Nuclear architecture dictates HIV-1 integration site selection

B. Marini, A. Kertesz-Farkas, H. Ali, B. Lucic, K. Lisek, L. Manganaro, S. Pongor, R. Luzzati, A. Recchia, F. Mavilio, et al.

Nuclear architecture dictates HIV-1 integration site selection

Bruna Marini[1], Attila Kertesz-Farkas[2]*, Hashim Ali[1]*, Bojana Lucic[1]†, Kamil Lisek[1]†, Lara Manganaro[1]†, Sandor Pongor[2]†, Roberto Luzzati[3,4], Alessandra Recchia[5], Fulvio Mavilio[5,6], Mauro Giacca[1,4]§ & Marina Lusic[1]§†

Long-standing evidence indicates that human immunodeficiency virus type 1 (HIV-1) preferentially integrates into a subset of transcriptionally active genes of the host cell genome[1–4]. However, the reason why the virus selects only certain genes among all transcriptionally active regions in a target cell remains largely unknown. Here we show that HIV-1 integration occurs in the outer shell of the nucleus in close correspondence with the nuclear pore. This region contains a series of cellular genes, which are preferentially targeted by the virus, and characterized by the presence of active transcription chromatin marks before viral infection. In contrast, the virus strongly disfavours the heterochromatic regions in the nuclear lamin-associated domains[5] and other transcriptionally active regions located centrally in the nucleus. Functional viral integrase and the presence of the cellular Nup153 and LEDGF/p75 integration cofactors are indispensable for the peripheral integration of the virus. Once integrated at the nuclear pore, the HIV-1 DNA makes contact with various nucleoporins; this association takes part in the transcriptional regulation of the viral genome. These results indicate that nuclear topography is an essential determinant of the HIV-1 life cycle.

One important aspect of the interaction between HIV-1 and its target cells is the encounter between the viral complementary DNA (cDNA) with the complex architecture of the mammalian nucleus, in which chromosomes and genes are spatially arranged to occupy preferred positions within the three-dimensional space[6].

We analysed the lists of human genes targeted by HIV-1 from six different studies (Extended Data Table 1), containing altogether 1,136 unique gene integration sites in activated T cells carrying the CD4 antigen ($CD4^+$); 126 of these genes recurred in two lists, 24 in three, and six in at least four lists, for a total of 156 genes, which we named HIV recurrent integration genes (RIGs). The probability of detecting this number of specific genes by chance was extremely low ($P < 1 \times 10^{-9}$; Extended Data Fig. 1a). RIGs were also highly represented in another list of approximately 12,000 integration sites[4], 5,221 of which were unique genes, as well as in two integration lists generated from patients' $CD4^+$ T cells[7,8] ($P < 0.001$ of detecting these genes by chance). Thus, RIGs are bona fine the hottest spots of HIV-1 integration.

We then ranked RIGs according to their frequency and plotted them onto the human chromosome map[9]. Unexpectedly, they appeared to cluster into specific chromosomal regions (Extended Data Fig. 1b). In five out of eight cases, RIGs were also in proximity to the 'hotter zones', previously defined as regions with remarkably high HIV-1 integration density[1] (Supplementary Table 1). In these areas, observations hinted at the possibility that the topological distribution of these chromosomal regions inside the nucleus could determine HIV-1 integration.

By applying three-dimensional immuno-DNA fluorescence *in situ* hybridization (FISH), we assessed the position of RIGs and hotter zones in primary $CD4^+$ T cells from healthy donors. Selected FISH probes, listed in the Supplementary Information, provided topological information for a total of 169 RIGs and other integration sites located within 10 megabases (Mb) from the centre of the probe (Extended Data Fig. 2).

When the radial positions of the RIG FISH signals were binned into three zones of equal area[10,11] (Fig. 1a), a clear gradient in signal localization was observed, which decreased from the nuclear envelope towards the interior (images of 14 RIGs in Fig. 1b, c; four hotter zones in Fig. 1d). The global distribution of RIGs ($n = 1{,}420$ analysed alleles) was remarkably different from that of control genes, all of which were expressed in $CD4^+$ T cells[12,13] ($n = 522$): 44% of RIGs mapped in zone 1, 41.5% in zone 2 and only 14.5% in zone 3 versus 25.6%, 47.6% and 26.8% for control genes, respectively (Fig. 1f; representative images of control genes are shown in Fig. 1e). Considering an average of about 7 μm for the nuclear diameter in $CD4^+$ T cells, 63% of RIGs and hotter-zone alleles were concentrated within about 1 μm below the nuclear membrane.

We wanted, therefore, to visualize the position of the HIV-1 DNA itself in infected, primary $CD4^+$ T-cell nuclei. At 4 days after infection with the VSV-G-pseudotyped HIV-$1_{NL4\text{-}3/E\text{-}R\text{-}}$[14], the vast majority of the proviral immuno-FISH signals were in zone 1 (75.2% within 1 μm under the nuclear envelope) (Fig. 2c). The visualized viral DNA was integrated[15], as also detected by real-time Alu PCR (Fig. 2a), and transcriptionally active (Fig. 2b). A similar distribution was observed in primary macrophages and the monocytic cell line U937 (Extended Data Fig. 3a, b, respectively). Peripheral localization was also observed for a fully competent virus carrying the HIV-1_{BRU} envelope[16] (Fig. 2d) and, notably, for the wild-type viruses found in $CD4^+$ T-cells from two HIV-infected patients (Fig. 2e, f). Peripheral localization was also a feature of lentiviral vectors, irrespective of their transcriptional activity (Extended Data Fig. 3c, d), but not of the MoMLV gammaretrovirus, which localized preferentially inside the nuclear interior (Extended Data Fig. 3e).

In contrast, when integration was impaired, the viral cDNA roamed around the nucleus. This was the case for two HIV-1 clones harbouring single-point mutations in the integrase catalytic domain (class I IN mutations: IN(D64E) and IN(D116N))[17,18] or for HIV-$1_{NL4\text{-}3/E\text{-}R\text{-}}$ in the presence of the integrase inhibitor raltegravir; under these conditions, only 10–20% of viruses were found in zone 1 (Fig. 2g). In these cases, the detected viral genomes did not correspond to integrated DNA (Fig. 2h) but were highly enriched in circular forms of viral DNA containing two long terminal repeats (2-LTR circles) (Fig. 2i). We also downregulated the chromatin tethering factor LEDGF/p75 (ref. 19) and the inner nuclear basket protein Nup153 (ref. 20), which are involved in viral DNA integration (Fig. 2j). FISH was performed 48 h after infection when there

[1]Molecular Medicine Laboratory, International Centre for Genetic Engineering and Biotechnology (ICGEB), 34149 Trieste, Italy. [2]Protein Structure and Bioinformatics Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), 34149 Trieste, Italy. [3]Struttura Complessa Malattie Infettive, Azienda Ospedaliero-Universitaria, 34134 Trieste, Italy. [4]Department of Medical, Surgical and Health Sciences, University of Trieste, 34129 Trieste, Italy. [5]Department of Life Sciences, University of Modena and Reggio Emilia, 41121 Modena, Italy. [6]Genethon, 91002 Evry, France. †Present addresses: Department of Infectious Diseases, Integrative Virology, University Hospital Heidelberg and German Center for Infection Research, 69120 Heidelberg, Germany (B.L.; M.L.); Laboratorio Nazionale Consorzio Interuniversitario per le Biotecnologie (LNCIB), 34149 Trieste, Italy (K.L.); Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA (L.M.); Pazmany University, Budapest 1083, Hungary (S.P.).

*These authors contributed equally to this work.

§These authors jointly supervised this work.

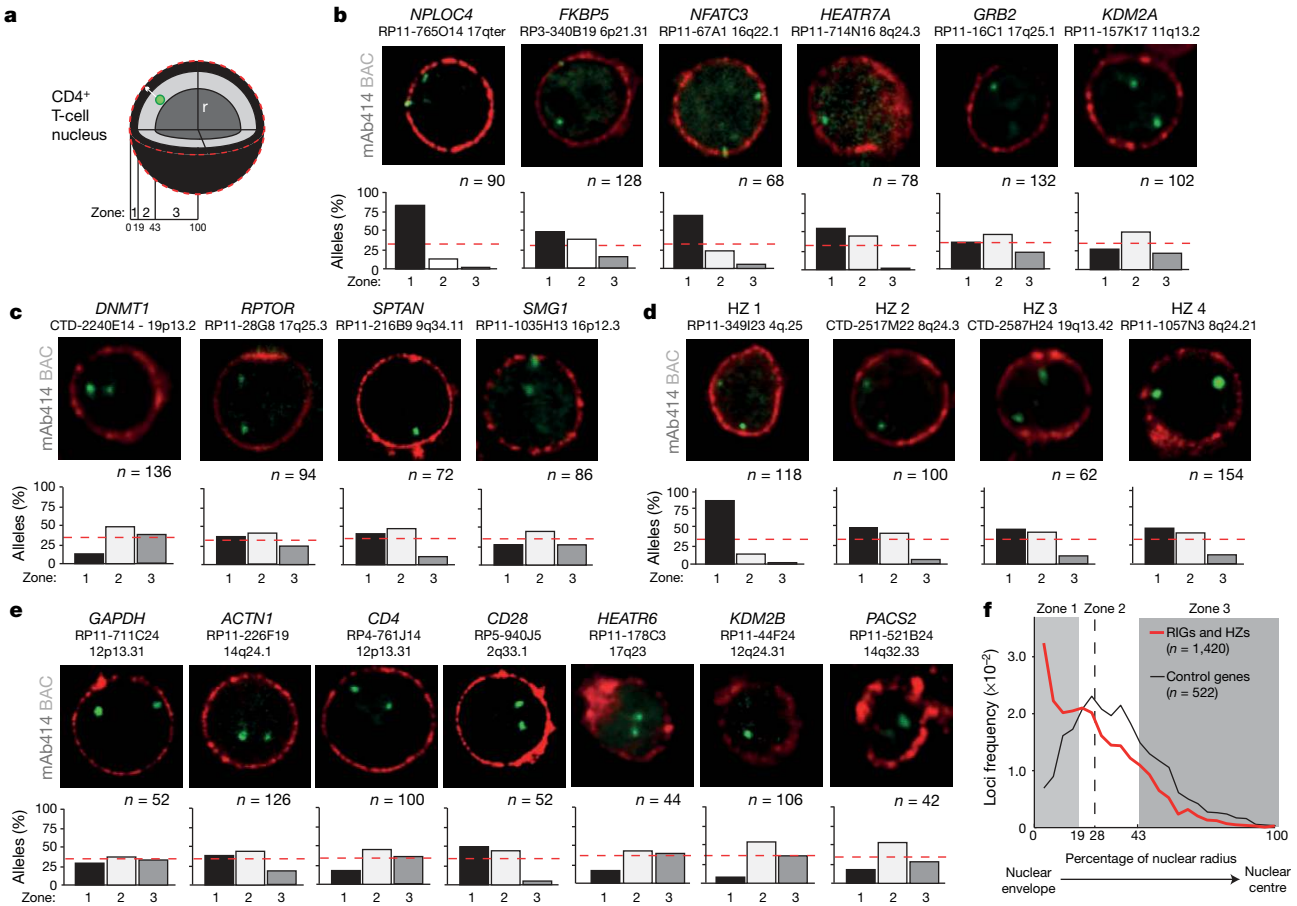


Figure 1 | Localization of HIV RIGs at the nuclear periphery. **a**, Subdivision of nucleus into three concentric zones of equal area. **b**–**e**, Three-dimensional immuno-DNA FISH of ten HIV RIGs (**b**, **c**), four hotter zones (**d**) and seven control genes (**e**) in activated CD4$^+$ T cells (green: bacterial artificial chromosome (BAC) probe labelled with DIG (dUTP-digoxygenin) and FITC (fluorescein isothiocyanate); red: NPC staining by mAb414). Below each representative image, the distribution of the analysed alleles into the three nuclear zones is shown, normalized over nuclear radius. Evenly distributed random genes would be enriched equally in the three zones (red dashed line). The number of alleles analysed is shown at the bottom of each panel. HZ, hotter zone. **f**, Distribution of the relative distances of all measured alleles from the nuclear envelope (HIV RIGs and hotter zones: $n = 1{,}420$; control genes: $n = 522$). The three zones are shown by grey shading. The dashed line indicates approximately 1 μm from the nuclear edge of the T-cell nucleus.

was a marked reduction in HIV-1 integration (Fig. 2k), and the majority of the signals labelled unintegrated viral DNA[16]: 72% and 77% of the FISH signals were in zones 2 and 3 for the LEDGF/p75 and Nup153 knockdowns, respectively (Fig. 2l). The effect of the Nup153 knockdown was rescued by transfecting an expression plasmid coding for an RNA interference (RNAi)-resistant Nup153 (Extended Data Fig. 4).

Most of the HIV-1 targets are common in different cell types; however, subtle differences exist. For example, HIV-1 almost never targets the *IKZF3* locus in CD34$^+$ haematopoietic stem cells[21] ($P < 1 \times 10^{-12}$), whereas the *TAP2* gene from the major histocompatibility complex class II locus is never targeted in CD4$^+$ T cells ($P < 1 \times 10^{-13}$). Strikingly, we observed that *IKZF3* localized in zones 1 and 2 in peripheral blood CD4$^+$ T cells (>80% of alleles), while it was almost absent from zone 1 in cord blood CD34$^+$ cells (<6% alleles; $P < 0.001$). Conversely, the *TAP2* locus was absent from zone 1 in CD4$^+$ T cells (<8% of alleles), while it was distributed between zones 1 and 2 in CD34$^+$ cells (>90% of alleles; $P < 0.001$; Fig. 3a).

To understand the chromatin features of RIGs, we compared the available data from chromatin immunoprecipitation sequencing (ChIP-seq) obtained in CD4$^+$ T cells for RIGs[22], cold genes (defined as transcriptionally inactive genes never targeted by HIV-1; F.M. and A.R., unpublished observations) and a list of genes corresponding to the 1,000 most expressed (active) and 1,000 least expressed (silent) genes from the GNF SymAtlas[13]. Association of RNA Pol2 with RIGs had a pattern superimposable to that of active genes, peaking at the transcription start sites (TSSs) (Fig. 3b). In a similar manner, distribution of markers of active transcription (H3K9ac, H3K36me3, H3K4me3, H4K16ac and H4K20me) was identical for RIGs and active genes (Fig. 3c–e and Extended Data Fig. 5). In contrast, markers of facultative (H3K9me2) and constitutive (H3K9me3 and H3K27me3) chromatin were found enriched both on cold genes (where HIV-1 never integrates) and on silent genes, but not on RIGs (Extended Data Fig. 5). Of interest, active genes and RIGs had a superimposable distribution of H3K4me2, which is enriched at the lamin-associated domain (LAD) borders[5] (Fig. 3f).

Heterochromatic LADs contain approximately 4,000 transcriptionally inactive genes[5]. We found that more than 90% of HIV RIGs lay outside LADs, while almost 80% of cold genes were inside LADs ($P < 0.001$ compared with a random gene distribution; Fig. 3g). Immuno-FISH images for three of these cold genes confirmed their localization close to the nuclear envelope in primary CD4$^+$ T cells (Fig. 3h). Finally, when all the 1,344 known LADs were aligned by their left or right borders, 87.2% of RIGs were found outside the LADs, in contrast to a random distribution of genes (68.2%; $P < 0.001$, also taking into account the lower gene density within LADs; Fig. 3i).

Transcriptionally active genes at the nuclear periphery are often associated with the nuclear pore complex (NPC)[23–27]. We therefore assessed interaction of the HIV-1 provirus with the NPC by ChIP assays in primary CD4$^+$ T cells (primer scheme and controls in Extended Data Fig. 6a, b). At 4 days after infection, when RNA Pol2 and the USF1 and p65/RelA transcription factors were associated with the viral DNA as expected[14], both the mAb414 antibody, which recognizes phenylalanine–glycine (FG)-repeats in nucleoporins, and specific antibodies against Nup153,

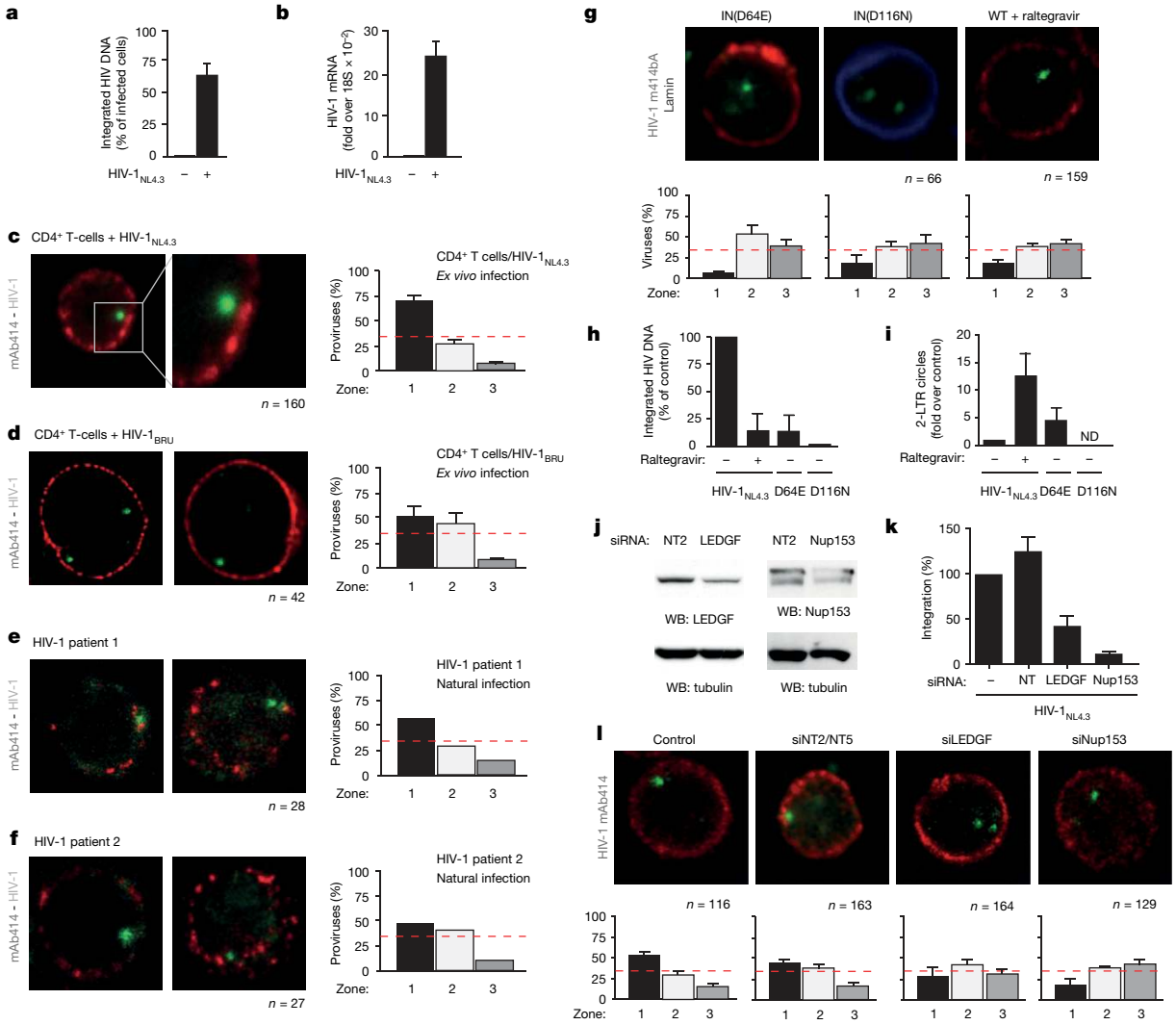


Figure 2 | Integrated, transcriptionally active HIV-1 is found at the nuclear periphery. **a**, **b**, Quantification of integrated HIV-$1_{NL4\text{-}3/E\text{-}R\text{-}}$ DNA (**a**) and HIV RNA (**b**) by real-time Alu PCR in infected $CD4^+$ T cells. **c**–**f**, Three-dimensional immuno-DNA FISH of HIV-1 DNA (green) in primary $CD4^+$ T cells infected *ex vivo* with HIV-$1_{NL4\text{-}3/E\text{-}R\text{-}}$ (**c**) and HIV-1_{BRU} (**d**), or directly obtained from two patients infected with HIV-1 (**e**, **f**). **g**, Three-dimensional immuno-DNA FISH of HIV-1 DNA in activated $CD4^+$ T cells infected with the mutant viruses IN(D64E) or IN(D116N) or with HIV-$1_{NL4\text{-}3/E\text{-}R\text{-}}$ upon raltegravir treatment. **h**, **i**, Real-time Alu PCR (**h**) and 2-LTR quantification (**i**) in the cells treated as in **g**. ND, not determined. **j**, Western blot (WB) showing protein levels for LEDGF/p75 and Nup153 at the moment of HIV-1 infection, 36 h after short interfering RNA (siRNA) transfection. NT2, non-targeting siRNA. **k**, Real-time Alu PCR in Jurkat cells infected with HIV-$1_{NL4.3}$ and previously transfected with a non-targeting siRNA (NT) or an siRNA targeting LEDGF/p75. Samples were normalized over control-infected cells. **l**, Three-dimensional immuno-DNA FISH for HIV-1 DNA visualization upon Jurkat cell treatment with the indicated siRNAs. NT2 and NT5 are two non-targeting siRNAs. All graphs, except those relative to patients' cells, show mean and s.e.m. of at least three independent experiments.

Nup98, Nup62 and Tpr all immunoprecipitated the HIV-1 DNA; binding was also observed for the *NPLOC4* RIG gene, but not for the LAD gene *PTPRD* (Extended Data Fig. 6c). When ChIP was performed on the IN-defective D64E virus, no viral DNA was detected using the mAb414 and anti-Nup153 antibodies (Extended Data Fig. 6d).

Next, we aimed to verify whether HIV-1 localization changed when the virus reverted from a transcriptionally inactive to an active state. In the latent T-cell J-Lat clone 15.4 (ref. 28), the HIV-1 DNA retained its gross peripheral localization both in inactive and in TPA (12-*O*-tetradecanoylphorbol-13-acetate) phorbol ester-reactivated conditions (Extended Data Fig. 7a, b). Similar results were obtained in a primary model of HIV-1 latency[14] (Extended Data Fig. 7c–e). However, when localization was analysed at molecular resolution by ChIP using the mAb414, anti-Tpr and anti-Nup153 antibodies, binding of the proviral region located downstream of the TSS to the nucleoporins was observed upon transcriptional activation but not in latent conditions (Extended Data Fig. 7f). We also observed that nucleoporins directly participated in HIV-1 transcriptional regulation. When Tpr and Nup153 were silenced by RNAi in latent J-Lat cells, proviral transcription was significantly reduced (Extended Data Fig. 7g, h). Similarly, downregulation of Tpr also blunted LTR-driven gene expression in HIV-1-infected HeLa cells (Extended Data Fig. 8a–e).

Our findings show that the cellular genes that are highly targeted by HIV-1 are distributed in a topologically non-random manner, being positioned within 1 μm from the nuclear edge; these genes are enriched in open chromatin marks, excluded from the LADs and associated with the NPC. Thus, the HIV-1 pre-integration complex preferentially targets those areas of open chromatin that are proximal to the nuclear pore, while excluding the internal regions in the nucleus as well as the peripheral regions associated with the nuclear lamina (model in Extended Data Fig. 9). The localization of HIV-1 proviral DNA in close association with the nuclear pore is consistent with several observations showing that different NPC components play a role in HIV-1 infection[20,29,30].

Why does the viral DNA integrate into the NPC compartment? A possibility that we favour is that the virus simply integrates into the first open chromatin regions it meets along its route into the nucleus. This is

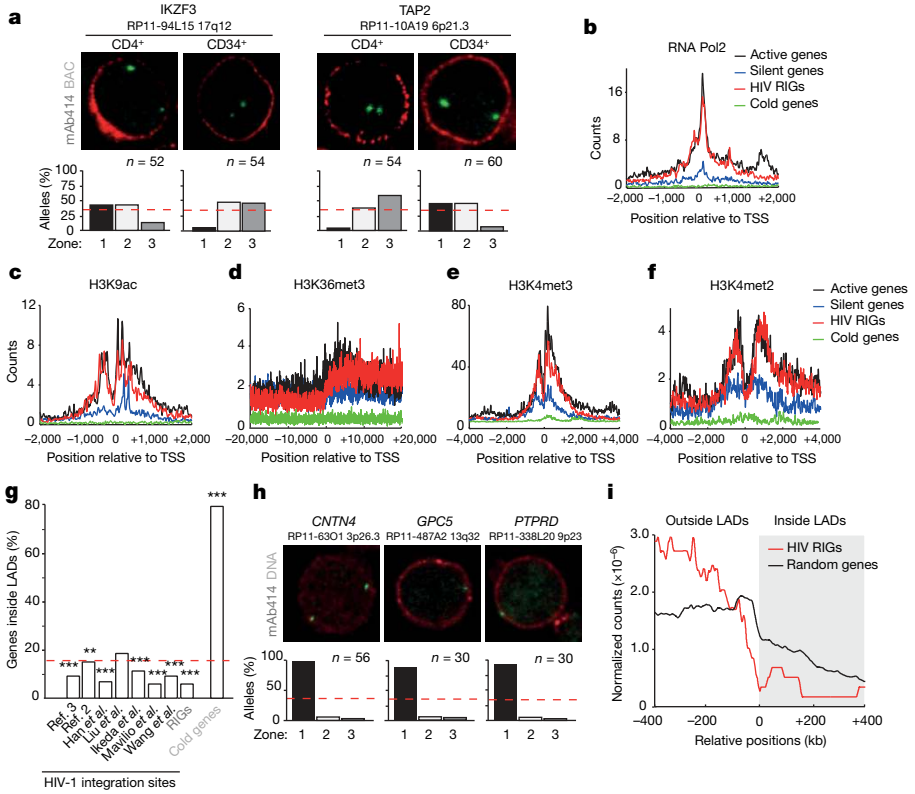


Figure 3 | HIV RIGs are transcriptionally active genes that are excluded from the LADs. **a**, Localization of the *IKZF3* (**a**) and *TAP2* (**b**) genes (green) in $CD4^+$ T cells and $CD34^+$ haematopoietic stem cells. **b**–**f**, Distributions of Pol2, acetylated H3K9, H3K36me3 and H3K4me2/3 around the TSSs of HIV RIGs (red) and cold genes (green), compared with highly active (black) and silent (blue) genes in activated $CD4^+$ T cells. **g**, Cross-comparison of different lists of integration loci, including HIV RIGs, with the lists of genes present inside LADs: HIV integration loci are significantly depleted in LADs compared with a null distribution (indicated by a red dotted line). $***P < 0.001$, $**P < 0.01$, $*P < 0.05$. References with authors' names can be found in the reference list at the end of the Methods section. **h**, Three-dimensional immuno-DNA FISH in activated $CD4^+$ T cells of three cold genes predicted to be inside LADs by bioinformatics analysis. **i**, Distribution of HIV RIGs (red) and of a random set of genes (black) around aligned LAD border regions. The light grey area with positive genomic coordinates indicates the regions inside LADs; the white area with negative coordinates is outside LADs.

likely to be related to the short life of viral integrase[16] and thus the need, for the pre-integration complex, to achieve rapid integration into genomic DNA upon its entry into the nucleus. This interpretation is consistent with our observation of more dispersed, unintegrated viral cDNA in all conditions in which integrase function is impaired.

Finally, while adding a three-dimensional view to the process of HIV-1 integration, our results also indicate that the localization of the HIV-1 DNA in close correspondence with the nuclear pore has functional relevance, since it appears important for productive HIV-1 gene expression.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 December 2013; accepted 9 January 2015.
Published online 2 March 2015.

1. Wang, G. P., Ciuffi, A., Leipzig, J., Berry, C. C. & Bushman, F. D. HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res.* **17,** 1186–1194 (2007).
2. Schroder, A. R. *et al.* HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110,** 521–529 (2002).
3. Brady, T. *et al.* HIV integration site distributions in resting and activated $CD4^+$ T cells infected in culture. *AIDS* **23,** 1461–1471 (2009).
4. Sherrill-Mix, S. *et al.* HIV latency and integration site placement in five cell-based models. *Retrovirology* **10,** 90 (2013).
5. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453,** 948–951 (2008).
6. Cavalli, G. & Misteli, T. Functional implications of genome topology. *Nature Struct. Mol. Biol.* **20,** 290–299 (2013).
7. Maldarelli, F. *et al.* HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* **345,** 179–183 (2014).
8. Wagner, T. A. *et al.* Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* **345,** 570–573 (2014).
9. Kin, T. & Ono, Y. Idiographica: a general-purpose web application to build idiograms on-demand for human, mouse and rat. *Bioinformatics* **23,** 2945–2946 (2007).
10. Nagai, S. *et al.* Functional targeting of DNA damage to a nuclear pore-associated SUMO-dependent ubiquitin ligase. *Science* **322,** 597–602 (2008).
11. Hediger, F., Neumann, F. R., Van Houwe, G., Dubrana, K. & Gasser, S. M. Live imaging of telomeres: yKu and Sir proteins define redundant telomere-anchoring pathways in yeast. *Curr. Biol.* **12,** 2076–2089 (2002).
12. Wu, C. *et al.* BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.* **10,** R130 (2009).
13. Su, A. I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA* **101,** 6062–6067 (2004).
14. Lusic, M. *et al.* Proximity to PML nuclear bodies regulates HIV-1 latency in $CD4^+$ T cells. *Cell Host Microbe* **13,** 665–677 (2013).
15. Butler, S. L., Hansen, M. S. & Bushman, F. D. A quantitative assay for HIV DNA integration *in vivo*. *Nature Med.* **7,** 631–634 (2001).
16. Manganaro, L. *et al.* Concerted action of cellular JNK and Pin1 restricts HIV-1 genome integration to activated $CD4^+$ T lymphocytes. *Nature Med.* **16,** 329–333 (2010).
17. Lu, R., Limon, A., Ghory, H. Z. & Engelman, A. Genetic analyses of DNA-binding mutants in the catalytic core domain of human immunodeficiency virus type 1 integrase. *J. Virol.* **79,** 2493–2505 (2005).
18. Negri, D. R. *et al.* Successful immunization with a single injection of non-integrating lentiviral vector. *Mol. Ther.* **15,** 1716–1723 (2007).
19. Shun, M. C. *et al.* LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev.* **21,** 1767–1778 (2007).
20. Matreyek, K. A., Yucel, S. S., Li, X. & Engelman, A. Nucleoporin NUP153 phenylalanine-glycine motifs engage a common binding pocket within the HIV-1 capsid protein to mediate lentiviral infectivity. *PLoS Pathog.* **9,** e1003693 (2013).
21. Cattoglio, C. *et al.* High-definition mapping of retroviral integration sites defines the fate of allogeneic T cells after donor lymphocyte infusion. *PLoS ONE* **5,** e15688 (2010).
22. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129,** 823–837 (2007).
23. Capelson, M. *et al.* Chromatin-bound nuclear pore components regulate gene expression in higher eukaryotes. *Cell* **140,** 372–383 (2010).

24. Kalverda, B., Pickersgill, H., Shloma, V. V. & Fornerod, M. Nucleoporins directly stimulate expression of developmental and cell-cycle genes inside the nucleoplasm. *Cell* **140,** 360–371 (2010).
25. Vaquerizas, J. M. *et al.* Nuclear pore proteins nup153 and megator define transcriptionally active regions in the *Drosophila* genome. *PLoS Genet.* **6,** e1000846 (2010).
26. Liang, Y., Franks, T. M., Marchetto, M. C., Gage, F. H. & Hetzer, M. W. Dynamic association of NUP98 with the human genome. *PLoS Genet.* **9,** e1003308 (2013).
27. Light, W. H. *et al.* A conserved role for human Nup98 in altering chromatin structure and promoting epigenetic transcriptional memory. *PLoS Biol.* **11,** e1001524 (2013).
28. Jordan, A., Bisgrove, D. & Verdin, E. HIV reproducibly establishes a latent infection after acute infection of T cells *in vitro*. *EMBO J.* **22,** 1868–1877 (2003).
29. Brass, A. L. *et al.* Identification of host proteins required for HIV infection through a functional genomic screen. *Science* **319,** 921–926 (2008).
30. Konig, R. *et al.* Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell* **135,** 49–60 (2008).

Acknowledgements This work was supported by grants from the Italian National Research Programme on AIDS of the Istituto Superiore di Sanità, Italy, to M.G. and M.L. and from the Young Investigator Grant RF2007-16 of the Italian Ministry of Health to M.L. The authors are grateful to S. Kerbavcic for editorial assistance.

Author Contributions B.M., B.L., K.L. and M.L. performed the immuno-DNA FISH and ChIP experiments; A.K.-F., B.M., S.P., M.L. and M.G. analysed the data; A.K.-F., B.M. and S.P. performed the bioinformatics analysis; H.A. and M.L. performed the experiments using infectious virus; L.M. generated and analysed integrase-defective HIV-1 molecular clones; R.L. contributed to studies in primary cells from patients with HIV; A.R. and F.M. generated lentiviral vectors and analysed integration into $CD4^+$ T cells and $CD34^+$ bone marrow cells; M.L. and M.G. conceived and supervised the experiments and wrote the paper with help from the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.G. (giacca@icgeb.org) or M.L. (marina.lusic@med.uni-heidelberg.de).

METHODS

Virus production. For the production of viral stocks, we used a plasmid obtained from the Env- molecular clone pNL4-3/E$^-$R$^-$, a gift from N. Landau. This viral clone harbours a frameshift mutation introduced near the 5′ end of the *env* gene[31] and performs a single-round infection once pseudotyped with vesicular stomatitis virus-G (VSV-G); this renders the virus incapable of spreading (and causing massive cell death).

We produced viral clone HIV-1_{BRU} as previously described[16]. The integrase (IN) defective packaging plasmid pcHelp/IN-, a gift from A. Cara, contains a D116N mutation in the IN genome, preventing the function of the IN protein[18]. The D64E mutant plasmid, which is similarly integration-defective, was obtained from the National Institutes of Health (NIH) AIDS Research and Reference Reagent Program[17].

Lentiviral vector pLV-THM[32] was obtained from Addgene, whereas pCCL18–green fluorescent protein (GFP) was modified from ref. 33 to become promoterless. The gammaretroviral MoMLV vector was a gift from G. Towers.

Infectious viral stocks were generated by transfecting viral DNA in HEK 293T cells and collecting supernatants after 48 h. Viral production was quantified by measuring viral p24 in the supernatants using the Innotest HIV antigen mAB kit (Innogenetics).

Primary cell isolation, culture and infection. Primary human CD4$^+$ T cells were isolated by Ficoll gradient separation, followed by purification with CD4 MicroBeads (Miltenyi Biotec). Cells were activated with a cocktail of beads containing 4.5×10^5 beads coated with αCD3 and αCD28 antibodies (Dynabeads Human T-Activator CD3/CD28 Dynal/Invitrogen), and plated in complete medium with interleukin-2 (IL-2; 30 U ml^{-1}, Sigma-Aldrich) for 4 days at 37 °C.

Activated CD4$^+$ T cells (1×10^6) were infected with 0.5–0.75 μg ml^{-1} of viral p24 for 4–5 h at 37 °C. After infection, cells were kept in culture at 1×10^6 cells per millilitre in complete RPMI 1640 medium supplemented with IL-2 and CD3/CD28 beads.

For raltegravir treatment, 10 μM raltegravir (obtained from the NIH AIDS Research and Reference Reagent Program) was added together with the virus during the infection, and it was later supplemented in the medium.

For generating the primary model of latency, naive CD4$^+$ T cells were isolated, cultured and infected as described in ref. 14.

Primary human CD34$^+$ cells were isolated, cultured and infected as described in ref. 34. Primary human macrophages were isolated and cultured as described in ref. 35.

Patients infected with HIV and having a CD4 count less than 3×10^5 ml^{-1} were enrolled before starting highly active antiretroviral therapy, following informed consent. Peripheral blood mononuclear cells obtained from healthy or infected blood donors according to a study protocol approved by the Ethical Committee of the Azienda Ospedaliero-Universitaria 'Ospedali Riuniti di Trieste', Italy, were isolated as described previously[16].

Cell culture and transfection. The Jurkat lymphoblastoid cell line, Jurkat J-Lat 15.4 clone and U937 monocytic cell line were kept in culture in complete RPMI 1640 medium with the addition of 10% fetal bovine serum (FBS). Cells were tested for mycoplasmal cell culture contaminants by using a MycoAlert kit from Lonza.

Transfection of Jurkat with p-eGFP–Nup153 expression plasmid[36] obtained from Euroscarf was done using Fugene HD (Promega), according the manufacturer's instructions. For the RNAi experiments, 3×10^6 cells were transfected with siRNA smart pools targeting LEDGF/p75 (PSIP1) (Dharmacon M-015209) or Nup153 (Dharmacon M-005283) proteins or with a non-targeting (NT) siRNA as a negative control (Dharmacon, Thermo Scientific). Transfection was performed with the Amaxa Nucleofection Device II (Amaxa), using an Amaxa nucleofection Kit V according to the manufacturer's instructions.

For western blot analysis, cells were harvested and homogenized in lysis buffer (20 mM Tris-HCl, pH 7.4, 1 mM EDTA, 150 mM NaCl, 0.5% Nonidet P-40, 0.1% SDS, 0.5% sodium deoxycholate) supplemented with protease inhibitors (Roche) for 10 min at 4 °C and sonicated (Bioruptor) for 5 min. Equal amounts of total cellular proteins (30 μg), as measured with Bradford reagent (Biorad), were resolved by 8% SDS–polyacrylamide gel electrophoresis (SDS–PAGE), transferred onto polyvinylidene difluoride membranes (GE Healthcare) and then probed with primary antibody (anti-Nup153 (Santa Cruz, sc-101544), anti-LEDGF/p75 (BD Biosciences, 611714), anti-GFP (LifeTechnologies, A6455)), followed by secondary antibody conjugated with horseradish peroxidase. The immunocomplexes were visualized with enhanced chemiluminescence kits (GE Healthcare).

Once protein silencing was assessed, cells were infected with 0.5–0.75 μg ml^{-1} of p24 of viral clone $NL_{4\text{-}3/E\text{-}R\text{-}}$ as described. At 24 and 48 h after infection, samples were collected for further analysis.

Integration assay (Alu PCR). Infected cells were tested for integration of HIV-1 by isolating genomic DNA from 1×10^6 cells with a DNeasy Tissue Kit (Qiagen). Genomic DNA (100 ng) was subjected to quantitative Alu-LTR PCR for integrated provirus or for 2-LTR circles as previously described[16].

Quantitative reverse transcription PCR. For the quantification of HIV transcript levels, RNA was purified from the cells with a Nucleospin RNA II purification kit (Macherey-Nagel). The messenger RNA (mRNA) levels were quantified by TaqMan quantitative reverse transcription PCR (qRT–PCR) using HIV-1 or interleukin-2 (IL-2) primers and probe[14], and housekeeping gene 18S and GAPDH as controls.

Luciferase activity assay. Cells were harvested 48 h after infection, and luciferase activity was measured using the Luciferase Assay Kit (Promega). Viral expression was expressed after normalization over micrograms of total cell extracts.

Cell preparation for three-dimensional immuno-DNA FISH. Three-dimensional FISH combined with immunostaining was performed according to protocols in ref. 37. Culture or primary cells were resuspended at 3×10^6 cells per millilitre in 5% FBS in PBS and cell suspension was allowed to attach to the glass cover slips, previously coated with poly-L-lysine. Cells were fixed in 4% paraformaldehyde in 0.3× hypotonic PBS for 10 min, permeabilized with PBS/0.5% Triton X-100 for 10 min and left in PBS/20% glycerol for 1 h. Cells were then blocked in PBS/5% fetal horse serum (FHS) for 45 min, and primary antibody (anti-NPC mAb414, Covance; anti-LaminB, Abcam ab16048; anti-GFP, LifeTechnologies A6455) was added for an overnight incubation at +4 °C in a humid chamber. The subsequent day, cells were washed five times in PBS-T (PBS with 0.05% Tween) and the secondary antibody (Jackson Laboratories) was used for 45 min at 22 °C (1/1,000 dilution). After five washings in PBS-T, cells were additionally crosslinked with EGS (ethylene glycol-bis(succinic acid *N*-hydroxysuccinimide ester) (Sigma E-3257) for 10 min, washed and permeabilized again in PBS-T/0.5% Triton X-100. After washing in PBS/0.05% Triton X-100, cells were rinsed and incubated in 0.1 N HCl (freshly prepared) for 10 min. Cells were left in PBS/20% glycerol for at least 45 min, and then subjected to five cycles of freeze and thaw in liquid nitrogen and PBS/20% glycerol. Additional washings in PBS/0.05% Triton X-100 preceded an overnight incubation in 50% formamide/2× SSC (hybridization buffer). The subsequent day, cells were treated with RNase A (100 μg ml^{-1} in 2× SSC) in a humid chamber at 37 °C for 1 h, were rinsed again in 50% formamide/2× SSC for at least 1 h (or overnight) and were then subjected to hybridization with the appropriate probe.

Probes for hybridization in three-dimensional immuno-DNA FISH. For visualization of the loci of interest, specific BAC clones (selected from CHORI (Children's Hospital Oakland Research Institute in Oakland, California) sites and purchased from Invitrogen) were isolated according to the manufacturer's instructions. The listing of the BACs with their identities and the genes they contain is provided in the Supplementary Information. BAC DNA (2 μg) was labelled with digoxigenin by Dig-Nick Translation (Roche) at 15 °C.

For the visualization of HIV-1, lentiviral or gamma-retroviral vector DNA integrated inside Jurkat or primary CD4$^+$ T cells, 2 μg of the respective plasmids was labelled by nick translation in the presence of 16-dUTP Biotin nucleotides at 15 °C for 3 h.

In both cases, probes were checked on agarose gel and then cleared by using an Illustra Microspin G-25 column (GE Healthcare) and precipitated in the presence of Cot-1 DNA (Roche) and DNA from herring sperm (Sigma). Finally, after ethanol precipitation, the probes were resuspended in 10 μl formamide, incubated at 37 °C for 15–20 min and 10 μl of 20% dextran in 4× SSC was added to a final volume of 20 μl.

Hybridization set up and development. The probe (1–10 μl) was loaded onto glass cover slips with the cells, followed by sealing with rubber cement, and heat-denatured on a heat block at 75 °C for 4 min. Hybridization was performed for 48 h at 37 °C in a humid chamber. Three washings in 2× SSC (10 min each) were followed with three washings in 0.5× SSC at 56 °C.

FISH development for Dig-labelled BACs was performed by using FITC-labelled anti-digoxigenin antibody (Roche), whereas biotin-labelled HIV-1 probes were detected by a TSA Plus system from Perkin Elmer, allowing signal amplification, by using an anti-biotin antibody (SA-HRP) and a secondary antibody with a fluorescent dye (usually FITC for HIV).

Microscopy. Three-dimensional-stacks of slides with fixed cells were captured on a Zeiss LSM 510 META confocal microscope (Carl Zeiss Microimaging) with a ×63 numerical aperture 1.4 Plan-Apochromat oil objective. The pinhole of the microscope was adjusted to obtain an optical slice of less than 1.0 μm for any wavelength acquired.

Distances observed between the FISH signals and the nuclear envelope were measured using LSM 510 Image Examiner Software (Zeiss) and Volocity (Perkin Elmer); measurements were normalized over nuclear radius (defined as half of the middle of the mAb414-TRITC ring), and then binned into three classes of equal surface area[11].

Measurements were acquired for the alleles of the following genes in activated CD4$^+$ T cells: *NPLOC4* ($n = 90$), *FKBP5* ($n = 128$), *NFATC3* ($n = 68$), *HEATR7A* ($n = 78$), *RPTOR* ($n = 94$), *SPTAN* ($n = 72$), *SMG1* ($n = 80$), *GRB2* ($n = 132$), *KDM2A* ($n = 102$), *DNMT1* ($n = 136$), hotter zone 1 ($n = 118$), hotter zone 2 ($n = 100$), hotter zone 3 ($n = 62$), hotter zone 4 ($n = 154$), *GAPDH* ($n = 52$), *ACTN1* ($n = 126$),

CD4 ($n = 100$), *CD28* ($n = 52$), *HEATR6* ($n = 44$), *KDM2B* ($n = 106$), *PACS2* ($n = 42$), *IKZF3* ($n = 52$), *TAP2* ($n = 54$), *CNTN4* ($n = 56$), *GPC5* ($n = 30$), *PTPRD* ($n = 30$). Measurements were acquired for the alleles of the following genes in $CD34^+$ cells: *IKZF3* ($n = 54$) and *TAP2* ($n = 60$). Measurements were acquired for the proviruses in primary macrophages ($n = 18$) and the U937 cell line ($n = 30$). Measurements were acquired for the proviruses in primary cells upon several conditions: 4 days after infection of activated $CD4^+$ T cells ($n = 160$ HIV-$1_{NL4\text{-}3/E\text{-}R\text{-}}$ and $n = 42$ HIV-1_{BRU} measured in three independent experiments); HIV-1 in $CD4^+$ T cells from infected patients ($n = 28$ and $n = 27$); 4 days after infection with mutant viruses IN(D64E) ($n = 30$) or IN(D116N) ($n = 66$), or cells infected with HIV-$1_{NL4.3}$ upon raltegravir treatment ($n = 159$); latent $CD4^+$ cells (2 weeks after infection) with or without CD3/CD28 stimulation ($n = 40$ and $n = 33$, respectively). Measurements were acquired for proviruses in Jurkat or J-Lat cell lines 4 days after infection of Jurkat with HIV-$1_{NL4\text{-}3/E\text{-}R\text{-}}$, in different conditions: no transfection (control, $n = 116$), transfections with non-targeting siRNA (siNT2/NT5, $n = 163$), LEDGF/p75 siRNA (siLEDGF, $n = 164$), Nup153 (siNup153, $n = 129$), Nup153 siRNA + enhanced GFP (eGFP)–Nup153 ($n = 52$). The corresponding graphs show the average results from three independent experiments. Proviral DNA was analysed in the J-Lat clone 15.4 with and without TPA ($n = 74$ and $n = 150$, respectively); Jurkat + Lentiviral promoter ($n = 19$); Jurkat + transcription-less lentiviral vector ($n = 51$); Jurkat + gammaretroviral vector ($n = 88$).

ChIP. $CD4^+$ T cells (40×10^6) were washed twice in PBS before crosslinking with 1% final formaldehyde for 10 min at room temperature, followed by termination of the reaction with 125 mM glycine on ice. The cell pellet was washed twice with PBS and was lysed in 0.5% NP-40 buffer (10 mM Tris-Cl pH7.4, 10 mM NaCl, 3 mM $MgCl_2$, 1 mM PMSF and protease inhibitors). The nuclei obtained were washed once in the same buffer without NP-40. Lysis of the nuclei was performed using the same buffer containing 4% of NP-40 at 37 °C for 15 min, upon which micrococcal nuclease was added (120 units of the enzyme), and the reaction was stopped with 3 mM EGTA. DNA was additionally sheared by sonication to an average size of DNA fragments below 500 base pairs. Extracts were pre-cleared by two rounds of incubation with immunoglobulin-γ and agarose beads, followed by centrifugation at 1,200*g* for 5–10 min. The lysate (400 μl) was then incubated with 2–4 μg of the indicated antibody overnight at 4 °C, followed by incubation for 4 h with MagnaChIP Protein A/G Magnetic Beads (Millipore). Beads were then washed thoroughly with RIPA150, with LiCl-containing buffer and with TE, RNase treated for at least 30 min at 37 °C, and treated with proteinase K for at least 2 h at 56 °C. De-crosslinking of protein–DNA complexes was performed by an overnight incubation at 65 °C. DNA was then extracted by phenol–chloroform extraction followed by ethanol precipitation and was quantified by real-time PCR. The following antibodies were used in the ChIP experiments: mAb414 (Covance, MMS-120R), anti-Pol2 (Santa Cruz, sc-9001X), anti-USF1 (Santa Cruz, sc-229X), anti-NF-κB p65 subunit (Santa Cruz, sc-109X), anti-Nup153 [SA1] (Abcam, ab-96462) and anti-Nup153 (QE5) (Abcam, ab-24700), anti-Nup98 (Cell Signaling, 2598), anti-Nup62 (BD Biosciences, 610497), anti-Tpr (Abcam ab-58344), anti-Mcm2 (Abcam), mouse IgG (Santa Cruz, sc-2025). The graphs of ChIP figures show the mean and s.e.m. from at least three independent experiments.

Bioinformatics and statistical analysis. No statistical methods were used to predetermine sample size.

Five lists of HIV-1 integration sites were collected from published work[2,3,38–40] and an unpublished list of integration sites in $CD4^+$ T cells provided by A.R. and F.M. (Extended Data Table 1). HIV-1 RIGs ($n = 156$) were genes found in more than one list (Supplementary Information and Extended Data Table 1); their genomic position was plotted onto the chromosome map using the Idiographica webtool (http://www.ncrna.org/idiographica (ref. 9); Extended Data Fig. 2). Genomic coordinates of eight selected hotter zones, into which HIV-1 integration density was found higher than expected, were downloaded from the Bushman Lab website (http://www.bushmanlab.org/tutorials/ucsc)[1] and reported in the Supplementary Information.

The calculation of the probability of finding 156 genes present in more than one list by chance was performed by computer simulation. A program was written to randomly draw, from 25,000 genes, 265, 329, 294, 32, 158 and 58 genes, and to count the genes drawn more than once. The simulation was repeated 1×10^9 times. The distribution obtained is shown in Extended Data Fig. 1. Calculations were performed using the Matlab 2011R software (http://www.mathworks.com/).

Expression of HIV RIGs and control genes was derived from published transcriptomic data in $CD4^+$ T cells, using biogps.org (as in refs 12, 13, 22), and was compared with a random sets of genes. Using the non-parametric Mann–Whitney-Wilcoxon test, it was concluded that that HIV RIGs and control genes are more transcribed than a random sets of genes ($P < 0.005$).

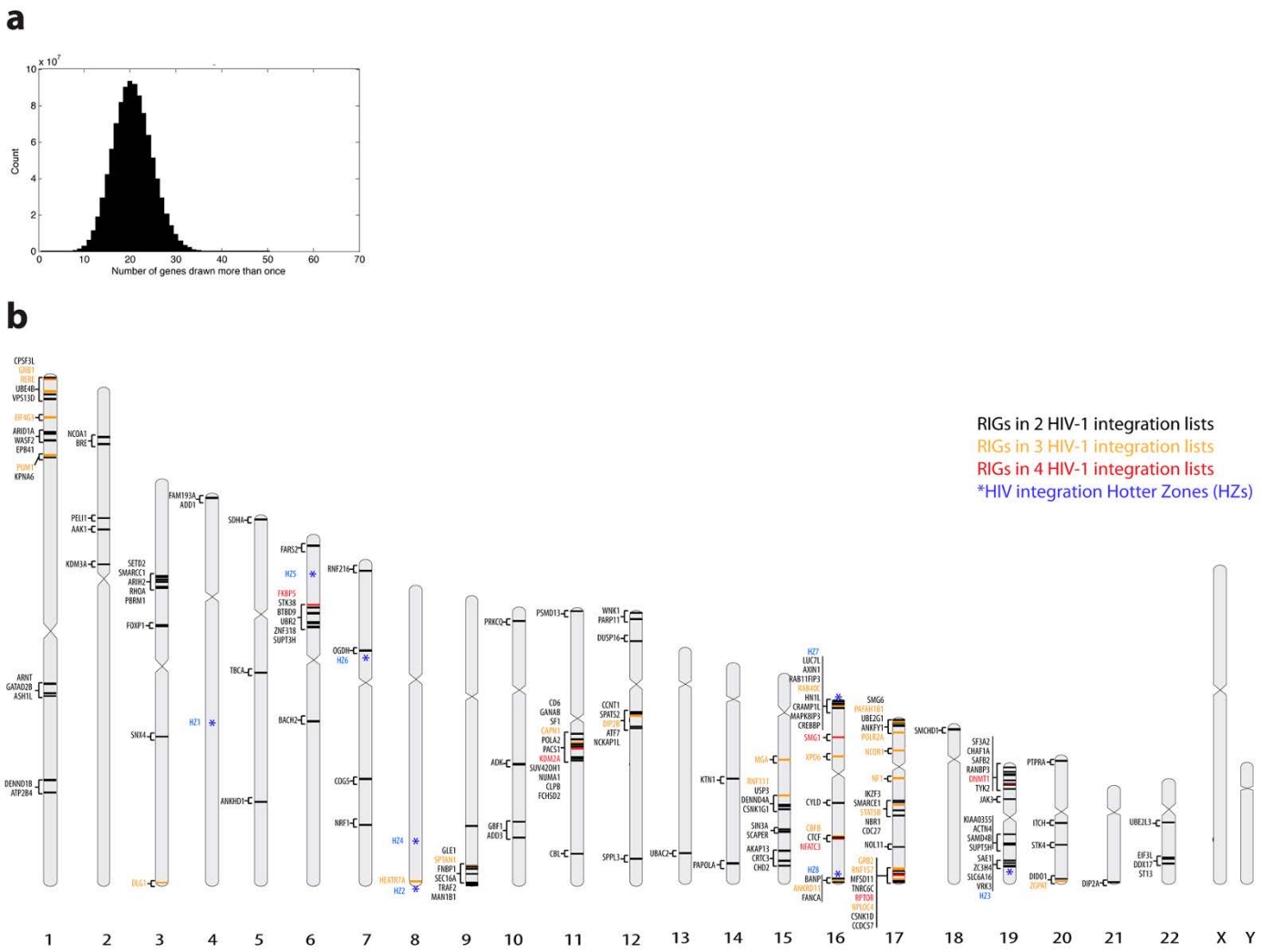
LAD coordinates were obtained from ref. 5, whereas genes inside LADs were derived from BioMart Ensembl and named LAD genes. Then, the *P* value of the common genes in the LAD genes, HIV RIGs, the six integration lists and cold genes were calculated by pairwise comparison of each combination, followed by hypergeometric test (Fig. 3g). The following *P* values were obtained: ref. 3, $P = 1.27 \times 10^{-8}$; ref. 2, $P = 0.008$; ref. 38, $P = 0.001$; ref. 40, $P = 0.36$; ref. 39, $P = 0.0007$; F.M. *et al.* (unpublished observations), $P = 1.32 \times 10^{-15}$; ref. 1 hotter zones, $P = 0.0003$; HIV RIGs, $P = 2.09 \times 10^{-10}$. Eighty per cent of genes that are never targeted by HIV-1 (cold genes) are significantly enriched inside LADs ($P = 3.25 \times 10^{-19}$).

The profile of aligned LAD border regions (Fig. 3i) was performed as described in ref. 5. A χ^2 test was applied to compare the distribution across the LAD border of HIV RIGs with the one of 3,000 random genes that were generated without replacement using the RSA-tool (http://floresta.eead.csic.es/rsat/random-genes_form.cgi)[41].

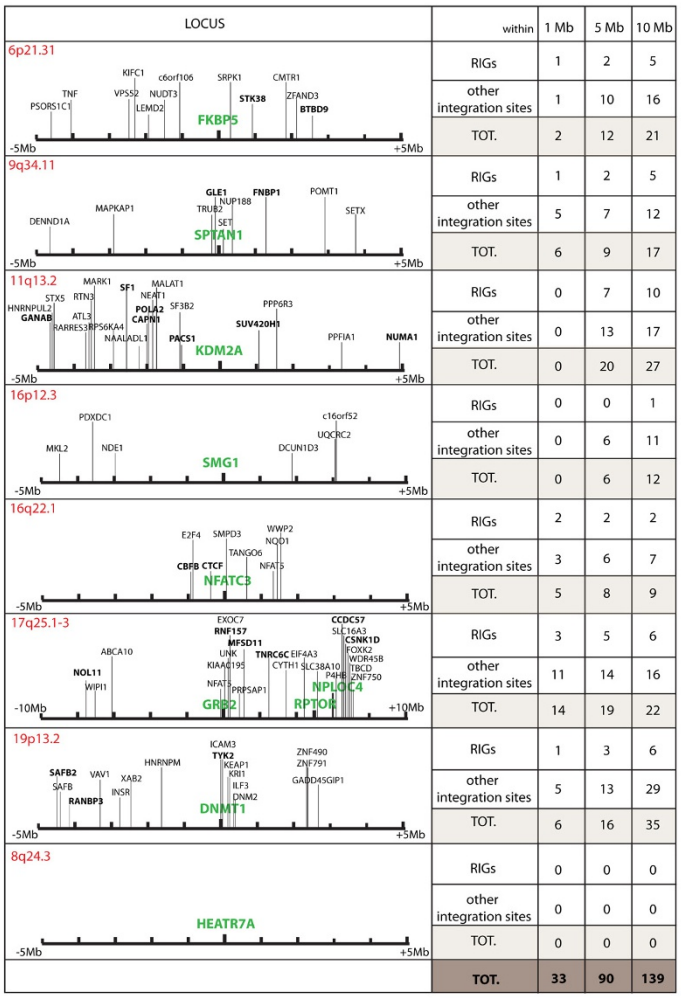
ChIP-seq profile analyses (Fig. 3b–f and Extended Data Fig. 6) were performed as in refs 22, 42. The 1,000 most expressed and 1,000 least expressed genes were obtained as in ref. 22. and named active and silent genes, respectively. The TSS coordinates of these genes were obtained using the University of California, Santa Cruz (UCSC) Table Browser.

Comparison between groups for expression data was performed using a non-parametric Mann–Whitney–Wilcoxon rank sum test; comparison of gene distributions was by a χ^2 test, with the exception of data shown in Fig. 3g, which were analysed by a hypergeometric test. For the FISH, ChIP and real-time PCR results, the reported values are means and s.e.m., calculated from at least three independent samples. For statistical comparison of three or more groups, one-way analysis of variance followed by Tukey's post-hoc test was used. A value of $P < 0.05$ was considered significant.

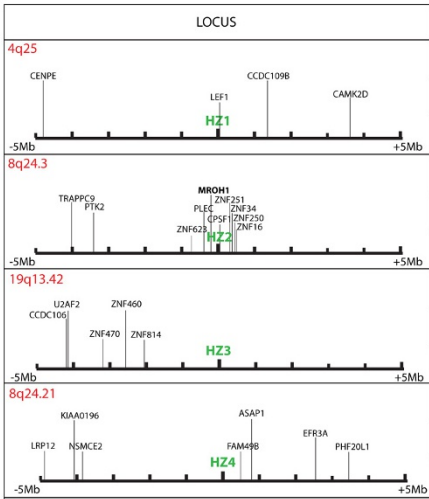
31. Connor, R. I., Chen, B. K., Choe, S. & Landau, N. R. Vpr is required for efficient replication of human immunodeficiency virus type-1 in mononuclear phagocytes. *Virology* **206,** 935–944 (1995).
32. Wiznerowicz, M. & Trono, D. Conditional suppression of cellular genes: lentivirus vector-mediated drug-inducible RNA interference. *J. Virol.* **77,** 8957–8961 (2003).
33. Dull, T. *et al.* A third-generation lentivirus vector with a conditional packaging system. *J. Virol.* **72,** 8463–8471 (1998).
34. Cattoglio, C. *et al.* Hot spots of retroviral integration in human $CD34^+$ hematopoietic cells. *Blood* **110,** 1770–1778 (2007).
35. Repnik, U., Knezevic, M. & Jeras, M. Simple and cost-effective isolation of monocytes from buffy coats. *J. Immunol. Methods* **278,** 283–292 (2003).
36. Daigle, N. *et al.* Nuclear pore complexes form immobile networks and have a very low turnover in live mammalian cells. *J. Cell Biol.* **154,** 71–84 (2001).
37. Solovei, I. & Cremer, M. 3D-FISH on cultured cells combined with immunostaining. *Methods Mol. Biol.* **659,** 117–126 (2010).
38. Han, Y., Wind-Rotolo, M., Yang, H. C., Siliciano, J. D. & Siliciano, R. F. Experimental approaches to the study of HIV-1 latency. *Nature Rev. Microbiol.* **5,** 95–106 (2007).
39. Ikeda, T., Shibata, J., Yoshimura, K., Koito, A. & Matsushita, S. Recurrent HIV-1 integration at the BACH2 locus in resting $CD4^+$ T cell populations during effective highly active antiretroviral therapy. *J. Infect. Dis.* **195,** 716–725 (2007).
40. Liu, H. *et al.* Integration of human immunodeficiency virus type 1 in untreated infection occurs preferentially within genes. *J. Virol.* **80,** 7765–7768 (2006).
41. van Helden, J. Regulatory sequence analysis tools. *Nucleic Acids Res.* **31,** 3593–3596 (2003).
42. Wang, Z. *et al.* Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138,** 1019–1031 (2009).



Extended Data Figure 1 | HIV-1 RIGs. **a**, Probability of recurrence of a random set of genes in different lists of HIV-1 integration sites. The histogram shows the distribution of the number of genes present at least twice in the six HIV-1 integration site lists considered; 1×10^9 independent drawings were evaluated. The distribution peaks around 20 genes, with a maximum observed of 50. The number of RIGs detected experimentally in at least two lists was instead 156 ($P < 1 \times 10^{-9}$). **b**, Human chromosome map showing the localization of 156 HIV RIGs. Genes found in four, three and two HIV-1 integration lists are highlighted in red, orange and black, respectively. Hotter genomic regions, favoured for HIV-1 integration as described in ref. 1, are highlighted in blue and indicated by a star.



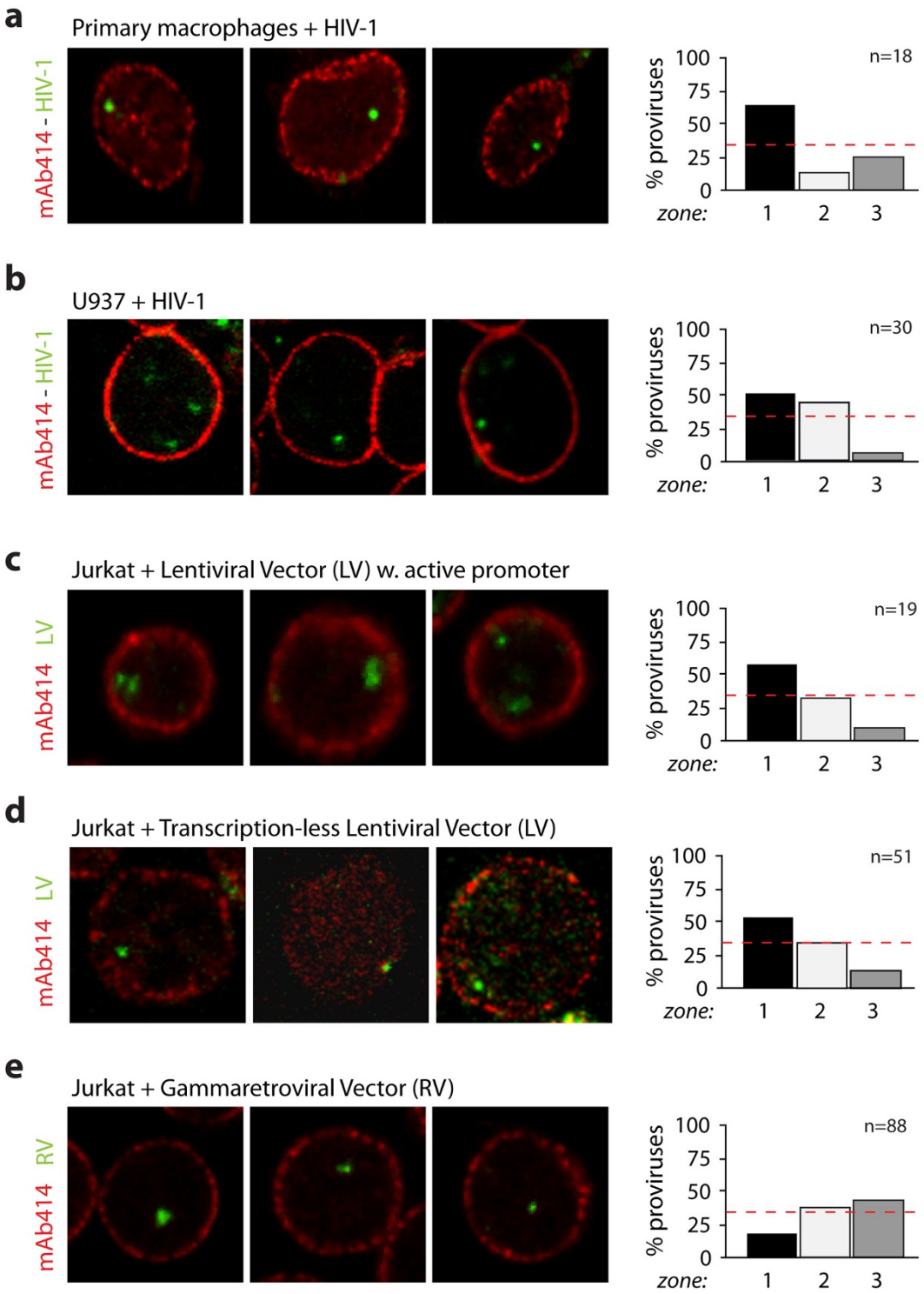
LOCUS	within	1 Mb	5 Mb	10 Mb
6p21.31	RIGs	1	2	5
	other integration sites	1	10	16
	TOT.	2	12	21
9q34.11	RIGs	1	2	5
	other integration sites	5	7	12
	TOT.	6	9	17
11q13.2	RIGs	0	7	10
	other integration sites	0	13	17
	TOT.	0	20	27
16p12.3	RIGs	0	0	1
	other integration sites	0	6	11
	TOT.	0	6	12
16q22.1	RIGs	2	2	2
	other integration sites	3	6	7
	TOT.	5	8	9
17q25.1-3	RIGs	3	5	6
	other integration sites	11	14	16
	TOT.	14	19	22
19p13.2	RIGs	1	3	6
	other integration sites	5	13	29
	TOT.	6	16	35
8q24.3	RIGs	0	0	0
	other integration sites	0	0	0
	TOT.	0	0	0
	TOT.	**33**	**90**	**139**



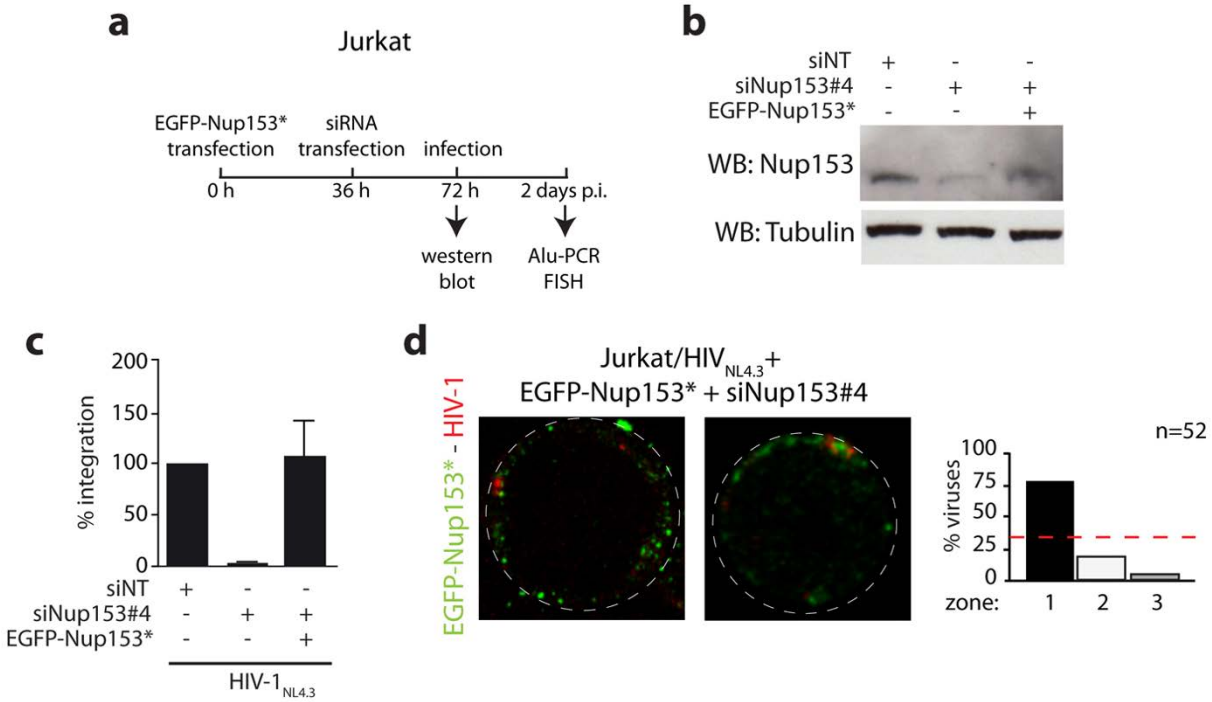
LOCUS	within	1 Mb	5 Mb	10 Mb
4q25	RIGs	0	0	0
	other integration sites	1	4	5
	TOT.	1	4	5
8q24.3	RIGs	1	1	1
	other integration sites	7	9	9
	TOT.	8	10	10
19q13.42	RIGs	0	0	1
	other integration sites	0	5	7
	TOT.	0	5	8
8q24.21	RIGs	0	0	0
	other integration sites	2	7	7
	TOT.	2	7	7
	TOT.	**11**	**26**	**30**

within	1 Mb	5 Mb	10 Mb
total number of RIGs/integrants in the proximity of the BAC probes	**44**	**116**	**169**

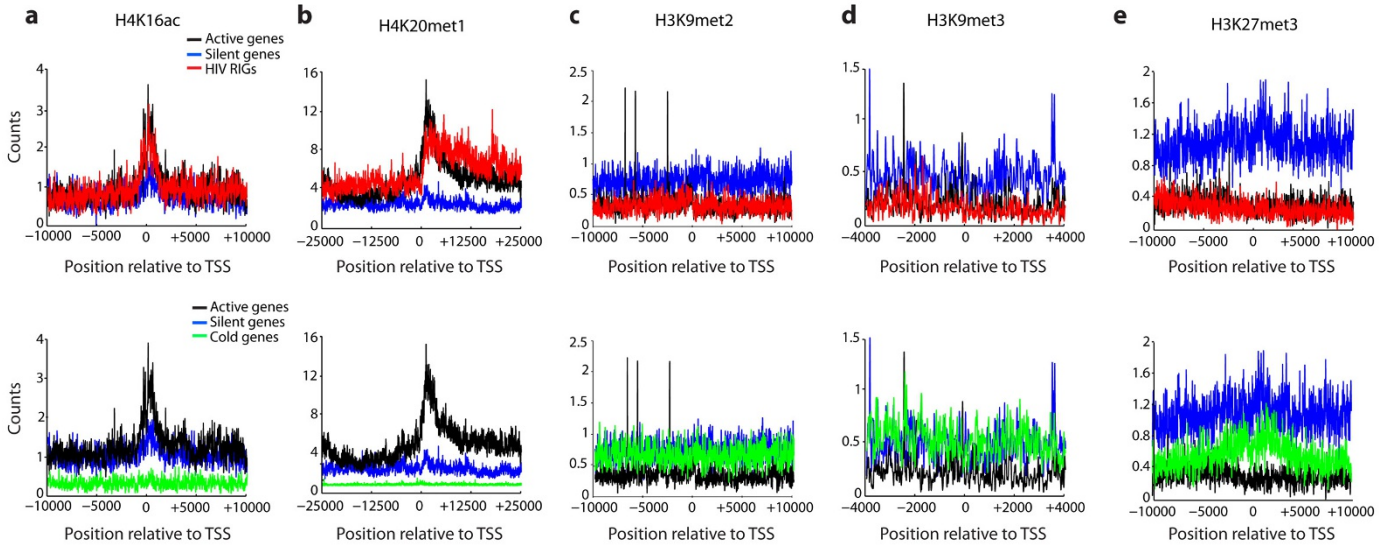
Extended Data Figure 2 | Distribution of RIGs or individual integration sites all over the loci analysed by FISH. The scheme describes the distribution of RIGs (bold) or simple integration sites (regular) around the locus analysed by FISH in Fig. 1; RIGs are in the left panel and hotter zones are in the right panel. As indicated on the side, the total number of RIGs/integrants was calculated within 1, 5 or 10 Mb from the locus analysed by FISH. In total, considering all the RIGs and hotter zones, there are 44 other RIGs/integrants within a window of 1 Mb, 116 within 5 Mb and 169 within 10 Mb around the analysed locus.



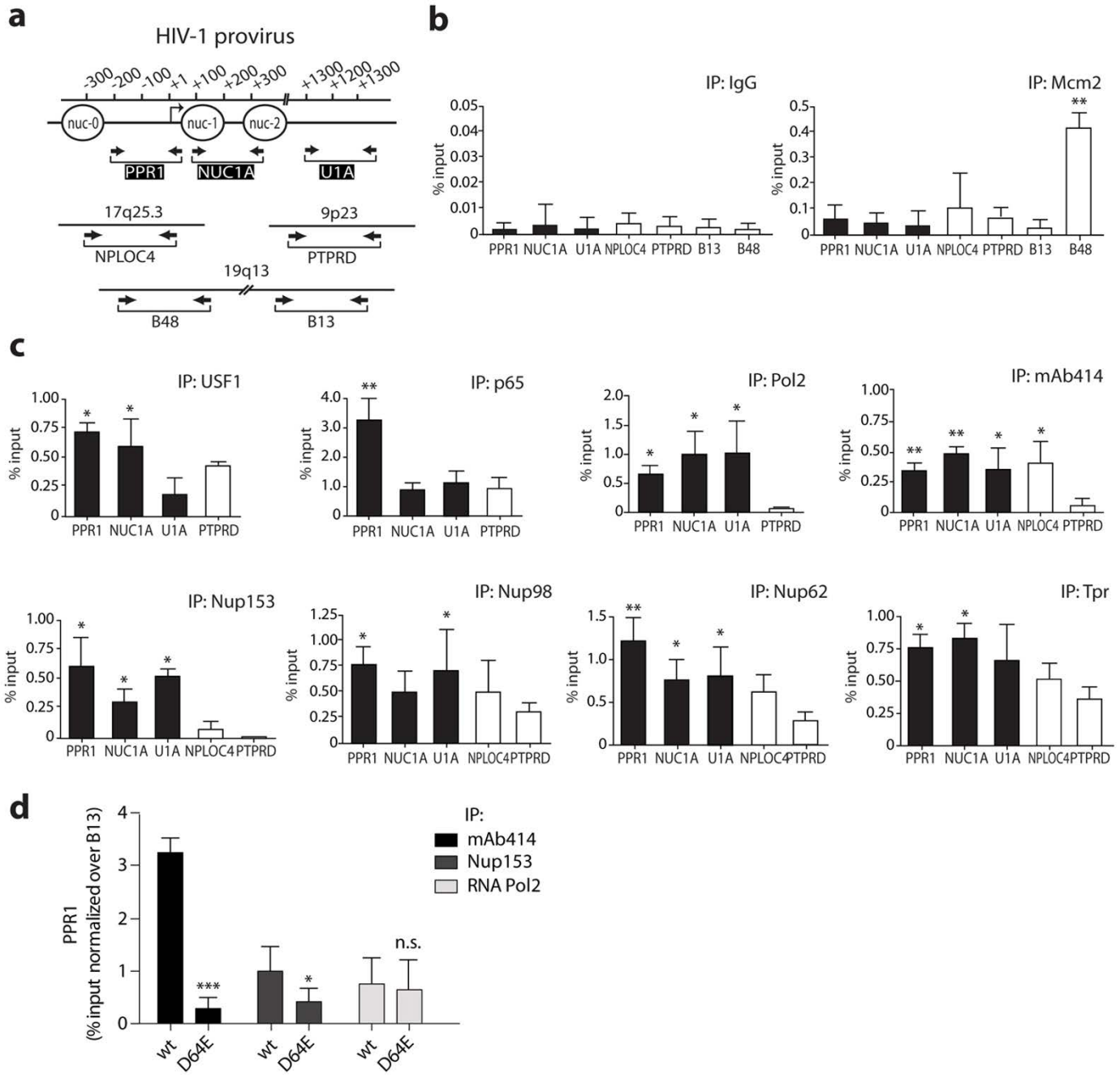
Extended Data Figure 3 | FISH analysis. **a**, Representative images of three-dimensional immuno-DNA FISH of HIV-1 DNA (green) in human primary HIV-1 macrophages stained for mAb414 (red), with relative distribution of FISH signals according to the three concentric zones. **b**, FISH of HIV-1 DNA (green) in the HIV-1 infected U937 monocytic cell line. **c**, Representative images of three-dimensional immuno-DNA FISH of lentiviral vector pLV-THM (green) in Jurkat cells. **d**, Representative images of three-dimensional immuno-DNA FISH of the promoter-less lentiviral vector pCCL–18GFP (green) in Jurkat cells. **e**, Representative images of three-dimensional immuno-DNA FISH of a gammaretroviral vector (green) in Jurkat cells. For all panels, the graphs are organized as described in the main text.



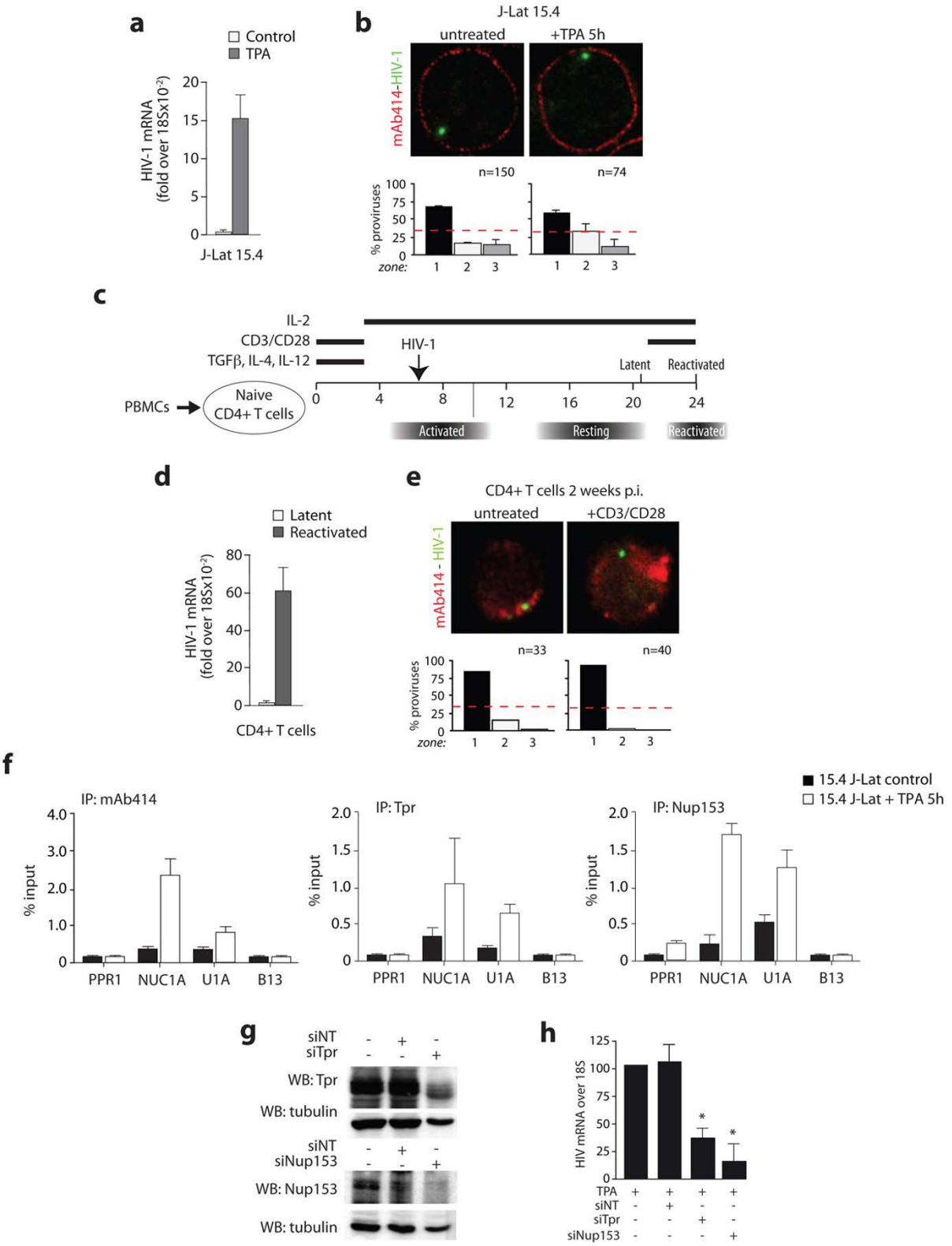
Extended Data Figure 4 | Reconstitution of Nup153 by transfection of an siRNA-resistant plasmid coding for eGFP–Nup153. **a**, Scheme of the experiment performed in Jurkat cells. eGFP–Nup153* contains the coding region for Nup153 tagged with eGFP, but is devoid of the 3′ untranslated region of the mRNA, which is the target of the anti-Nup153 siRNA4. **b**, Western blot showing Nup153 protein level at the moment of infection. siNT, not targeting siRNA. **c**, Real-time Alu-PCR in Jurkat cells 2 days after infection with HIV-$1_{NL4.3}$. Values are mean and s.e.m. of three experiments after normalization over Jurkat transfected with a control, non-targeting siRNA (siNT). **d**, Representative images of three-dimensional immuno-DNA FISH of HIV-1 DNA (red) in Jurkat cells transfected first with the eGFP–Nup153* expression plasmid and then with the siRNA4, targeting endogenous Nup153. The graph on the right side shows the distribution of HIV-1 FISH signals according to the three concentric zones in cells expressing eGFP.



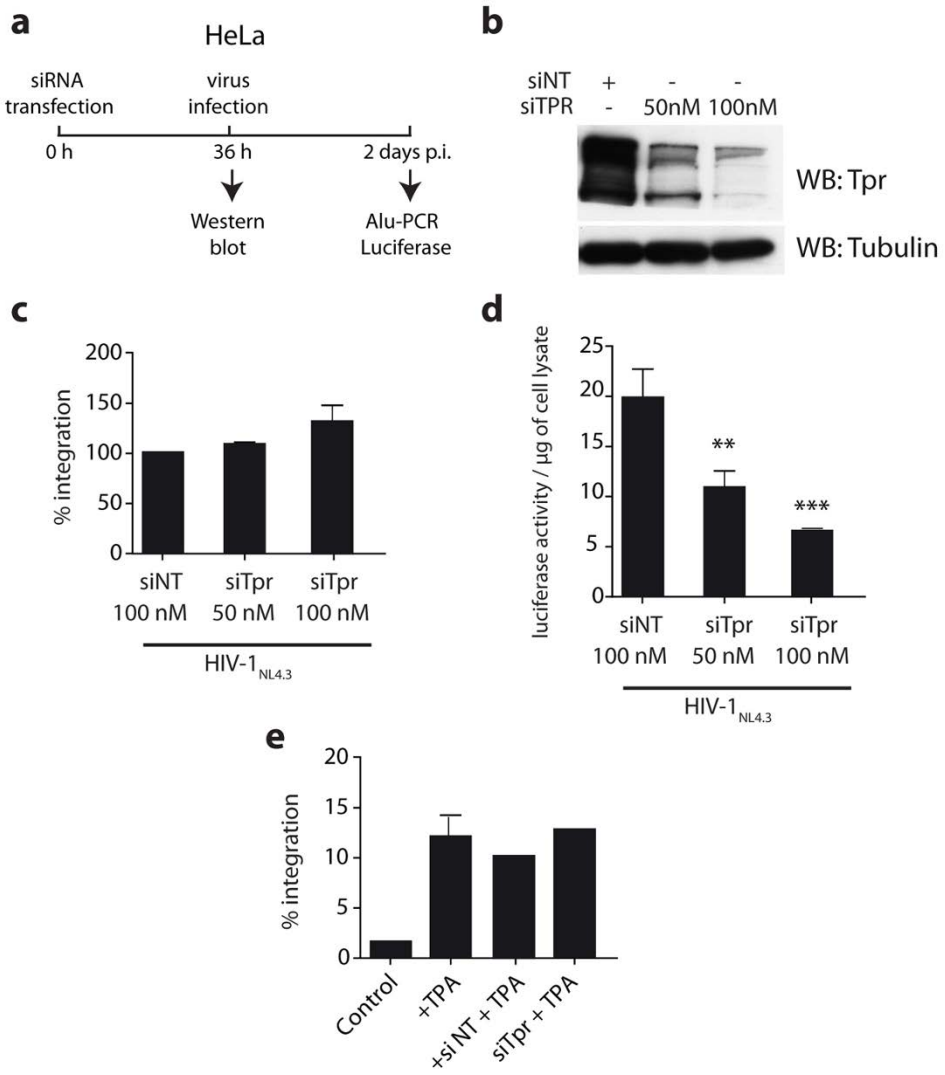
Extended Data Figure 5 | ChIP-seq profiles for HIV RIGs, cold genes and controls. **a**–**e**, Profiles of chromatin modifications around the TSS for HIV RIGs (red) and cold genes (green) compared with highly active (black) and silent (blue) genes in activated $CD4^+$ T cells. Each panel reports results for a specific modification, as indicated.



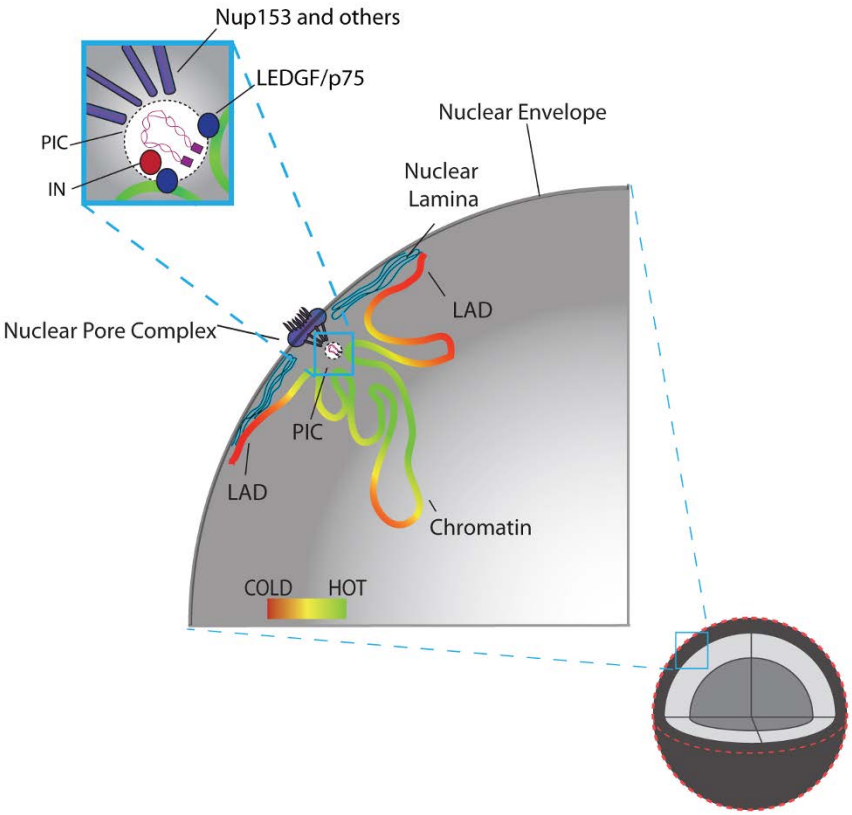
Extended Data Figure 6 | Association of HIV-1 provirus with nucleoporins. **a**, Positions of primers used for ChIP on the HIV-LTR (numbering is according to the TSS and nucleosomes are shown), the NPLOC4 RIG, the PTPRD cold gene, B48 and B13 genomic controls for DNA standardization. B48 maps within the human lamin B2 origin of DNA replication. **b**, Control ChIP data in $CD4^+$ T cells infected with HIV-$1_{NL4\text{-}3/E\text{-}R\text{-}}$, using total immunoglobulin-γ and an antibody against the unrelated Mcm2 cellular protein. For each analysed region, the amount of immunoprecipitated chromatin using the indicated antibodies was normalized according to the input amount of chromatin. Mean and s.e.m. from at least three independent experiments. $^{**}P < 0.01$. **c**, ChIP results in $CD4^+$ T cells, 4 days after HIV-1 infection, using the indicated antibodies. The amount of immunoprecipitated chromatin was normalized according to input. Mean and s.e.m. from at least three independent experiments. $^{**}P < 0.01$,$^{*}P < 0.05$. **d**, ChIP results in $CD4^+$ T cells, 4 days after infection with wild-type HIV-1 or the IN(D64E) mutant virus. For the PPR1 region, corresponding to the viral promoter, the amount of immunoprecipitated chromatin using the indicated antibodies (mAb414, Nup153 and Pol2) was calculated according to the input amount of chromatin, and then normalized over the B13 control genomic region. The graphs show the mean and s.e.m. from three independent experiments. $^{***}P < 0.001$; $^{*}P < 0.05$.



Extended Data Figure 7 | HIV-1 transcriptional activation is concomitant with, and requires, nucleoporins binding to the provirus. **a**, Quantitative reverse transcription PCR measurement of HIV-1 mRNA in mock- or TPA-treated J-Lat 15.4 cells. **b**, Three-dimensional immuno-DNA FISH of HIV-1 DNA (green) in J-Lat 15.4 cells stained for NPC (red) before and after TPA reactivation. **c**, Scheme of the experiment for the generation of a primary, cellular model of HIV-1 latency to study HIV-1 DNA localization in activated and resting primary $CD4^+$ T cells. **d**, Quantitative reverse transcription PCR measurement of HIV-1 mRNA levels in primary infected $CD4^+$ T cells before and after reactivation, normalized over the 18S housekeeping gene. Latent versus reactivated: $P < 0.001$. **e**, Three-dimensional immuno-DNA FISH of HIV-1 DNA (green) in latently infected $CD4^+$ T cells stained for the NPC (red) before and after reactivation, with relative distribution of HIV-1 FISH signals according to the three concentric zones considered in this work. **e**, ChIP in control and TPA-stimulated J-Lat 15.4 cells, with the indicated antibodies. Mean and s.e.m. from at least three independent experiments. **g**, Immunoblot for Tpr (upper panel) and Nup153 (lower panel), 36 h after transfection of the indicated siRNAs (NT, non-targeting control). **h**, Levels of HIV-1 RNA in siRNA-treated J-Lat 15.4 cells after TPA activation. Mean and s.e.m. from three independent experiments. *$P < 0.05$.



Extended Data Figure 8 | Silencing of Tpr in HeLa cells and 15.4 J-Lat clones. **a**, Scheme of the experiment to study HIV-1 integration in infected HeLa cells after Tpr silencing. **b**, Western blot showing Tpr protein level at the moment of infection, after treatment of HeLa cells with a non-targeting siRNA (siNT) or an siRNA targeting Tpr at two different doses. Values are mean and s.e.m. of three experiments after normalization over HeLa cells transfected with a control non-targeting siRNA. **c**, Real-time Alu PCR in HeLa cells infected with HIV-$1_{NL4.3}$ and previously transfected with a non-targeting siRNA (siNT) or an siRNA targeting Tpr at two different doses. Values are mean and s.e.m. of three experiments after normalization over HeLa cells transfected with a control non-targeting siRNA. **d**, Luciferase activity assay in HeLa infected with HIV-$1_{NL4.3}$ and previously transfected with a non-targeting siRNA (siNT) or an siRNA targeting Tpr at two different doses. Values are mean and s.e.m. of three experiments. Statistical significance: $***P < 0.001$; $**P < 0.01$. **e**, Real-time PCR quantification of IL-2 mRNA levels in J-Lat 15.4 cells. The following conditions were tested: untreated cells, plus TPA (4 h), transfection with non-targeting siRNA or an siRNA targeting Tpr for 24 h, followed by treatment with TPA (4 h). Values are mean and s.e.m. of three experiments after normalization over GAPDH. Transcription of interleukin-2 (IL-2) was not significantly altered upon Tpr downregulation.



Extended Data Figure 9 | Model for HIV-1 integration site selection. After entry into the nucleus through the nuclear pore, the viral DNA integrates into the active chromatin closest to the NPC (green zones), avoiding both LADs and the inner part of the nucleus (red zones).

Extended Data Table 1 | List of HIV-1 integration sites considered in this work

List	Source	Nr. Published Sequences	Nr. Unique Intragenic Sites	Reference
Brady et al.	Primary, activated CD4+ T cells, in vitro infection	524	265	(Brady et al., 2009)
Mavilio et al.	Primary, activated CD4+ T cells, in vitro infection	638	329	Unpublished
Schroeder et al.	Sup T1, in vitro infection	642	294	(Schroder et al., 2002)
Liu et al.	PBMCs and tissues from HIV patients	42	32	(Liu et al., 2006)
Ikeda et al.	CD4+ T cells from HIV patients	463	158	(Ikeda et al., 2007)
Han et al.	CD4+ T cells from HIV patients	74	58	(Han et al., 2004)
TOTAL			**1136**	
N. Genes in 4 lists			6	
N. Genes in 3 lists			24	
N. Genes in 2 lists			126	
TOTAL N. RECURRENT GENES			**156**	

Out of the indicated numbers of sequences identified by the six considered studies, 1,136 were within individual genes; of these, 156 recurred in two or more studies.