated retroviral DNA and that binding was reduced in NP220 knockout cells (Fig. 4f, g). Re-expression of wild-type NP220 in NP220 knockout cells decreased the levels of acetylated H3 on unintegrated DNA, whereas NP220(ΔDB) and NP220(AZnF) failed to do so (Fig. 4h). These results indicate that NP220 utilizes its C-terminal zinc finger to recruit HDAC4, and probably HDAC1, to deacetylate histone H3 and thereby silence the expression of unintegrated retroviral DNA. Deacetylation appears to be mechanistically upstream of H3 methylation: HDAC inhibition and H3 acetylation led to decreased H3K9me3 (Extended Data Fig. 2c, d).
To test the importance of these DNA sequences in NP220-mediated silencing of unintegrated DNA, we generated a panel of variant reporters (Fig. 5a). Mutations M1 and M2 that delete the first three NP220-binding sites had no measurable effect on the LTR promoter activity when integration was allowed, but allowed higher expression of unintegrated DNAs (Extended Data Fig. 6b, c). Expression of unintegrated MLV DNA in which the U3 region was replaced by RSV U3—which lacks cytidine clusters—or in which the NP220-binding sequences made unintegrated HIV-1 DNA less responsive to silencing—decreased upon the replacement, deletion or mutation of the NP220-binding DNA sequences (Fig. 5d). In the case of HIV-1, we identified five consensus NP220-binding sequences in the U3 promoter region (Extended Data Fig. 7a, b). Deletion of these putative binding sequences made unintegrated HIV-1 DNA less responsive to NP220 knockout (Extended Data Fig. 7c, d). It should be noted that the last three binding sequences overlap with SP1-binding sites and deletion of these sequences severely diminished HIV-1 LTR basal promoter activity (Extended Data Fig. 7e).

NP220 not only silenced non-integrating viral DNA vectors, but also influenced infection by integration-competent vectors and even replicating viruses. MLV DNA was silenced, marked with histone deacetylation and bound by NP220, and NP220 knockout markedly increased the expression of viral DNA at 12 h after infection, when most of the viral DNA is unintegrated, but not 48 h after infection, when most of the viral
DNA is integrated into the host genome (Fig. 5e, f and Extended Data Fig. 6d–g). The rate at which MLV and HIV-1 spread in NP220 KO cells was faster than in control cells (Fig. 5g, h and Extended Data Fig. 7f).

Here we define the mechanisms and the machinery by which silencing is imposed on unintegrated retroviral DNA (Extended Data Fig. 8). These findings define new functions for NP220 and the HUSH complex. The restriction is sufficiently strong that many viruses have evolved means to evade or inactivate this machinery: the Vpr and Vpx proteins from primate immunodeficiency viruses mediate degradation of the HUSH complex to relieve silencing of proviruses, and the ICP0 protein of herpes simplex virus type 1 relieves HUSH-mediated silencing of viral DNA. The silencing mechanisms and machinery are illustrated in Fig. 8. These findings define new functions for NP220 and the HUSH complex. The restriction is sufficiently strong that many viruses have evolved means to evade or inactivate this machinery: the Vpr and Vpx proteins from primate immunodeficiency viruses mediate degradation of the HUSH complex to relieve silencing of proviruses and the ICP0 protein of herpes simplex virus type 1 relieves HUSH-mediated silencing of viral DNA. The silencing mechanisms and machinery are illustrated in Fig. 8.

Fig. 4 | NP220 recruits HDAC1 and HDAC4 to deacetylate histone H3 on unintegrated retroviral DNA. **Note**: a–c, HDAC1 and HDAC4 are required for the silencing of unintegrated retroviral DNA. HeLa cells were transfected with indicated siRNAs and then infected with an integrase-deficient (IN(D184A)) MLV-Luc virus. a, b, Luciferase activities were measured and luciferase activity of non-targeting (NT) siRNA-transfected cells was set to 1. The expression of HDAC1 and HDAC4 was determined by western blot (b, bottom). c, ChIP was performed using antibodies against pan-acetyl H3 followed by qPCR using primers that target the LTR. d, e, NP220 interacts with HDAC4. d, Endogenous NP220 was immunoprecipitated from lysates of the indicated HeLa cell lines. e, Haemagglutinin (HA)-tagged NP220 or NP220 in which the zinc finger domain was deleted (ΔZnF) were introduced into NP220 knockout HeLa cells. NP220 was then immunoprecipitated from HeLa cell lysates using an anti-haemagglutinin antibody. The co-immunoprecipitating protein HDAC4 was analysed by western blot. Images are representative of two independent experiments. NS, not significant (P > 0.05); *P < 0.05; **P < 0.01. P values are from paired two-sided Student’s t-tests. Exact P values are included in the Source Data associated with this figure.

Fig. 5 | The binding specificity of NP220 to unintegrated retroviral DNA. **Note**: a, Locations of putative NP220 binding sites in the MLV U3 region. Putative NP220-binding sites are indicated in red. T, C nucleotides mutated to A, B, NP220 binds a DNA fragment from the MLV U3 region. Indicated DNA fragments were incubated with the NP220 DNA-binding domain (NP220 DB) and shifted bands were probed with an antibody that recognized biotin. c, d, Specific DNA sequences are responsible for NP220 silencing and binding of unintegrated retroviral DNA. Indicated cells were infected with integrase-deficient (IN(D184A)) MLV-Luc viruses bearing the indicated deletions or mutations in the U3 region. c, Luciferase activities were measured and the fold increase (NP220 knockout/NT wild type) was calculated as the ratio of luciferase activity in knockout cells compared to wild-type cells. d, NP220 was performed to assess the association of NP220 with unintegrated MLV-Luc DNA. RSV U3, replacement of MLV U3 with RSV U3. Viral spreading was monitored using an assay for reverse transcriptase activity in knockout cells compared to wild-type cells. d, ChIP was performed to assess the association of NP220 with unintegrated MLV-Luc DNA. RSV U3, replacement of MLV U3 with RSV U3.
uncovered here may have wide-ranging implications for the design of non-integrating retroviral vectors for gene therapy and induction of gene expression after plasmid DNA transfection.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0750-6.

Received: 8 May 2018; Accepted: 22 October 2018; Published online 28 November 2018.

1. Orzalli, M. H. & Knipe, D. M. Cellular sensing of viral DNA and viral evasion mechanisms. *Annu. Rev. Microbiol.* **68**, 477–492 (2014).

2. Schwartzberg, P., Colicelli, J. & Goff, S. P. Construction and analysis of deletion mutants in the pol gene of Moloney murine leukemia virus: a new viral function required for productive infection. *Cell* **37**, 1043–1052 (1984).

3. Schneider, W. M., Wu, D. T., Amin, V., Aiyer, S. & Roth, M. J. MuLV IN mutants responsive to HDAC inhibitors enhance transcription from unintegrated retroviral DNA. *Virology* **426**, 188–196 (2012).

4. Šakai, H. et al. Integration is essential for efficient gene expression of human immunodeficiency virus type 1. *J. Virol.* **67**, 1169–1174 (1993).

5. Nakajima, N., Lu, R. & Engelman, A. Human immunodeficiency virus type 1 replication in the absence of integrase-mediated DNA recombination: definition of permissive and nonpermissive T-cell lines. *J. Virol.* **75**, 7944–7955 (2001).

6. Tchasovnikarova, I. A. et al. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science* **348**, 1481–1485 (2015).

7. Liu, N. et al. Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature* **553**, 228–232 (2018).

8. Robbez-Masson, L. et al. The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. *Genome Res.* **28**, 836–845 (2018).

9. Varmus, H. Retroviruses. *Science* **240**, 1427–1435 (1988).

10. Yoshimura, F. K. & Weinberg, R. A. Restriction endonuclease cleavage of linear and closed circular murine leukemia viral DNAs: discovery of a smaller circular form. *Cell* **16**, 323–332 (1979).

11. Kilzer, J. M. et al. Roles of host cell factors in circularization of retroviral DNA. *Virology* **314**, 460–467 (2003).

12. Inagaki, H. et al. A large DNA-binding nuclear protein with RNA recognition motif and serine/arginine-rich domain. *J. Biol. Chem.* **271**, 12525–12531 (1996).

13. Chang, Y., Horton, J. R., Bedford, M. T., Zhang, X. & Cheng, X. Structural insights for MPP8 chromodomain interaction with histone H3 lysine 9: potential effect of phosphorylation on methyl-lysine binding. *J. Mol. Biol.* **408**, 807–814 (2011).

14. Kokura, K., Sun, L., Bedford, M. T. & Fang, J. Methyl-H3K9-binding protein MPP8 mediates E-cadherin gene silencing and promotes tumour cell motility and invasion. *EMBO J.* **29**, 3673–3687 (2010).

15. Chouguì, G. et al. HIV-2/SIV viral protein X counteracts HUSH repressor complex. *Nat. Microbiol.* **3**, 891–897 (2018).

16. Yurkovetskiy, L. et al. Primate immunodeficiency virus proteins Vpx and Vpr counteract transcriptional repression of proviruses by the HUSH complex. *Nat. Microbiol.* https://doi.org/10.1038/s41564-018-0256-x (2018).

17. Gu, H. & Roizman, B. The two functions of herpes simplex virus 1 ICP0, inhibition of silencing by the CoREST/REST/HDAC complex and degradation of PML, are executed in tandem. *J. Virol.* **83**, 181–187 (2009).

18. Poon, A. P., Gu, H. & Roizman, B. ICP0 and the US3 protein kinase of herpes simplex virus 1 independently block histone deacetylation to enable gene expression. *Proc. Natl Acad. Sci. USA* **103**, 9993–9998 (2006).

Acknowledgements Cell sorting was performed with the assistance of A. Figueroa in the Core Facility of Department of Microbiology and Immunology at Columbia University. This work was supported by the Howard Hughes Medical Institute and by National Institutes of Health (NIH) grant R01 CA30488 (to S.P.G.).

Reviewer information Nature thanks A. Engelman, M. Lorincz, L. Naldini and M. Ott for their contribution to the peer review of this work.

Author contributions Y.Z., G.Z.W. and S.P.G. designed the experiments. Y.Z., G.Z.W. and S.P.G. wrote the manuscript, with input from O.C. Y.Z. and S.P.G. performed experiments in Fig. 2 and ChIP–qPCR experiments. All other experiments were performed by Y.Z.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0750-6.

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0750-6.

Reprints and permissions information is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to S.P.G.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
METHODS

Cell Lines. Mammalian cell lines, HeLa (CCL-2), NIH3T3 (CRL-1658), HEK293T (CRL-11268), COS-7 (CRL-1651) and the chicken cell line DF-1 (CRL-12203) were purchased from ATCC. All lines and their derivatives were maintained at 37 °C and 5% CO₂ in DMEM supplemented with 10% heat-inactivated fetal bovine serum, 2 mM glutamine, penicillin and streptomycin. MT-4 (C120) cells were obtained through the NIH AIDS Reagent Program and cultured at 37 °C and 5% CO₂ in RPMI 1640 supplemented with 10% heat-inactivated fetal bovine serum.

DNA construction. The replication-defective retroviral vectors pNCA-Luc (MLV vector that expresses firefly luciferase), pNCA-GFP (MLV vector that expresses GFP) and pSARM-Luc (MMPV vector that expresses firefly luciferase) have been described previously 29. pRCAS-Luc (RSV vector that expresses luciferase) was constructed by replacing GFP in pRCAS-GFP with firefly luciferase. pNL4.3-Luc (HIV-1 vector that expresses firefly luciferase) was obtained from the NIH AIDS Reagent Program. The plasmid pCMV-Neo-expresses wild-type Gag and Pol from the plasmid pHCMV-Gag-Pol 19. pMD2.G—was cloned into pLVX-EF1-IRES-Neo (CLONTECH, 632181) with the EF1 promoter. The coding sequence of MPR8 (wild type or with a W08A mutation) was cloned into pLVX-EF1-1RES-Neo to produce pLVX-EF1-1RES-Neo-MPR8 and pLVX-EF1-1RES-Neo-MPR8(W08A). The coding sequences of NP220 with (including the putative Np220-binding sites) were constructed as follows: the MLV(U3) region in the wild type (MT-4 cell line) was inserted into pLVX-EF1-1RES-Neo to produce pLVX-EF1-RES-Neo-MP8 and pLVX-EF1-1RES-Neo-MP8Δ(U3)(W08A).

Quantitative PCR. Quantitative PCR (qPCR) was performed in an ABI 7500 Fast Real-Time PCR System using FastStart Universal SYBR Green Master (Roche). The PCR cycle program was: (1) 50 °C, 2 min, 1 cycle; (2) 95 °C, 10 min, 1 cycle; (3) from 95 °C, 15 s to 60 °C, 30 s, 30 to 72 °C, 30 s, 40 cycles; (4) 72 °C, 10 min, 1 cycle. The primers used for qPCR are listed in Supplementary Table 2.

ChIP. In brief, 2 × 10⁸ cells were seeded in 10-cm dishes and infected with VSV-G pseudotyped, integrate-deficient or integrate-proficient MLV-Luc viruses for two days. The virus used for infection was pretreated with 5 U ml⁻¹ DNase (Promega, M6101) supplemented with 10 mM MgCl₂ at 37 °C for 1 h to remove any residual plasmid DNA. Cells were crosslinked in 1% formaldehyde for 10 min, quenched in 0.125 M glycine for 5 min and lysed in 1 ml of ChIP lysis buffer (50 mM Tris-Cl, pH 8.0, 1% SDS, 10 mM EDTA) at 4°C for 10 min. The supernatant of approximately 50 µg sonicated chromatin was then immunoprecipitated overnight using 5 µg of the specific antibodies in ChIP dilution buffer (10 mM Tris-Cl, pH 8.0, 1% Triton X-100, 0.1% SDS, 150 mM NaCl, 2 mM EDTA). The next day, 25 µl of Dynabeads (12.5 µl protein A and 12.5 µl protein G) was added and the mixture was incubated for an additional 2 h. The beads were washed twice each in ChIP low-salt buffer (20 mM Tris-Cl, pH 8.0, 1% Triton X-100, 0.1% SDS, 150 mM NaCl, 2 mM EDTA), ChIP high-salt buffer (20 mM Tris-Cl, pH 8.0, 1% Triton X-100, 0.1% SDS, 500 mM NaCl, 2 mM EDTA), ChIP LiCl buffer (10 mM Tris-Cl, pH 8.0, 1% NP-40, 250 mM LiCl, 1 mM EDTA) and TE buffer (10 mM Tris-Cl, pH 8.0, 1 mM EDTA). Protein–DNA complexes were eluted from beads in 200 µl elution buffer (TE buffer containing 1% SDS, 100 mM NaCl, 5 mM DTT), treated with RNase A (5 µg per ml) for 30 min at 37 °C, and protein was removed by proteinase K digestion (5 µg ml⁻¹) and phenol–chloroform extraction (30 s, 2°C). DNA was precipitated with ethanol.

Luciferase assay. Luciferase activity was assayed 40 h after infection, using the Luciferase Assay System (Promega). The ecotropic MLV infectious molecular clone (pNCS) has been deposited to Addgene (plasmid 17362). The amphotropic MLV infectious molecular clone (pNCA-Luc(IN(D121A)), pSARM-Luc(IN(D121A)), pCMV-intron-IN(D184A), pNL4.3-Luc(IN(D64A)) were created by site-directed mutagenesis.

Luciferase assay. Luciferase activity was assayed 40 h after infection, using the Luciferase Assay System (Promega). The ecotropic MLV infectious molecular clone (pNCS) has been deposited to Addgene (plasmid 17362). The amphotropic MLV infectious molecular clone (pNCA-Luc(IN(D121A)), pSARM-Luc(IN(D121A)), pCMV-intron-IN(D184A), pNL4.3-Luc(IN(D64A)) were created by site-directed mutagenesis.

Luciferase assay. Luciferase activity was assayed 40 h after infection, using the Luciferase Assay System (Promega). The ecotropic MLV infectious molecular clone (pNCS) has been deposited to Addgene (plasmid 17362). The amphotropic MLV infectious molecular clone (pNCA-Luc(IN(D121A)), pSARM-Luc(IN(D121A)), pCMV-intron-IN(D184A), pNL4.3-Luc(IN(D64A)) were created by site-directed mutagenesis.

Luciferase assay. Luciferase activity was assayed 40 h after infection, using the Luciferase Assay System (Promega). The ecotropic MLV infectious molecular clone (pNCS) has been deposited to Addgene (plasmid 17362). The amphotropic MLV infectious molecular clone (pNCA-Luc(IN(D121A)), pSARM-Luc(IN(D121A)), pCMV-intron-IN(D184A), pNL4.3-Luc(IN(D64A)) were created by site-directed mutagenesis.

Luciferase assay. Luciferase activity was assayed 40 h after infection, using the Luciferase Assay System (Promega). The ecotropic MLV infectious molecular clone (pNCS) has been deposited to Addgene (plasmid 17362). The amphotropic MLV infectious molecular clone (pNCA-Luc(IN(D121A)), pSARM-Luc(IN(D121A)), pCMV-intron-IN(D184A), pNL4.3-Luc(IN(D64A)) were created by site-directed mutagenesis.
ZNT1 (L-007522-01), DUOX2 (L-032210-02), TESC (L-020896-01), GP5 (L-012671-00), SETDB1 (L-020070-00), SETD2B (L-014751-00), SUV39H1 (L-009604-00), SUV39H2 (L-008512-00), EHTM1 (L-007065-02), EHTM2 (L-006937-00), EZH2 (L-004218-00) and non-targeting control siRNA (NT, D-001810-10).

For siRNA transfection, 10^6 HeLa cells were seeded in six-well plates. After 24 h, siRNAs were transfected into the cells using Lipofectamine RNAiMax (Invitrogen) according to the manufacturer’s protocol. After another 24 h, the same siRNA transfection was performed for the second time. Then, 6 h after second transfection, the siRNA transfected cells were infected with virus for further experiments.

**Co-immunoprecipitation and immunoblotting.** For co-immunoprecipitation, approximately 5 × 10^6 HeLa cells were lysed in 1 ml Pierce IP Lysis Buffer for 10 min and the lysate was clarified by centrifugation at 4 °C for 15 min at 12,000 r.p.m. For nuclelease treatment, cell lysates were treated with benzonase (250 U ml^−1, supplemented with 10 mM MgCl2). DNaSe (5 U ml^−1 supplemented with 10 mM MgCl2) or RNase A (5 μg ml^−1). First, 40 μl Dynabeads (20 μl protein A and 20 μl protein G beads) were mixed with 1 μg of the specific antibody for 10 min at room temperature and then washed twice with TBST. Antibody-coated Dynabeads were incubated with 800 μl cell lysate at 4 °C for 4 h. The beads were then washed three times with TBST. The bound proteins on the beads were eluted with 40 μl 1× SDS sample buffer and subjected to SDS–PAGE and western blot analysis. For immunoblotting, cells were lysed in RIPA Lysis and Extraction Buffer for 10 min. The lysate was clarified by centrifugation at 4 °C for 15 min at 12,000 r.p.m. The samples were heated at 95 °C in SDS sample buffer and resolved by SDS–PAGE electrophoresis, transferred to a PVDF membrane and probed with specific antibodies by western blot.

**Genome-wide CRISPR–Cas9 screen.** The human CRISPR knockout pooled libraries (Brunello, two-vector system) was obtained from Addgene and the screen was performed broadly as described previously.21 The CRISPR sgRNA library virus was packaged by transfecting HEK293T cells with library DNA, a HIV-1-Gag–Pol–amphotropic MLV at a low MOI. Culture medium was changed at 3 h after infection. Culture supernatants (50 μl) were taken every two days for 12 days after infection. The culture medium was assayed for reverse transcriptase activity to monitor the production of progeny virus.

For amphotropic MLV replication in HeLa cells, NP220 wild-type or knock-out cells (3 × 10^6 cells per well) were seeded in six-well plates and infected with amphotropic MLV at a low MOI. Culture medium was changed at 3 h after infection. Culture supernatants (50 μl) were taken every two days for 12 days after infection. The culture medium was assayed for reverse transcriptase activity to monitor the production of progeny virus. For the reverse transcriptase assay, 5 μl of culture supernatant was incubated with 20 μl hot/cold mix at room temperature for 40 min and then 5 μl of the reaction mix was spotted on DEAE paper. The DEAE paper was washed with 2× SSC buffer for 20 min twice, dried under a heat lamp, exposed and visualized by phosphor imaging (GE). The formula for the buffers are as follows: hot/cold mix (1 ml cold mix, 72 μl hot mix, 2 μl MnCl2, (0.5 M), 1 μl 5′-dTTTP, all reagents in the indicated order); cold mix (50 μM Tris–HCl pH 8.3, 75 mM NaCl, 0.06% NP-40); hot mix (7.6 mg ml^−1 oliga(DT), 16.6 μg ml^−1 TTP, 166 μg ml^−1 poly(A), 0.5 M DTT); oliga(DT) (GE healthcare, 27-7858-2); poly(A) (GE healthcare, 27-4110).

**HIV-1 replication.** HIV-1 viruses were produced by transfecting HEK293T cells with NL4.3 using Lipofectamine 2000 (Invitrogen) following the manufacturer’s protocol. For HIV-1 replication in MT-4 cells, 10^6 NP220 wild-type or knockout cells were transduced with HIV-1 virus (1 ng capsid protein p24) by spin infection (3,500 r.p.m. at room temperature for 2 h) and then cultured at 37 °C. After 3 h of infection, cells were washed twice with PBS and then cultured in 1 ml medium in a 24-well plate every two days. After 10 days, aliquots of the culture supernatants (50 μl) were taken for p24 measurement and half of the cells and medium (500 μl) were transferred to a new well containing 500 μl fresh medium, p24 levels were determined by enzyme-linked immunosorbent assay (ELISA) using the HIV1 p24 ELISA kit (Abcam, ab218268), to monitor the production of progeny virus.

**Statistical analysis.** Sample sizes are provided in the figure legends. Statistical significance was determined by a two-tailed Student’s t-test. The experiments were not randomized. The investigators were not blinded to allocation during experiments or outcome assessment.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

**Data availability.** The data that support the findings of this study are available from the corresponding author upon request. Source Data is available for the Figs 3a–g, 4c–h, 5b, d and Extended Data Figs 1e–j, 3a–e, 4a–d, 6f, g. For gel source data, see Supplementary Fig. 1.

19. Wang, G. Z. & Goff, S. P. Postentry restriction of Mason–Pfizer monkey virus in mouse cells. J. Virol. 89, 2813–2819 (2015).
20. Wolf, D. & Goff, S. P. Embryonic stem cells use ZFP809 to silence retroviral DNAs. Nature 458, 1201–1204 (2009).
21. Junqu, J. et al. Genome-scale CRISPR–Cas9 knockout and transcriptional activation screening. Nat. Protoc. 12, 828–833 (2017).
22. Diaz, A. A., Qin, H., Ramalho-Santos, M. & Song, S. J. HitSelect: a comprehensive tool for high-complexity-pooled screen analysis. Nucleic Acids Res. 43, e16 (2015).
Extended Data Fig. 1 | Silencing of unintegrated, but not integrated, retroviral DNA. a, b, Silencing of unintegrated retroviral DNA. HeLa cells were infected with VSV-G pseudotyped, integrase-proficient (IN(WT)) or integrase-deficient (IN(D184A)) MLV-Luc viruses. Total viral DNA levels (a) and luciferase activities (b) were measured 40 h after infection. The viral DNA concentration and luciferase activity of IN(WT) MLV-Luc were set to 1. Data are mean ± s.d.; n = 3 independent experiments. c, d, Silencing of unintegrated retroviral DNA is dependent on histone deacetylation. HeLa (c) and NIH3T3 (d) cells were infected with VSV-G pseudotyped, integrase-proficient (IN(WT)) or integrase-deficient (IN(D184A)) MLV-Luc viruses and treated with DMSO (TSA−) or 1 μM HDAC inhibitor trichostatin A (TSA+). Luciferase activities were measured 40 h after infection. Data are mean ± s.d.; n = 3 independent experiments. e–j, Repressive epigenetic marks are present on unintegrated retroviral DNA, whereas active epigenetic marks are present on integrated retroviral DNA. HeLa cells were infected with VSV-G pseudotyped, integrase-deficient (IN(D184A)) (e–g) or integrase-proficient (IN(WT)) (h–j) MLV-Luc viruses. At 40 h after infection, ChIP—using antibodies against pan-acetyl H3, H3K9me3 or H3K27me3—was performed to assess H3ac (e, h), H3K9me3 (f, i) and H3K27me3 (g, j) modifications across the LTR of unintegrated and integrated MLV-Luc DNA. qPCR data from each ChIP were calculated as the percentage of input DNA. Data are mean ± s.d.; n = 3 independent experiments. ns, P > 0.05; *P < 0.05; **P < 0.01. P values are from paired two-sided Student’s t-tests. Exact P values are included in the Source Data associated with this figure. k, Knockdown of indicated genes has no effect on viral DNA levels. HeLa cells were first transfected with the indicated siRNA and then infected with VSV-G pseudotyped, integrase-deficient (IN(D184A)) MLV-Luc virus. Viral DNA levels were measured 40 h after infection. DNA levels in HeLa cells transfected with a non-targeting (NT) control siRNA was set to 1. Data are mean ± s.d.; n = 3 independent experiments.
Extended Data Fig. 2 | NP220 and SETDB1 are required for the silencing of unintegrated MLV DNA. a, b, HeLa (a) or NIH3T3 (b) cells were transfected with indicated siRNAs and then infected with VSV-G pseudotyped, integrase-deficient (IN(D184A)) MLV-Luc virus. Luciferase activities were measured 40 h after infection and luciferase activity in non-targeting (NT) control siRNA-transfected cells was set to 1 (top). Data are mean ± s.d.; n = 3 independent experiments. The expression of NP220 was determined by western blot (bottom). c, HeLa cells were first transfected with the indicated siRNAs targeting histone methyltransferases and then infected with VSV-G pseudotyped, integrase-deficient (IN(D184A)) MLV-Luc virus. Luciferase activities were measured 40 h after infection and luciferase activity in non-targeting (NT) control siRNA-transfected HeLa cells was set to 1. Data presented are mean ± s.d.; n = 3 independent experiments. d, Knockdown efficiency of siRNAs used in c. HeLa cells were transfected with the indicated siRNAs targeting histone methyltransferases and mRNA levels of siRNA-targeted genes were measured by qPCR with reverse transcription (RT-qPCR). mRNA levels in non-targeting (NT) control siRNA-transfected HeLa cells was set to 1. Data are mean ± s.d.; n = 3 independent experiments.
Extended Data Fig. 3 | NP220 recruits the HUSH complex and SETDB1 to silence unintegrated MLV DNA. a, b, HA-tagged NP220, NP220 with zinc finger deletion (ΔZnF) (a) or indicated fragments of NP220 (b) were introduced into NP220 knockout HeLa cells and then immunoprecipitated using an HA antibody. Co-immunoprecipitated MPP8 was analysed by western blot. Images are representative of two independent experiments with similar results. c, Parental wild-type (WT) HeLa cells, NP220 knockout HeLa cells, NP220 knockoout HeLa cells and NP220 knockout HeLa cells that were reconstituted with indicated variants of NP220 were subsequently infected with VSV-G pseudotyped, integrase-deficient (IN(D184A)) MLV-Luc virus. At 40 h after infection, ChIP was performed to assess the association of NP220 across the LTR of unintegrated MLV-Luc DNA. qPCR data from each ChIP were calculated as the percentage of input DNA. Data are mean ± s.d.; n = 3 independent experiments. ns, P > 0.05; *P < 0.05. P values are from paired two-sided Student’s t-tests. Exact P values are included in the Source Data associated with this figure. d, Parental HeLa cells, MPP8 knockout HeLa cells and MPP8 knockout HeLa cells that were reconstituted with indicated variants of MPP8 were infected with VSV-G pseudotyped, integrase-deficient (IN(D184A)) MLV-Luc virus. At 40 h after infection, ChIP was performed to assess the association of MPP8 with the LTR of unintegrated MLV-Luc DNA. qPCR data from each ChIP were calculated as the percentage of input DNA. Data are mean ± s.d.; n = 3 independent experiments. ns, P > 0.05; *P < 0.05. P values are from paired two-sided Student’s t-tests. Exact P values are included in the Source Data associated with this figure. e, HeLa cells were infected with VSV-G pseudotyped, integrase-deficient (IN(D184A)) MLV-Luc virus. At 40 h after infection, ChIP was performed using the indicated antibodies followed by qPCR using primers targeting 2LTR circles. qPCR data from each ChIP were calculated as the percentage of input DNA. Data are mean ± s.d.; n = 3 independent experiments.
Extended Data Fig. 4 | Interaction between NP220 and HDACs.

**a**, Screen for interactions between NP220 and HDACs. Endogenous NP220 were immunoprecipitated (IP) from indicated HeLa cell lines and the indicated co-immunoprecipitated HDAC proteins and MPP8 were analysed by western blot using specific antibodies. MPP8 serves as a positive control. Images are representative of two independent experiments with similar results.

**b**, NP220–MPP8 or NP220–HDAC4 interactions are independent of DNA or RNA. Cell lysates from indicated HeLa cell lines were treated with benzonase, DNase or RNase A, endogenous NP220 were immunoprecipitated and co-immunoprecipitating proteins were analysed by western blot using specific antibodies. Images are representative of two independent experiments with similar results.

**c, d**, Interrelationship of histone deacetylation and H3K9 trimethylation on unintegrated viral DNA. 

- **c**, HeLa cells were infected with VSV-G pseudotyped, integrase-deficient (IN(D184A)) MLV-Luc virus and treated with DMSO (TSA−) or 1 μM HDAC inhibitor trichostatin A (TSA+).
- **d**, Parental HeLa cells and SETDB1 knockout HeLa cells (SETDB1 KO) were infected with VSV-G pseudotyped, integrase-deficient (IN(D184A)) MLV-Luc virus. At 40 h after infection, ChIP was performed using the indicated antibodies followed by qPCR using primers targeting LTR. qPCR data from each ChIP were calculated as the percentage of input DNA. Data are mean ± s.d.; n = 3 independent experiments. **P** < 0.05; ***P** < 0.01. 

P values are from paired two-sided Student’s t-tests. Exact P values are included in the Source Data associated with this figure.
Extended Data Fig. 5 | NP220 mediates silencing of unintegrated retroviral DNA from HIV-1 and MPMV, but not RSV. a, Parental MT-4 cells and NP220 knockout MT-4 cell line (NP220 KO) were infected with VSV-G pseudotyped, integrase-deficient (IN(D64A)) HIV-1 vector NL4.3-Luc. Luciferase activities were measured 40 h after infection and luciferase activity in parental (wild-type) MT-4 cells was set to 1 (top). Data are as mean ± s.d.; n = 3 independent experiments. The expression of NP220 was determined by western blot (bottom). b, COS-7 cells were first transfected with indicated siRNAs and then infected with VSV-G pseudotyped, integrase-deficient (IN(D127A)) MPMV vector SARM-Luc. Luciferase activities were measured 40 h after infection and luciferase activity in non-targeting (NT) control siRNA-transfected cells was set to 1 (top). Data are mean ± s.d.; n = 3 independent experiments. The expression of NP220 was determined by western blot (bottom). c–e, Indicated HeLa cell lines were infected with VSV-G pseudotyped, integrase-deficient HIV-1 vector NL4.3-Luc (c), MPMV vector SARM-Luc (d) or RSV vector RCAS-Luc (e). Luciferase activities were measured 40 h after infection. Luciferase activity in parental HeLa cells was set to 1. Data are mean ± s.d.; n = 3 independent experiments. 

f, g, HDAC1 and HDAC4 are involved in silencing of unintegrated HIV-1 and MPMV DNA. HeLa cells were first transfected with the indicated siRNAs and then infected with VSV-G pseudotyped, integrase-deficient (IN(D64A)) HIV-1 vector NL4.3-Luc (f) or VSV-G pseudotyped, integrase-deficient (IN(D127A)) MPMV vector SARM-Luc (g). Luciferase activities were measured 40 h after infection and luciferase activities in HeLa cells transfected with control non-targeting (NT) siRNA were set to 1. Data are mean ± s.d.; n = 3 independent experiments.
Extended Data Fig. 6 | The binding specificity of NP220 bound to unintegrated MLV DNA. **a**, The sequence of MLV U3 region. Putative NP220-binding sites are indicated in red. The sequence of the 84-nucleotide probe used for EMSA is italicized and underlined. **b, c**, Parental or NP220 knockout HeLa cells were infected with VSV-G pseudotyped, integrase-proficient (IN(WT)) (**b**) or integrase-deficient (IN(D184A)) (**c**) MLV-Luc viruses bearing the indicated deletions or mutations in the U3 region. Luciferase activities were measured 40 h after infection. Data are mean ± s.d.; n = 3 independent experiments. **d, e**, Indicated HeLa cell lines were infected with VSV-G pseudotyped, integrase-proficient (IN(WT)) MLV-Luc virus. At the indicated times after infection, luciferase activities were measured. **e**, Fold increase (KO/WT) was calculated as the ratio of the luciferase activities of knockout cells compared to the luciferase activities of wild-type cells. Data are mean ± s.d.; n = 3 independent experiments. 

© 2018 Springer Nature Limited. All rights reserved.
Extended Data Fig. 7 | The binding specificity of NP220 bound to unintegrated HIV-1 DNA. **a, b.** The sequences and locations of putative NP220-binding sites in HIV-1 U3 region. Putative NP220-binding sites are indicated in red. **c, d.** Parental or NP220 knockout HeLa cells were infected with VSV-G pseudotyped, integrase-deficient (IN(D64A)) MLV-Luc viruses bearing indicated deletions in the U3 region. Luciferase activities were measured 40 h after infection. Fold increase (NP220 KO/NP220 WT) was calculated as the ratio of the luciferase activities of knockout cells compared to the luciferase activities of wild-type cells. Data are mean ± s.d.; n = 3 independent experiments. **e, f.** Knockout HeLa cells were infected with VSV-G pseudotyped, integrase-proficient (IN(WT)) HIV-Luc viruses bearing the indicated deletions in the U3 region. Luciferase activities were measured 40 h after infection. Data are mean ± s.d.; n = 3 independent experiments. **f.** Knockout of NP220 increases the rate of HIV-1 spreading. **g.** Parental and NP220 knockout MT-4 cells were infected with HIV-1(NL4.3). Viral spreading was monitored by assay for p24 concentration in the culture medium. Data are mean from two technical ELISA replicates and are representative of two independent experiments.
Extended Data Fig. 8 | Schematic of the silencing of unintegrated retroviral DNAs. Retroviral infection results in the synthesis of a linear double-stranded DNA in the cytoplasm, which is delivered into the nucleus to give rise to two circular forms and the integrated provirus (top). The unintegrated nuclear DNAs are rapidly loaded with nucleosomal histones (blue). In the case of MLV, NP220 binds to the unintegrated viral DNA and is responsible for attracting histone deacetylases (HDACs), the HUSH complex (consisting of MPP8, TASOR and PPHLN1) and the histone methyltransferase SETDB1. HDACs remove the activation marks of histone acetylation and SETDB1 introduces repressive H3K9me3 marks. MPP8 binds H3K9me3 to strengthen the association with the viral chromatin.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- n/a  Confirmed
- X  The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- X  An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- X  The statistical test(s) used AND whether they are one- or two-sided
- X  Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- X  A description of all covariates tested
- X  A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- X  A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- X  For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
-  Give P values as exact values whenever suitable.
- X  For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- X  For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- X  Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated
- X  Clearly defined error bars
- X  State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection
Deep-sequencing of the sgRNA sequences in the bulk genomic DNA was performed on the Illumina HiSeq (2x150bp configuration, per lane). The count of sgRNA was analyzed with count_spacer.py using Python 2.7. (The Python script of count_spacer.py was downloaded from Supplementary information session in "Joung J. et al, Nat Protoc. 2017 Apr;12(4):828-863. PMID: 28333914").

Data analysis
The degree of sgRNA enrichment and gene hit rank was analyzed by software HiTSelect (HiTSelect_Windows.exe, modified on 2014-07-08, available at http://sourceforge.net/projects/hitselect/, see PMID: 25428347 for detailed description of the HiTSelect). Statistical analysis was performed using GraphPad Prism (Version 7.04).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. The following figures have associated raw data: Figures 3a-g, 4c-h, 5b, 5d; Extended Data Figures 1e-j, 3a-e, 4a-d, 6f-g.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☑ Life sciences  ☐ Behavioural & social sciences  ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Sample size was based on traditional experimental approach in molecular and cell biology. In general, the sample sizes for quantitative luciferase assays, qPCR, ChIP-qPCR were 3 (repeated independently for 3 times).

Data exclusions

No data were excluded from analysis.

Replication

All attempts at replication were successful. In Figure legends, number of replications was stated for each experiment. Quantitative luciferase assays, qPCR, ChIP-qPCR experiments were repeated independently for 3 times; Western Blots and Virus replication experiments were repeated twice. The CRISPR screening was performed only once, without replication, but the top screening hits were further validated by CRISPR knockout or siRNA knockdown experiment.

Randomization

The experiments did not require sample randomization. Samples were handled the same way in all experiments.

Blinding

The investigators were not blinded to group allocation during data collection or subsequent analysis. This approach is considered standard for biochemical experiments performed in this study.

Reporting for specific materials, systems and methods

Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ☑   | Unique biological materials |
| ☐   | Antibodies |
| ☑   | Eukaryotic cell lines |
| ☑   | Palaeontology |
| ☐   | Animals and other organisms |
| ☑   | Human research participants |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ☑   | ChIP-seq |
| ☑   | Flow cytometry |
| ☑   | MRI-based neuroimaging |

Antibodies

Antibodies used

- Histone H3ac (pan-acetyl), abcam, ab47915 (ChIP: 5 ug / 50 ug DNA)
- Histone H3K9me3, abcam, ab8898  (ChIP: 5 ug / 50 ug DNA)
- Histone H3K27me3 (Lys27), EMD Millipore, 07-449 (ChIP: 5 ug / 50 ug DNA)
- Histone H3, abcam, ab1791 (ChIP: 5 ug / 50 ug DNA)
- Rabbit IgG, abcam, ab171870 (ChIP: 5 ug / 50 ug DNA)
- SETDB1, Proteintech, 11231-1-AP (WB: 1:1000, ChIP: 5 ug / 50 ug DNA)
- HDAC1, Proteintech, 10197-1-AP (WB: 1:1000, ChIP: 5 ug / 50 ug DNA)
- HDAC4, Proteintech, 17449-1-AP (WB: 1:1000, ChIP: 5 ug / 50 ug DNA)
Validation

The antibodies suitable for specific purposes (validation was performed by the vendors) were purchased for our experiments. Key antibodies, including NP220, MPP8, TASOR, PPHLN1, SETDB1, HDAC1, HDAC4 antibodies, were validated by western blot upon depletion of the proteins.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s) Cell lines HeLa (CCL-2), NIH3T3 (CRL-1658), Hek293T (CRL-11268), COS-7 (CRL-1651), and chicken cell line DF-1 (CRL-12203) were purchased from ATCC. MT-4 (#120) cells were obtained through the NIH AIDS Reagent Program.

Authentication None of the cell lines used were authenticated.

Mycoplasma contamination Cell lines were not tested for Mycoplasma contamination.

Commonly misidentified lines (See ICLAC register) No commonly misidentified cell lines were used.

Flow Cytometry

Plots

Confirm that:
- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation HeLa cells infected with MLV-GFP IN-virus

Instrument FACSARiaD (P69500149)

Software FACSDiva Version 8.0; FlowJo

Cell population abundance 100% of cells were analyzed.

Gating strategy 5% brightest GFP expressing cells were sorted out and collected. No gating strategy were used for analysis.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.