Supplemental Data

The Bromodomain and extraterminal (BET) family of proteins targets Moloney Murine Leukemia Virus integration near transcription start sites

Jan De Rijck*[1], Christine de Kogel*[1], Jonas Demeulemeester[1], Sofie Vets[1], Sara El Ashkar[1], Nirav Malani[2], Frederic D Bushman[2], Bart Landuyt[4], Steven J. Husson[4,5], Katrien Busschots[3], Rik Gijsbers[#1] and Zeger Debyser[#1]

[1]Laboratory for Molecular Virology and Gene Therapy, KU Leuven, 3000 Leuven, Belgium.

[2]Department of Microbiology, University of Pennsylvania School of Medicine, Philadelphia 19104, Pennsylvania, USA.

[3]Current address: Hasselt University, Biomedical Research Institute, 3590 Diepenbeek, Belgium.

[4]Functional Genomics and Proteomics, Department of Biology, KU Leuven, 3000 Leuven, Belgium

[5]Systemic Physiological & Ecotoxicological Research (SPHERE), Department of Biology, University of Antwerp, 2000 Antwerp, Belgium.

Figure S1. BET proteins interact with MLV integrase and are important for viral replication, related to Figure 1.

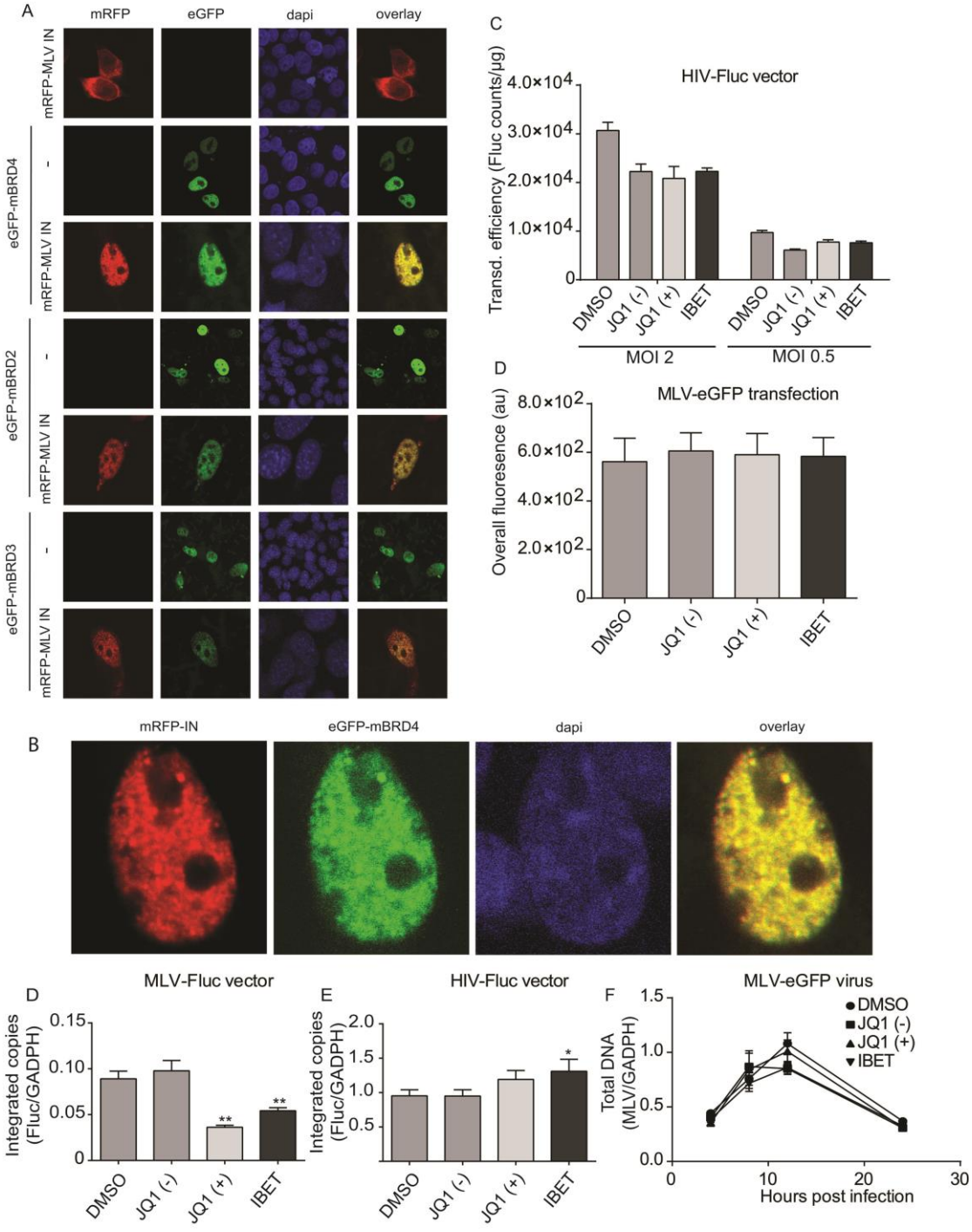


Figure S2. The BET protein ET domain interacts with the MLV-IN C-terminus, related to Figure 2.

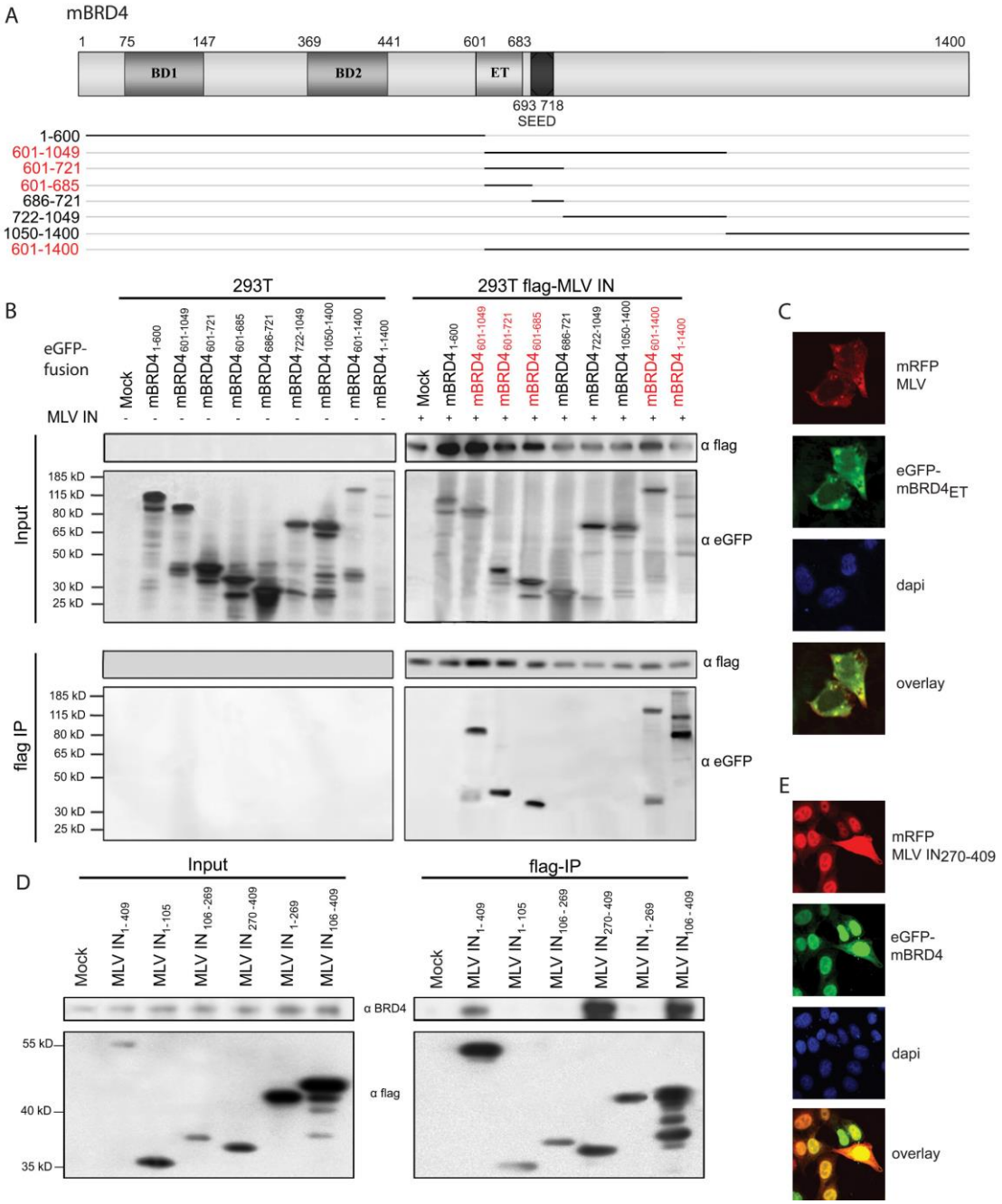


Figure S3. A LEDGF-BRD4 chimeric protein shifts the integration pattern of MLV towards that of HIV, related to Figure 4.

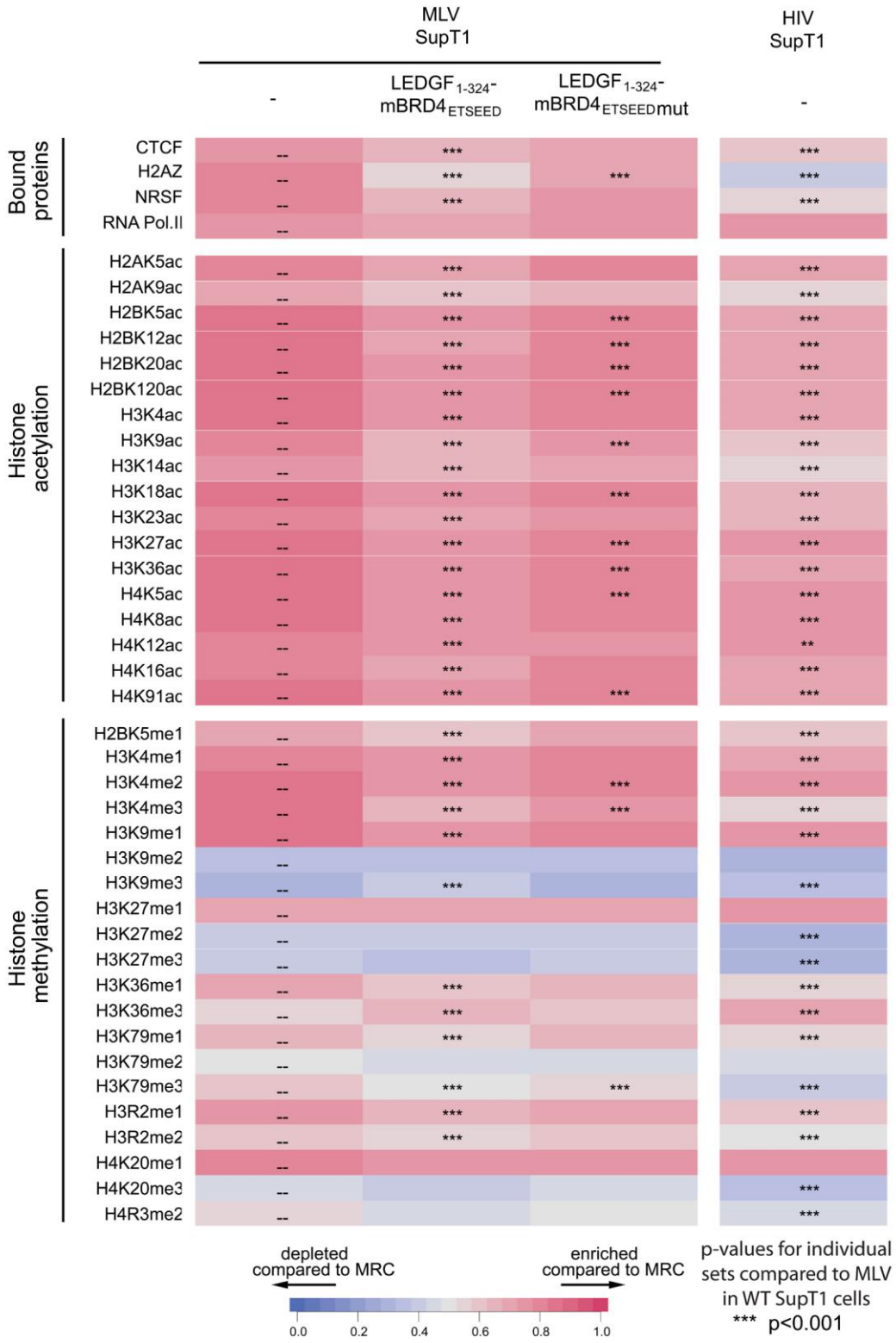
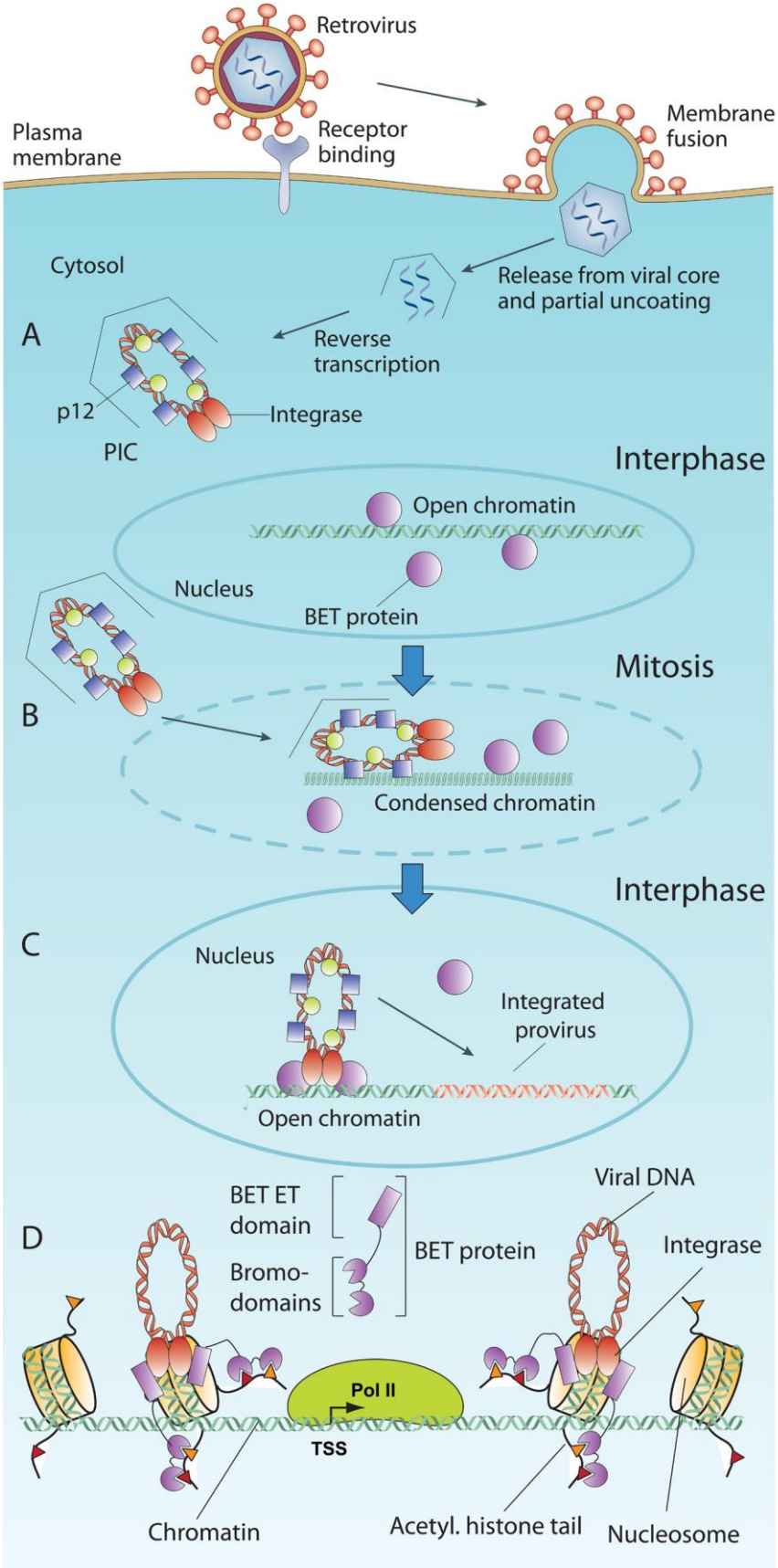


Figure S4. Model for MLV integration targeting.



Supplementary figure legends

Figure S1. BET proteins interact with MLV integrase and are important for viral replication, related to Figure 1. A) Co-localization of eGFP-tagged mBRD4, mBRD3 and mBRD2 with mRFP-tagged MLV IN in NIH3T3 cells. NIH3T3 cells were transfected with plasmids encoding eGFP-tagged proteins and/or mRFP-tagged MLV IN. Cells were fixed 24 h after transfection and analyzed by confocal microscopy. B) Magnification of the co-localization of eGFP-tagged mBRD4 with mRFP-tagged MLV IN in NIH3T3 cells. C) NIH3T3 cells were transduced with HIV-Fluc vector in the presence of 200 nM JQ1(-) or JQ1(+), 500 nM I-BET or an equivalent amount of DMSO as a negative control. Different multiplicities of infection (MOI) are presented. 24 h after transduction, cells were washed. 48 h post transduction, Fluc expression was determined. Data are plotted as average ± standard deviation of triplicate measurements. D) NIH3T3 cells were transfected with the MLV-eGFP molecular clone. Analysis was performed as in panel C. eGFP expression was determined by FACS. E-F) NIH3T3 cells were transduced with MLV-Fluc or HIV-Fluc vectors. Subsequently, cells were expanded and split until 10 days post infection. The number of integrated copies was determined via qPCR using specific primer probe sets and normalized to *GADPH*. Average values and standard deviations of a triplicate measurement are shown. G) NIH3T3 cells were infected with MLV-eGFP virus in the presence of 200 nM JQ1 (-) or JQ1 (+), 500 nM I-BET or an equivalent concentration of DMSO as a control. At 4 h post infection, cells were washed and medium with equivalent amounts of compound or DMSO was added. At 4, 8, 12 and 24 h post infection, gDNA was isolated and total viral DNA production was determined using a MLV-eGFP qPCR with *GADPH* as a housekeeping gene control. In all panels, statistical differences were determined using a student's T-test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure S2. The BET protein ET domain interacts with the MLV-IN C-terminus, related to Figure 2. A) Schematic representation of different mBRD4 deletion mutants analyzed.

Numbers correspond to aa residues. Deletion mutants shown in red interact with MLV IN B) Co-IP of eGFP-labeled mBRD4 truncation mutants with flag-tagged MLV IN in 293T cells. Two days post transfection, cells were lysed, whole cell extracts were prepared and samples were split in two to IP MLV IN or the mBRD4 domains, respectively, and analyzed by Western blot. MLV IN was visualized with a flag antibody and the different mBRD4 domains were visualized with an eGFP antibody. C) Co-localization of eGFP-mBRD$_{ET}$ (aa 601-685) and mRFP-MLV IN in NIH3T3 cells. D) Co-IP of endogenous BRD4 with flag-tagged MLV IN domains. 293T cells were transfected to express the flag-tagged MLV IN domains. Two days post transfection, cells were lysed, whole cell extracts were prepared and flag-tagged MLV IN domains were immunoprecipitated and analyzed by Western blot. MLV IN domains were visualized with flag antibody. A specific BRD4 antibody was used for visualization of endogenous BRD4. E) Co-localization of flag-tagged MLV IN C-terminal domain (aa 270-409) and eGFP-tagged mBRD4$_{ET}$ in NIH3T3 cells. Cells were fixed 24 h after transfection and analyzed by confocal microscopy.

Figure S3. A LEDGF-BRD4 chimeric protein shifts the integration pattern of MLV towards that of HIV, related to Figure 4. Heat map of integration frequency relative to epigenetic features in SupT1 cells. A heat map summarizes the relationships of proviral integration site data sets to epigenetic features. Integration data sets are indicated *above the columns*. Epigenetic features analyzed are shown to the *left* of the corresponding row of the heat map. Tile color indicates whether a chosen feature is favored (*red*, enrichment compared to MRC) or disfavored (*blue*, depletion compared to MRC) for integration for the respective data sets relative to their MRCs, as detailed in the colored ROC area scale at the bottom of the panel. *p* values show significance of departures from MLV integration in WT SupT1 cells and are indicated with *asterisks* (***, $p < 0.001$, Wald statistics referred to χ^2 distribution).

Figure S4. Model for MLV integration targeting. (A) After entry of the virus in the cell, reverse transcription is initiated resulting in the formation of a pre-integration complex (PIC)

composed of the viral cDNA, IN, p12 and several other viral and cellular components. It is unclear to which extent capsid is removed from the PIC during its journey toward the nuclear membrane. (B) Breakdown of the nuclear membrane during mitosis allows the PIC to attach to condensed chromosomes assisted by the viral p12 protein (Elis et al., 2012). (C) After mitosis the PIC becomes trapped in the nucleus and cellular chromatin decondenses. The PIC engages BET proteins to integrate its viral DNA in close proximity to a TSS. It is currently not clear at which step in the replication cycle BET proteins bind MLV IN or at which step of the cell cycle the virus completes the integration step. (D) Detailed model of MLV integration around the TSS. Active TSSs are locally depleted of nucleosomes while the surrounding nucleosomes contain hyperacetylated histone tails (small triangles) bound to bromodomains of BET proteins (Filippakopoulos et al., 2012; Leroy et al., 2012; Wu et al., 2013). Through its interaction with the BET ET-domain the MLV PIC binds nucleosomes and integrates into outward-facing major grooves on nucleosome-wrapped DNA (Roth et al., 2011) on both sides of the TSS, resulting in a bimodal integration pattern around TSSs (Adapted from (Stoye, 2012)).

Supplementary tables

Supplementary Table 1. Determination of the activity and toxicity of the BET inhibitors used in this study, related to Figure 1.

Cell type	Compound	IC_{50} (µM) (95%CI)	CC_{50} (µM) (95% CI)	Selectivity Index
NIH3T3	JQ1 (-)	>50	>50	≤ 1
	JQ1 (+)	0.122 (0.078-0.191)	0.652 (0.387-1.097)	5.3
	I-BET	1.173 (0.791-1.880)	8.038 (4.584-14.090)	6.9

50% inhibitory concentration (IC_{50}) values were determined by transducing NIH3T3 cells with a MLV-based vector encoding luciferase in the presence of a dilution series of compound ranging from 50 to 0.098 µM. After 24 h, the cells were washed to remove the compound and vector. Cells were analyzed 48 h post transduction. 50% cytotoxic concentration (CC_{50}) values were determined by culturing NIH3T3 cells in the presence of a dilution series of compounds ranging from 50 to 0.098 µM in serum-free medium conditions. After 24 h, supernatants were collected and cytotoxicity was determined with an assay measuring LDH release in the supernatant using DMSO as a control. CI ; confidence interval.

Supplementary Table 2 related to Figure 3: Data sets used in this study

Cell type	Data type	GEO or SRA accession	Reference
HEK293T	BRD2 (ChIP-seq)	GSM971946	(Leroy et al., 2012)
	BRD3 (ChIP-seq)	GSM971947	(Leroy et al., 2012)
	BRD4 (ChIP-seq)	GSM971948	(Leroy et al., 2012)
	H3K4me3 (ChIP-seq)	GSM945288	(Thurman et al., 2012)
	DNase I hypersensitivity (DNase-seq)	GSM1008573	(Natarajan et al., 2012; Thurman et al., 2012)
	PHF8 (ChIP-seq)	GSM520382	(Liu et al., 2010)
	RNA Polymerase II (ChIP-seq)	GSM935534	ENCODE Transcription Factor Binding Sites from Stanford/Yale/USC/Harvard
	ELK4 (ChIP-seq)	GSM935590	ENCODE Transcription Factor Binding Sites from Stanford/Yale/USC/Harvard
	KAP1 (ChIP-seq)	GSM935592	ENCODE Transcription Factor Binding Sites from Stanford/Yale/USC/Harvard
	TCF7L2 (ChIP-seq)	GSM782124	ENCODE Transcription Factor Binding Sites from Stanford/Yale/USC/Harvard
GM12878	Nucleosome positions (MNase-seq)	GSM920558	(Kundaje et al., 2012)
K562	Nucleosome positions (MNase-seq)	GSM920557	(Kundaje et al., 2012)
CD4+ T	BRD4 (ChIP-seq)	GSM823378	(Zhang et al., 2012)
	RNA Polymerase II (ChIP-seq)	SRA000206	(Barski et al., 2007)
	Various histone methylations (ChIP-seq)	SRA000206	(Barski et al., 2007)
	Various histone acetylations (ChIP-seq)	SRP000200	(Wang et al., 2008)

Supplementary Table 3. Distance from MLV integration and MRC sites to BRD2-4, Pol II, H3K4me3 and CpG islands, related to Figure 3.

Cell type	Feature	% MLV (MRC) sites in feature	Median distance (bp)	AUC
HEK 293T	BRD2	42.7 (1.9)	149 (35,386)	0.8511
	BRD3	23.5 (0.9)	1,647 (66,843)	0.8126
	BRD4	15 (0.6)	9,816 (126,130)	0.7617
	BRD2–4	44.7 (2.6)	111 (26,146)	0.8404
	Pol II	9.1 (0.4)	13,999 (208,320)	0.7747
	H3K4me3	6.9 (0.2)	1,537 (72,504)	0.8137
	CpG	4.4 (0.2)	5,242 (86,883)	0.7715
CD4+ T	BRD4	18.8 (0.7)	3,489 (95,928)	0.8436
	Pol II	17.0 (0.5)	8,610 (145,000)	0.8266
	CpG	2.3 (0.7)	19,462 (81,029)	0.7262

This Table gives the percentage of MLV integration sites (third column) for and the median distance (fourth column) to regions enriched for the depicted features. The fifth column gives the area under the curves (AUC) calculated from the receiver-operating characteristic (ROC) curves shown in Figure 3 as a measure for each feature to predict MLV integration sites.

Supplementary Table 4. Oligonucleotides used in this study, related to the Materials and Methods.

#	Name	Sequence (5'-3')
1	Fw-HFA MLV IN-SpeI	GCGACTAGTATGGCTTACCCTTAC
2	Rev-HFA MLV IN-AscI	GGCGCGCCTTAGGGGGCCTCCGCG
3	Fw-MLV INs-HindIII	ACAAAGCTTTTATCGAGAACAGCAGGCCCTA
4	Rev-MLV INs-BglII	AAAAGATCTTCAGGGGGCCTCTCTGGTCAG
5	Fw-hBRD4-HindIII	TTCAAGCTTACATGTCTGCGGAGAGCGGC
6	Rev-hBRD4-BamHI	TGGATCCTCAGAAAAGATTTTCTTCAAATATTGAC
7	Fw-mBRD4-HindIII	CTCAAGCTTACATGTCTACGGAGAGCGGCC
8	Rev-mBRD4-BamHI	CTGGGATCCTTAAAAAAGATTTTCTTCAAAT
9	Rev-mBRD4$_{600}$-BamHI	TGGATCCTTAATATGTGGGAGGCGGCTTG
10	Fw-mBRD4$_{601}$-HindIII	CTCAAGCTTACGAATCAGAAGAGGAGGATA
11	Rev-mBRD4$_{685}$-BamHI	TTTTGGATCCTCATTCAGCTTGAGGTTTCCTTTTCTTC
12	Fw-mBRD4$_{686}$-HindIII	CTCAAGCTTACAAAGTTGACGTGATTGCT
13	Rev-mBRD4$_{721}$-BamHI	CGCGGATCCCTACTCTGTTTCAGAGTCTTC
14	Fw-mBRD4$_{722}$-HindIII	CGCAAGCTTACATGGCTCCCAAGTCAAAA
15	Rev-mBRD4$_{1049}$-BamHI	GATGGATCCTTAGGGGGAAGGGTGATGCTGG
16	Fw-mBRD4$_{1050}$-HindIII	CTCAAGCTTACCGGCACCACAAGTCAGACC
17	Fw-hBRD2-HindIII	ATAAAGCTTATGCTGCAAAACGTGACTCC
18	Rev-hBRD2-BamHI	TACGGATCCTTAGCCCGAGTCTGAATCGC
19	Fw-hBRD3-HindIII	CTAAAGCTTCCAGAAATGGGATGCCAAG
20	Rev-hBRD3-BamHI	CGAGGATCCACGTCCATCTGTCCAGCTCTGTC
21	FW-MLV IN$^{s}_{001}$-HindIII	ACAAAGCTTTTATCGAGAACAGCAGGCCCTA
22	Rev-MLV IN$^{s}_{105}$-BglII	GCGAGATCTTTAGCTCTTGCTGGCGTTGA
23	Fw-MLV IN$^{s}_{106}$-HindIII	ATAAAGCTTTTGCCGTGAAGCAGGGCACAAG
24	Rev-MLV IN$^{s}_{269}$-BglII	ATAAGATCTTTAATACAGGATCTCGTAGGGGGT
25	Fw-MLV IN$^{s}_{270}$-HindIII	ATAAAGCTTAAGGCGCCCCTCCCCCCCTGGT
26	Rev-MLV IN$^{s}_{409}$-BglII	ATAAGATCTTCAGGGGGCCTCTCTGGTCAG
27	Fw-mBRD4$_{601}$-AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTTCAGAAGAGGAGGATAAGTG

28	Rev-mBRD4$_{721}$-AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTCTATTCAGCTTGAGGTTTCCT
29	Fw-HsBRD2 ET AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTCTGGAAGTTCTGTTCCAGGGGCCCCGAACTGCAAAACGTGACTCCC
30	Rev-HsBRD2 ET AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTCTAAATGGTGTAGGGCTTCCGG
31	Fw-HsBRD3 ET AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTCTGGAAGTTCTGTTCCAGGGGCCCGAATCAGAGGAAGAGGAGGAGG
32	Rev-HsBRD3 ET AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTCTATGCTGAGAACGGTTTCC
33	Fw1-ET 4mut	AGACCCTGAAGCCATCTACACTACG
34	Fw2-ET 4mut	TGATCAGATTCAGATTAACTTTCAGACCCTGAAGCCATCTACACTACG
35	Rev1-ET 4mut	GGGTTGGAGTTTTTAAGTGATGG
36	Rev2-ET 4mut	GAAAGTTAATCTGAATCTGATCAGGGTTGGAGTTTTTAAGTGATGG
37	Fw-Brd4 601-BamHI AscI	AAAGGATCCAAAAGGCGCGCCGAATCAGAAGAGGAGG
38	Rev-Brd4 721-NheI	GCTAGCTCACTCTGTTTCAGAGTC
39	Fw-Flag-BamHI	AAAGGATCCATGGACTACAAAGACCATG
40	Rev-LEDGF 324-SpeI MluI	GGGACGCGTTTACTAGTAGTTTCCATTTGTTCCTCTTG
41	FLuc Q-PCR forward:	GAAGAGATACGCCCTGGTTCC
42	Fluc Q-PCR reverse	TGTGATTTGTATTCAGCCCATATCG
43	Fluc Q-PCR probe	[6FAM]TTCATAGCTTCTGCCAACCGAACGGACA[TAM]
44	MLV-GFP forward	GGTCCTGCTGGAGTTCGTG
45	MLV-GFP reverse2	GGGTCCCGCTACTAGACACA
46	MLV-GFP Q-PCR probe	[6FAM]TACCAAGCCCTCAACCTCAC[TAM]

Supplementary methods

Compounds

The BET compounds JQ1 (Filippakopoulos et al., 2010) (the active, positive (JQ1 (+)) and inactive, negative (JQ1 (-)) enantiomer) and I-BET (Nicodeme et al., 2010) were kindly provided by J. Bradner (Harvard University, Boston, USA) and dissolved in DMSO.

Plasmids and cloning

All oligonucleotides used are listed in Supplementary Table 4. All enzymes were purchased from Fermentas (Thermo Scientific, St. Leon-Rot, Germany). The expression construct used to identify MLV IN binding proteins (pCG-HFA-MLV IN), encoding HA-FLAG-AU(HFA)-tagged MLV IN (HFA-MLV IN), was kindly provided by J. Skowronski (Cold Spring Harbor Labs, New York, USA). To create stable cell lines, HFA-MLV IN was PCR amplified using primers 1 and 2 (Supplementary Table 4), digested with SpeI and AscI and ligated into digested pCHMWS-intronA-MCS-IRES-HygroR, a lentiviral vector transfer plasmid containing the promoter and intron A of the human cytomegalovirus (CMV) immediate early gene.

For co-localization and co-immunoprecipitation experiments pmRFP-MLV IN^S and p3xflag-MLV IN^S plasmids were constructed. A codon-optimized MLV IN synthetic cDNA (IN^S) was ordered from Addgene (Cambridge MA, USA), amplified using primers 3 and 4, digested with HindIII and BglII and ligated into HindIII/BamHI digested pmRFP-C1 or p3xflag-C1 (Clontech, Mountain View CA, USA). To generate eGFP-BRD4 fusion constructs the human and mouse BRD4 (hBRD4 and mBRD4, respectively) open reading frames were

amplified from HeLaP4 cDNA using primers 5 and 6 (hBRD4) or from NIH3T3 cell cDNA using primers 7 and 8 (mBRD4). Amplicons were digested using HindIII and BamHI and ligated into HindIII/BamHI digested peGFP-C1 (Clontech, Mountain View CA, USA). mBRD2 and mBRD3 cDNA was amplified likewise using primers 17-18 and 19-20 respectively, digested with BamHI and HindIII and ligated into pCHMWS-eGFP-MCS-IRES-HygroR to generate eGFP-fused expression constructs. Mouse BRD4 truncations were amplified from the peGFP-C1 mBRD4 construct using primers 7-16. All amplicons were digested with HindIII and BamHI and ligated into HindIII/BamHI digested peGFP-C1. MLV IN^S deletion mutants were amplified by PCR from p3xflag-MLV IN^S using primers 21-26, digested with HindIII and BglII and ligated into HindIII/BamHI digested pmRFP-C1 or p3xflag-C1. The plasmid encoding MLV IN for recombinant protein production in *e. Coli* cells was kindly provided by C. Johnson (Picathaway, NJ, USA). The plasmids for recombinant protein production of human immunodeficiency virus (HIV) IN and Rous sarcoma virus (RSV) IN were described before (Busschots et al., 2005; Maertens et al., 2003). The plasmid for recombinant protein production of prototype foamy virus (PFV) IN was kindly provided by P. Cherepanov, London, UK. Plasmids for recombinant protein production of the BET proteins were cloned using the Gateway system (Invitrogen, Merelbeke, Belgium). The ET-domains from mBRD2, mBRD3 and mBRD4 were amplified using primers 27-32 and recombined into pDONR221 (Invitrogen). Subsequently, constructs with a Glutathione-S-Transferase (GST) fusion or a Maltose-Binding Protein (MBP) fusion were generated by recombination from the pDONR plasmid into pGGWA or pMGWA (Busso et al., 2005), respectively. The ET domain single,

double, triple and quadruple mutants (E652Q, E652-654Q, E652-654Q-D656N and E652-654Q D656N-E658Q, respectively) were generated by SLIM mutagenesis (Chiu et al., 2004) on pDONR221-mBRD4$_{ET}$ plasmid using primers 33-36.

To create a lentiviral vector transfer plasmid expressing the LEDGF$_{1-324}$mBRD4 chimeric construct, first the cDNA of mBRD4$_{ETSEED}$ or its quadruple pointmutant (mBRD4$_{ETSEEDmut,}$ containing E652Q-E654Q-D656N-E658Q) was PCR amplified using primers 37 and 38, digested with BamHI and NheI and ligated into BamHI/NheI digested pCH_EF1a_MCS_IRES_BsdR transfer plasmid, creating pCH_EF1a_MCS_mBRD4$_{ETSEED}$_IRES_BsdR and pCH_EF1a_MCS_mBRD4$_{ETSEEDmut}$_IRES_BsdR, respectively. Subsequently, 3xflag-LEDGF$_{1-324}$ was amplified from pCHMWS_3xflag-LEDGF/p75_IRES_HygroR (Bartholomeeusen et al., 2009) using primers 39 and 40. The resulting PCR product was digested (BamHI/MluI) and ligated into BamHI/AscI digested pCH_EF1a_MCS_mBRD4$_{ETSEED}$_IRES_BsdR and pCH_EF1a_MCS_mBRD4$_{ETSEEDmut}$_IRES_BsdR, resulting in pCH_EF1a_3xflag_LEDGF$_{1-324}$mBRD4$_{ETSEED}$_IRES_BsdR and pCH_EF1a_3xflag_LEDGF$_{1-324}$mBRD4$_{ETSEEDmut}$_IRES_BsdR. All cloning steps and plasmid constructs were verified by DNA sequencing.

Cell culture and stable cell lines

NIH3T3 (ATCC CRL-1658), 293T (ATCC CRL-11268) and HeLa P4 (a kind gift from Pierre Charneau, Institut Pasteur, Paris, France) cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) (Gibco-BRL, Merelbeke, Belgium) supplemented with 7% heat inactivated fetal calf serum (Sigma-

Aldrich, Bornem, Belgium) and gentamicin (50 µg/ml, Gibco-BRL). SupT1 cells (ATCC CRL-1942) were cultured in Roswell Park Memorial Institutes medium (RPMI-1640, Gibco-BRL) supplemented with 10% heat inactivated fetal calf serum and gentamicin (50 µg/ml, Gibco-BRL). Stable cell lines overexpressing HFA-MLV IN or $LEDGF_{1-324}mBRD4_{ETSEED}$ chimera were generated using lentiviral vectors (see below) encoding a hygromycin (HygroR) or a blasticidin (BsdR) resistance gene, respectively. Cells were selected in medium containing 50 mg/ml hygromycin or 9 µg/ml blasticidin (Invivogen, Toulouse, France). After selection, blasticidin concentration was reduced to 3 µg/ml. All cells were grown in a humidified atmosphere with 5% CO_2 at 37°C.

Co-Immunoprecipitation

Six million 293T cells were plated in 8.5 cm dishes and transfected using linear polyethylenimine (PEI) with the indicated expression constructs. After 48 h, cells were washed and lysed in co-immunoprecipitation (co-IP) whole cell lysis buffer (20 mM Tris/HCl pH 7.5, 250 mM NaCl, 0.1% Triton-X100, 10% glycerol, 1 mM phenylmethanesulfonylfluoride (PMSF), and protease inhibitor cocktail (Complete: EDTA-free, 1 tablet/50 ml buffer, Roche Diagnostics, Mannheim, Germany). Lysates were cleared by centrifugation at 20,000 g for 10 min and incubated overnight at 4°C with tag-specific antibodies. Subsequently, protein-G agarose (Roche Diagnostics) was added and samples were incubated at 4°C for 3 h. Samples were washed 5 times with co-IP lysis buffer, before Western blot analysis. The same protocol was used to prepare cytoplasmic and nuclear extracts. In short, cells were lysed in

cytoplasmic extraction buffer (25 mM Tris/HCl pH 7.4, 150 mM NaCl, 5 mM dithiothreitol (DTT), protease inhibitor cocktail (Roche Diagnostics). Next, the cytoplasmic extract was cleared by centrifugation and the remaining nuclei were washed and lysed in nuclear lysis buffer (25 mM Tris/HCl pH 7.4, 300 mM NaCl, 5 mM DTT, protease inhibitor cocktail (Roche Diagnostics). To identify interaction partners of MLV IN, extracts from 12 cell culture dishes were pooled in a single precipitation experiment using flag-M2 affinity gel (Sigma-Aldrich, Erembodegem, Belgium). After immunoprecipitation, bound proteins were eluted in 1xTris-buffered saline (TBS) buffer supplemented with 250 µg/ml flag-peptide (Sigma-Aldrich) for 30 min. The eluted proteins were incubated with washed Protein-G agarose (Roche Diagnostics) for 20 min to remove excess of immunoglobulins. Cleared supernatants, containing MLV IN and binding partners, were precipitated overnight with 10 % (v/v) trichloroacetic acid. Dried protein pellets were stored at -20°C for subsequent protein identification.

Sample Preparation, NanoLC - Q-TOF MS/MS and Data Analysis

Lyophilized samples, corresponding to approximately 200 µg protein material were dissolved in 45 µl digestion buffer (100 mM ammonium bicarbonate, 5 mM DTT, 10% Acetonitrile, pH 8.0) containing 0.1% RapiGest (Waters, Etten-Leur, the Netherlands) and incubated for 10 min at 95°C. Next, 5 µg trypsin (dissolved in 5 µl digestion buffer containing 0.1% RapiGest) was added and incubated overnight at 37°C. To stop the reaction, 5 µl of 0.5 M HCl was added to the samples that were incubated for another hour at 37°C. Samples were filtered through a 22 µm spin-down filter (Ultrafree-MC, Millipore),

lyophilized in a Speedvac Concentrator, aliquoted in two duplicates and stored at -20°C until analysis.

Nanoscale liquid chromatography-tandem mass spectrometry (nanoLC-MS/MS) experiments were conducted using an Ultimate3000 LC instrument (LC Packings, Thermo Scientific, St Leon-Rot, Germany) and a microTOF-Q mass spectrometer (Bruker Daltonic GmbH, Bremen, Germany) The nanoscale LC system was directly coupled to a quadrupole time-of-flight mass spectrometer (Q-TOF). Samples were dissolved in 10 µl of MQ water containing 5% CH_3CN and 0.1% formic acid (FA) and sonicated briefly. Five µl of the samples was loaded on a µ-guard pre-column (MGU-30 C18, LC-Packings) at a flow rate of 30 µl/min to trap the peptides and to get rid of the salts when rinsing this column for 5 min with 5% CH_3CN and 0.1% FA. This pre-column was then placed in line with the analytical column (Symmetry C18, 3.5µm, 75µm x 100mm, Waters). Peptides were separated using a 90 min gradient from 5 to 90% CH_3CN containing 0.1% FA at a flow rate of 200 nl/min.

The column eluent was directed through a stainless steel needle (Proxeon, Thermo Scientific) to the electrospray source of the Q-TOF mass spectrometer that was set to automatic data-dependent MS to MS/MS switching. The resulting data were automatically processed using the DataAnalysis software package (Bruker Daltonics) to generate mgf peak list files that were submitted to an in-house Mascot (Matrixscience) server using the human protein database from NCBI to match the fragmentation data from the mgf peak list files with the FASTA format protein database. To identify the

proteins, only significant identifications, as annotated by the Mascot scoring method, were taken into account and manually refined.

Western blot analysis and antibodies

Western blot analysis was performed as described previously (Vandekerckhove et al., 2006). Briefly, cellular extracts were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Flag-MLV IN^S was detected using a purified IgG1 mouse anti-flag M2 monoclonal antibody (F3165, Sigma Aldrich). eGFP-tagged proteins were detected using a polyclonal goat-anti eGFP antibody (Abcam, Cambridge, UK Ab6673). Endogenous h/mBRD4 was detected using a rabbit anti-hBRD4 antibody (A301-985A; Bethyl Laboratories, ImTec Diagnostics, Antwerp, Belgium). Equal loading was verified with β-tubulin (T-4026; Sigma-Aldrich). Visualization was performed by chemiluminescence (ECL+; Amersham Biosciences, Uppsala, Sweden).

Confocal microscopy

Thirty thousand NIH3T3 cells were seeded in Lab-Tek™ microscopy chamber slides (Thermo Scientific) and transfected with 0.5 µg of a plasmid encoding eGFP-BRD2/3/4 and/or mRFP-MLV IN^S using Lipofectamin2000 reagent (Invitrogen). After 24 h, cells were fixed using 4% formalin, stained with 4',6-diamidino-2-phenylindole (DAPI), and mounted. Samples were analyzed using a LSM 510 META imaging unit (Carl Zeiss, Zaventem, Belgium) as described before (Cherepanov et al., 2003).

Retroviral vector production

Both lenti- and gammaretroviral vectors were produced as described before (Geraerts et al., 2006). Briefly, lentiviral vectors were produced by a PEI-based triple transfection of 293T cells with the envelope plasmid pVSV-G, the packaging construct p8.91 and one of the described transfer plasmids. MLV-based gammaretroviral vectors were produced similarly using triple transfection of the pVSV-G envelop, the pRVgagpol packaging plasmid and pLNC-eGFP-T2A-fLuc as transfer plasmid.

Virus production

A near complete eGFP-labeled MLV-virus, a molecular clone containing eGFP in the envelop (Sliva et al., 2004), was produced by transfection in 293T cells and subsequent passaging the supernatant to freshly seeded NIH3T3 cells to nullify potential plasmid contamination. Supernatant was harvested and concentrated using an Amicon Ultra-15™ filter unit (50 kDa cut-off) (Merck Millipore, Overijse, Belgium). To study viral replication, $2*10^6$ NIH3T3 cells were infected with 50 reverse transcriptase units (RTU) of MLV-eGFP virus as determined by SYBRGreen-I product-enhanced reverse transcriptase assay (SG-PERT (Pizzato et al., 2009)). Supernatants were sampled daily and viral replication was monitored using the SG-PERT assay.

IC_{50} and CC_{50} determination of BET inhibitors

50% inhibitory concentrations (IC_{50}) were determined by transducing $2*10^4$ NIH3T3 cells with a MLV-based vector encoding luciferase in the presence of a 1:2 dilution series of compound ranging from 50 to 0.098 µM. After 24 h, the

cells were washed to remove the compound. 48 h post transduction, cells were lysed and transduction efficiency was determined using a standard luciferase assay kit (ONE-Glo™, Promega, Leiden, The Netherlands) as described by the manufacturer. Fifty percent cytotoxic concentrations (CC_{50}) were determined by a lactate dehydrogenase (LDH) cytotoxicity assay kit (Cytotoxicity detection kit, Roche Applied Science, Vilvoorde, Belgium) as described by the manufacturer. Twenty-four hours after seeding, cells were incubated with a compound dilution series ranging from 50 to 0.098 µM. After 24 h, LDH release in the supernatant was measured and cell toxicity was determined compared to the negative (DMSO) control and a positive control (Triton X-100 lysed cells).

Luciferase assay

To measure luciferase activity, cells were lysed in luciferase lysis buffer (50 mM Tris/HCl pH 7.5, 200 mM NaCl, 0.2% NP-40, 10% glycerol) and luciferase activity was measured following the manufacturers protocol using ONE-Glo™ reagents (Promega, Leiden, The Netherlands). Luciferase activity was normalized for total protein determined with the bicinchoninic acid assay (BCA) (Pierce Protein Biology products, Erembodegem, Belgium).

gDNA isolation and Quantitative PCR

Two million cells were pelleted and genomic DNA was extracted using a mammalian genomic DNA miniprep kit (Sigma-Aldrich). Genomic DNA concentrations were determined using standard spectrophotometric methods. Samples corresponding to 700 ng genomic DNA were used for analysis. Each

reaction contained 12.5 µl iQ Supermix (Biorad, Nazareth, Belgium), 40 nM forward and reverse primer and 40 nM of probe in a final volume of 25 µl. The following primer/probe sets were used: Fluc; primers 41-43 (Supplementary Table 5), MLV-eGFP: primers 44-46. In all cases, *GADPH* was used as a house-keeping gene control (Applied Biosystems, rodent *GAPDH* quantification kit, VIC labeled probe). Samples were measured in triplicate for 3 min at 95°C followed by 50 cycles of 10 s at 95°C and 30 s at 55°C in a LightCycler® 480 (Roche-applied-science, Vilvoorde, Belgium). Analysis was performed using the LightCycler® 480 software supplied by the manufacturer.

Protein purification

E. coli BL21 CodonPlus chemically competent cells were transformed with prokaryotic expression constructs. Cultures were grown at 37°C to A_{600} = 0.5 and induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) at 30°C for 3 h. Cell pellets were lysed in a lysis buffer containing 50 mM Tris/HCl pH 7.3, 250 mM NaCl, 1 mM PMSF, 5 mM DTT, 10 IU recombinant DNAse/10 ml lysate, and sonicated. The lysate was cleared by centrifugation at 15,000 g for 30 min. Protein was purified from the lysate on a column containing an appropriate affinity resin for the specific tag. Glutathione Sepharose (GE life sciences, Diegem, Belgium) and Amylose Resin (New England Biolabs, Leiden, the Netherlands) were used for GST and MBP purifications, respectively. Ni^{2+}-NTA resin (Invitrogen) was used for His_6-tagged proteins. The column was washed in wash buffer containing 50 mM Tris/HCl pH 7.3, 250 mM NaCl and 5 mM DTT. Subsequently, proteins were eluted in wash buffer with addition of 50 mM reduced glutathione for GST purifications, 20

mM maltose for MBP-purifications or 250 mM imidazole for Ni^{2+} purifications, respectively. Eluted proteins were dialyzed overnight with wash buffer containing 10% glycerol and stored at -80°C.

Protein-protein interaction assay

AlphaScreen measurements were performed in a total volume of 25 µl in 384-well Optiwell microtiter plates (PerkinElmer, Zaventem, Belgium). All components were diluted to their desired concentrations in assay buffer (25 mM Tris/HCl pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$, 0.1% Tween-20 and 0.1% BSA and 1 mM DTT). For the K_d determinations, GST, GST-mBRD4$_{ET}$ or GST-mBRD4$_{ETSEED}$ were titrated against a background of 80 nM MLV IN-His$_{6.}$ This amount provided minimal binding curve perturbation while maintaining a good signal-to-noise ratio. The affinity of IN from several retroviral species was determined likewise against a fixed concentration of MBP-mBRD4$_{ET}$ (2 nM). After addition of the proteins, the plate was incubated for 1 h at 4°C and 20 µg/ml glutathione or anti-MBP donor and Ni^{2+}-chelate acceptor beads (PerkinElmer) were added, bringing the final volume to 25 µl. After 1 h incubation at room temperature, protected from light, the plate was read on an EnVision Multilabel Reader in AlphaScreen mode (PerkinElmer). Results were analyzed with Prism5.0 (GraphPad software) after non-linear regression with the appropriate equations: one-site specific binding, taking ligand depletion into account for the K_d measurements and dose-response with variable slope for the IC_{50} determination.

Recovery of integration sites and analysis of integration site distributions

Recovery of integration sites was performed essentially as previously described (Ciuffi et al., 2009; Wang et al., 2007). Briefly, linkers were ligated to restriction enzyme-digested (MseI) genomic DNA isolated from tranduced cells and virus-host DNA junctions were amplified by nested PCR. Different samples were individually barcoded with the second pair of PCR primers to generate 454 libraries. PCR products were purified by binding to beads and sequenced using 454/Roche pyrosequencing (titanium technology). Reads were quality-filtered by requiring perfect matches to the long terminal repeat (LTR) linker, barcode, and flanking LTR and subsequently mapped to the human/mouse genome. All sites were required to align to the reference genome within 3 bp of the LTR edge, with the great majority showing no gap. In order to control for possible biases in the datasets due to the choice of the MseI restriction endonuclease in cloning integration sites, matched random control (MRC) sites were generated *in silico*. To do so, each experimental integration site was paired with three sites in the genome, located at the same distance from a randomly selected MseI site in the genome. Association to genomic features and histone modifications were performed as described previously (Berry et al., 2006; Brady et al., 2009).

Integration frequencies surrounding RefSeq TSSs, CpG islands and DNAseI hypersensitive sites

For each integration site, the distance in kb to the respective genomic feature was calculated, where the midpoint of the CpG island, DHS or the TSS was set at 0 kb.

Integration sites left of a CpG island or TSS were given negative kb values, while integration sites right of a CpG island or TSS were calculated as positive. Subsequently, the integration site data where pooled in bins ranging from 0-1 kb, 1-2 kb, etc distance from this midpoint. The percentage of integration sites occurring at a certain distance from the CpG island midpoint was plotted vs the distance.

Supplementary references

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Bartholomeeusen, K., Christ, F., Hendrix, J., Rain, J.-C., Emiliani, S., Benarous, R., Debyser, Z., Gijsbers, R., and De Rijck, J. (2009). Lens epithelium-derived growth factor/p75 interacts with the transposase-derived DDE domain of PogZ. J. Biol. Chem. *284*, 11467–11477.

Berry, C., Hannenhalli, S., Leipzig, J., and Bushman, F.D. (2006). Selection of target sites for mobile DNA integration in the human genome. PLoS Comput. Biol. *2*, e157.

Brady, T., Lee, Y.N., Ronen, K., Malani, N., Berry, C.C., Bieniasz, P.D., and Bushman, F.D. (2009). Integration target site selection by a resurrected human endogenous retrovirus. Genes Dev. *23*, 633–642.

Busschots, K., Vercammen, J., Emiliani, S., Benarous, R., Engelborghs, Y., Christ, F., and Debyser, Z. (2005). The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. J. Biol. Chem. *280*, 17841–17847.

Busso, D., Delagoutte-Busso, B., and Moras, D. (2005). Construction of a set Gateway-based destination vectors for high-throughput cloning and expression screening in Escherichia coli. Anal. Biochem. *343*, 313–321.

Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E., and Debyser, Z. (2003). HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. J. Biol. Chem. *278*, 372–381.

Chiu, J., March, P.E., Lee, R., and Tillett, D. (2004). Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. Nucleic Acids Res. *32*, e174.

Ciuffi, A., Ronen, K., Brady, T., Malani, N., Wang, G., Berry, C.C., and Bushman, F.D. (2009). Methods for integration site distribution analyses in animal cell genomes. Methods *47*, 261–268.

Elis, E., Ehrlich, M., Prizan-Ravid, A., Laham-Karam, N., and Bacharach, E. (2012). p12 tethers the murine leukemia virus pre-integration complex to mitotic chromosomes. PLoS Pathog. *8*, e1003103.

Filippakopoulos, P., Picaud, S., Mangos, M., Keates, T., Lambert, J.-P., Barsyte-Lovejoy, D., Felletar, I., Volkmer, R., Müller, S., Pawson, T., et al. (2012). Histone recognition and large-scale structural analysis of the human bromodomain family. Cell *149*, 214–231.

Geraerts, M., Eggermont, K., Hernandez-Acosta, P., Garcia-Verdugo, J.-M., Baekelandt, V., and Debyser, Z. (2006). Lentiviral vectors mediate efficient and stable gene transfer in adult neural stem cells in vivo. Hum. Gene. Ther. *17*, 635–650.

Kundaje, A., Kyriazopoulou-Panagiotopoulou, S., Libbrecht, M., Smith, C.L., Raha, D., Winters, E.E., Johnson, S.M., Snyder, M., Batzoglou, S., and Sidow, A. (2012). Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements. Genome Res. *22*, 1735–1747.

Leroy, G., Chepelev, I., Dimaggio, P.A., Blanco, M.A., Zee, B.M., Zhao, K., and Garcia, B.A. (2012). Proteogenomic characterization and mapping of nucleosomes decoded by Brd and HP1 proteins. Genome Biol. *13*, R68.

Liu, W., Tanasa, B., Tyurina, O. V, Zhou, T.Y., Gassmann, R., Liu, W.T., Ohgi, K.A., Benner, C., Garcia-Bassets, I., Aggarwal, A.K., et al. (2010). PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. Nature *466*, 508–512.

Maertens, G., Cherepanov, P., Pluymers, W., Busschots, K., De Clercq, E., Debyser, Z., and Engelborghs, Y. (2003). LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. J. Biol. Chem. *278*, 33528–33539.

Natarajan, A., Yardimci, G.G., Sheffield, N.C., Crawford, G.E., and Ohler, U. (2012). Predicting cell-type-specific gene expression from regions of open chromatin. Genome Res. *22*, 1711–1722.

Pizzato, M., Erlwein, O., Bonsall, D., Kaye, S., Muir, D., and McClure, M.O. (2009). A one-step SYBR Green I-based product-enhanced reverse transcriptase assay for the quantitation of retroviruses in cell culture supernatants. J. Virol. Methods *156*, 1–7.

Roth, S.L., Malani, N., and Bushman, F.D. (2011). Gammaretroviral integration into nucleosomal target DNA in vivo. J. Virol. *85*, 7393–7401.

Sliva, K., Erlwein, O., Bittner, A., and Schnierle, B.S. (2004). Murine leukemia virus (MLV) replication monitored with fluorescent proteins. Virol. J. *1*, 14.

Stoye, J.P. (2012). Studies of endogenous retroviruses reveal a continuing evolutionary saga. Nat. Rev. Microbiol. *10*, 395–406.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Vandekerckhove, L., Christ, F., Van Maele, B., De Rijck, J., Gijsbers, R., Van den Haute, C., Witvrouw, M., and Debyser, Z. (2006). Transient and stable knockdown of the integrase cofactor LEDGF/p75 reveals its role in the replication cycle of human immunodeficiency virus. J. Virol. *80*, 1886–1896.

Wang, G.P., Ciuffi, A., Leipzig, J., Berry, C.C., and Bushman, F.D. (2007). HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. Genome Res. *17*, 1186–1194.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Peng, W., Zhang, M.Q., et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. *40*, 897–903.

Wu, S.-Y., Lee, A.-Y., Lai, H.-T., Zhang, H., and Chiang, C.-M. (2013). Phospho switch triggers Brd4 chromatin binding and activator recruitment for gene-specific targeting. Mol. Cell *49*, 843–857.

Zhang, W., Prakash, C., Sum, C., Gong, Y., Li, Y., Kwok, J.J.T., Thiessen, N., Pettersson, S., Jones, S.J.M., Knapp, S., et al. (2012). Bromodomain-containing protein 4 (BRD4) regulates RNA polymerase II serine 2 phosphorylation in human CD4+ T cells. J. Biol. Chem. *287*, 43137–43155.