Nuclear architecture dictates HIV-1 integration site selection

Bruna Marini, Attila Kertesz-Farkas, Hashim Ali, Bojana Lucic, Kamil Lisek, Lara Manganaro, Sandor Pongor, Roberto Luzzati, Alessandra Recchia, Fulvio Mavilio, Mauro Giacca & Marina Lusici

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. 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 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.

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-1NL4-3/E-R-16, 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 integrated15, 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_nrg envelope17 (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))15,16 or for HIV-1NL4-3/E-R- 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

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 area5,10,11 (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 cells12,13 (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 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 occurred in two lists, 24 in three, and six in at least four lists, for a total of 156 genes, which we named HIV recur-
was a marked reduction in HIV-1 integration (Fig. 2k), and the majority of the signals labelled unintegrated viral DNA\(^1\), 72% and 77% of the FISH signals were in zones 2 and 3 for the LEDGF/p75 and Nup153 expressed (active) and 1,000 least expressed (silent) genes from the GNF published observations) and a list of genes corresponding to the 1,000 most representative image, the distribution of the analysed alleles into the three chromosome (BAC) probe labelled with DIG (dUTP-digoxygenin) and FITC (fluorescein isothiocyanate); red: NPC staining by mAb414). Below each.

Figure 1 | Localization of HIV RIGs at the nuclear periphery. a, Subdivision of nucleus into three concentric zones of equal area. b-c, 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 (HZs and HZs (n = 1,400) was identical for RIGs and active genes (Fig. 3c–e and Extended Data Fig. 5).

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\(^2\) (P < 1 × 10\(^{-12}\)), whereas the TAP2 gene from the major histocompatibility complex class II locus is never targeted in CD4\(^{+}\) T cells (P < 1 × 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 CD4\(^{+}\) 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 CD4\(^{+}\) T cells (>90% of alleles; P < 0.001; Fig. 3a).

To understand the chromatin features of RIGs, we compared the available data from chromatin immunopurification sequencing (ChIP-seq) obtained in CD4\(^{+}\) T cells for RIGs\(^3\), 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\(^3\). 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\(^3\) (Fig. 3f).

Heterochromatic LADs contain approximately 4,000 transcriptionally inactive genes\(^3\). 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)\(^23\),\(^24\). 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 USFI 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,
Nup98, Nup62 and Tpr all immunoprecipitated the HIV-1 DNA; binding was also observed for the NPO4 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 active 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 (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 (26,29,30).

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 2 | Integrated, transcriptionally active HIV-1 is found at the nuclear periphery. a, b, Quantification of integrated HIV-1NL4.3 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-1NL4.3 (c) and HIV-1BRU (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-1NL4.3/E-R- 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-1NL4.3 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.
likely to be related to the short life of viral integrase 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.


Supplementary Information is available in the online version of the paper.

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).
Virus production. For the production of viral stocks, we used a plasmid obtained from the Env molecular clone pNL4-3/E8-R, a gift from N. Landau. This viral clone harbours a frasemesh mutation introduced near the 5’ end of the env gene 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-139L as previously described. 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. The D64E mutant plasmid, which is similarly integration-defective, was obtained from the National Institutes of Health (NIH) AIDS Research and Reference Reagent Program.

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, 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 × 10⁵ 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 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-NPC414, 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 anhydride) 1, 2-propanediol) 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) 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’s 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. The three washes 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).

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 FISH signals and the nuclear envelope were measured using LSM 510 Image Examiner Software (Zeiss) and Velocity (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.
CD4 (n = 100), CD28 (n = 52), HEATR6 (n = 44), KDM2B (n = 106), PACS2 (n = 42), JAZF3 (n = 52), TAP2 (n = 54), CNTN4 (n = 56), GPS3 (n = 30), PTPRD (n = 30). Measurements were acquired for the alleles of the following genes in CD43+ cells: JAZF3 (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-1NL4-3/R-E-R; n = 42 HIV-1NL4-3 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(D166N) (n = 66); or cells infected with HIV-1NL4-3 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-1NL4-3/R-E-R; 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 + Lentinial promoter (n = 19); Jurkat + transcriptionless 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% 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 pH 7.4, 10 mM NaCl, 3 mM MgCl2, 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 Lysis Solution with 100 μg/mL 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 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-6199), anti-BACH2 (Abcam, ab-69662) and anti-IKZF3 (Q595) (Abcam, ab-247000), 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 pre-determine sample size.

Five lists of HIV-1 integration sites were collected from published work33,34,40 and an unlisted number of integration sites in CD43+ 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) 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 × 10^6 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 set of genes. Using the non-parametric Mann–Whitney–Wilcoxon test, it was concluded that that HIV RIGs and control genes are more transcripted than a random set of genes (P < 0.005).

LAD coordinates were obtained from ref. 5, whereas genes in 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 × 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 × 10^{-12}; ref. 1 hotter zones, P = 0.0003; HIV RIGs, P = 2.09 × 10^{-10}. Eighty per cent of genes that are never targeted by HIV-1 (cold genes) are significantly enriched inside LADs (P = 3.25 × 10^{-10}).

The profile of aligned LAD border regions (Fig. 3i) was performed as described in ref. 5. A χ² 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).

ChIP-seq profile analyses (Fig. 3f 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 χ² 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.


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; $1 \times 10^9$ 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 ($P < 1 \times 10^{-5}$). 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.
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 siRNA. **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-1NL4.3. 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 siRNA, 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-1NL4-3/E-R-, using total immunoglobulin-c 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.
f, 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-1NL4.3 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-1NL4.3 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

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

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.