Cell PRESS
Open ACCESS

Cell Reports
Report

The BET Family of Proteins Targets Moloney Murine Leukemia Virus Integration near Transcription Start Sites

Jan De Rijck,[1,5] Christine de Kogel,[1,5] Jonas Demeulemeester,[1] Sofie Vets,[1] Sara El Ashkar,[1] Nirav Malani,[2] Frederic D. Bushman,[2] Bart Landuyt,[3] Steven J. Husson,[3,4] Katrien Busschots,[1,7] Rik Gijsbers,[1,6] and Zeger Debyser[1,6,*]

[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, PA, USA
[3]Functional Genomics and Proteomics, Department of Biology, KU Leuven, 3000 Leuven, Belgium
[4]Systemic Physiological & Ecotoxicological Research (SPHERE), Department of Biology, University of Antwerp, 2000 Antwerp, Belgium
[5]These authors contributed equally to this work and are co-first authors
[6]These authors contributed equally to this work and are co-last authors
[7]Present address: Hasselt University, Biomedical Research Institute, 3590 Diepenbeek, Belgium
*Correspondence: zeger.debyser@med.kuleuven.be
http://dx.doi.org/10.1016/j.celrep.2013.09.040

SUMMARY

A hallmark of retroviral replication is integration of the viral genome into host cell DNA. This characteristic makes retrovirus-based vectors attractive delivery vehicles for gene therapy. However, adverse events in gene therapeutic trials, caused by activation of proto-oncogenes due to murine leukemia virus (MLV)-derived vector integration, hamper their application. Here, we show that bromodomain and extraterminal (BET) proteins (BRD2, BRD3, and BRD4) and MLV integrase specifically interact and colocalize within the nucleus of the cell. Inhibition of the BET proteins' chromatin interaction via specific bromodomain inhibitors blocks MLV virus replication at the integration step. MLV integration site distribution parallels the chromatin binding profile of BET proteins, and expression of an artificial fusion protein of the BET integrase binding domain with the chromatin interaction domain of the lentiviral targeting factor LEDGF/p75 retargets MLV integration away from transcription start sites and into the body of actively transcribed genes, conforming to the HIV integration pattern. Together, these data validate BET proteins as MLV integration targeting factors.

INTRODUCTION

Integration of a DNA copy of the retroviral RNA genome into the host chromatin is a pivotal step in retroviral replication and links the fate of the invading virus with that of the infected cell. This characteristic makes retrovirus-based vectors suitable to deliver therapeutic genes into cells to correct genetic diseases. Murine leukemia virus (MLV)-derived vectors have been used successfully to correct primary immunodeficiency disorders like X-linked severe combined immunodeficiency (X-SCID) (Cavazzana-Calvo et al., 2000; Gaspar et al., 2004; Hacein-Bey-Abina et al., 2002). However, their use led to adverse events in a subset of patients due to long-terminal-repeat-driven activation of proto-oncogenes (i.e., insertional mutagenesis) resulting in uncontrolled clonal cell proliferation and leukemia (Deichmann et al., 2007). Retroviral integration site distribution is not random. Whereas the host protein lens epithelium-derived growth factor (LEDGF/p75) targets lentiviral (e.g., HIV) integration toward the body of active transcription units (Ciuffi et al., 2005), gammaretroviral (e.g., MLV) integration is independent of LEDGF/p75 and preferentially occurs near transcription start sites (TSSs), CpG islands, and DNase I-hypersensitive sites (DHSs) (Cattoglio et al., 2010; Felice et al., 2009; Mitchell et al., 2004; Wu et al., 2003). In addition, retroviral integration is favored on the outward-facing major groove of nucleosome-wrapped DNA (Roth et al., 2011; Wang et al., 2007, 2009). It is generally accepted that cellular proteins, cofactors, dictate target site selection. In this study, we identified the cellular determinants that target MLV integration. In earlier work with hybrid HIV viruses, it was shown that transferring the MLV integrase (IN) coding region into HIV caused the chimeras to integrate with a specificity close to that of MLV, revealing IN as the principal viral determinant of integration specificity (Lewinski et al., 2006). Therefore, we screened for cellular MLV IN interaction partners that could act as a MLV-specific tether. Although an earlier study picked up BRD2 as a MLV IN-interacting protein (Studamire and Goff, 2008), we show here that members of the bromodomain and extraterminal domain containing (BET) family of proteins (BRD2, BRD3, and BRD4) interact with MLV IN and orchestrate gammaretroviral integration, in agreement with a recent report by Sharma et al. (2013) and that engineered BET proteins can retarget MLV replication.

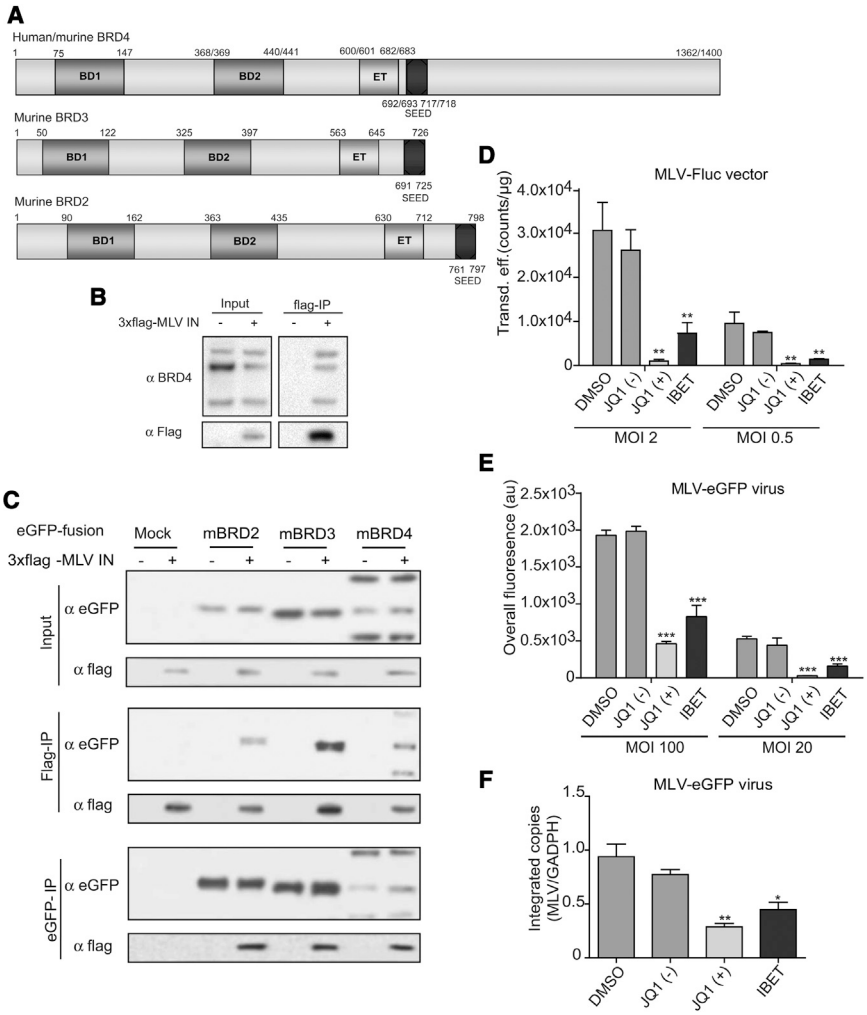


Figure 1. BET Proteins Interact with MLV IN and Are Important for Viral Integration

(A) Schematic representation of human/murine BET proteins. Numbers correspond to aa positions. BD, bromodomain; ET, extraterminal domain; SEED, serine, glutamic acid, aspartic acid-rich domain.

(B) Coimmunoprecipitation of endogenous BRD4 from 293T cells expressing flag-tagged MLV IN analyzed by western blot.

(C) 293T cells were cotransfected with flag-tagged MLV IN and eGFP-mBRD2, eGFP-mBRD3, or eGFP-mBRD4 expression constructs. MLV IN (Flag-IP) or mBRD proteins (eGFP-IP) were precipitated and analyzed by western blot.

(D and E) NIH 3T3 cells were transduced with MLV-derived vectors (MLV-Fluc) (D) or infected with an eGFP expressing viral clone (MLV-eGFP virus) (E) in the presence of 200 nM JQ1(−) or JQ1(+), 500 nM I-BET, or an equivalent amount of DMSO as a negative control. Two multiplicities of infection (MOI) are presented. Twenty-four hours after transduction/infection, cells were washed, and, in case of the viral clone, raltegravir (1 μM) was added to prevent multiple-round replication. Twenty-four hours later, transduction or infection efficiency was determined. Data are plotted as average ±SD of triplicate measurements.

(F) NIH 3T3 cells were infected with MLV-eGFP as in (E). Subsequently, cells were expanded and split until 10 days after infection. The number of integrated copies was determined via qPCR and normalized to GADPH. Average values and SDs of a triplicate measurement are shown.

In all panels, differences were determined using a Student's *t* test. *p < 0.05, **p < 0.01, ***p < 0.001.

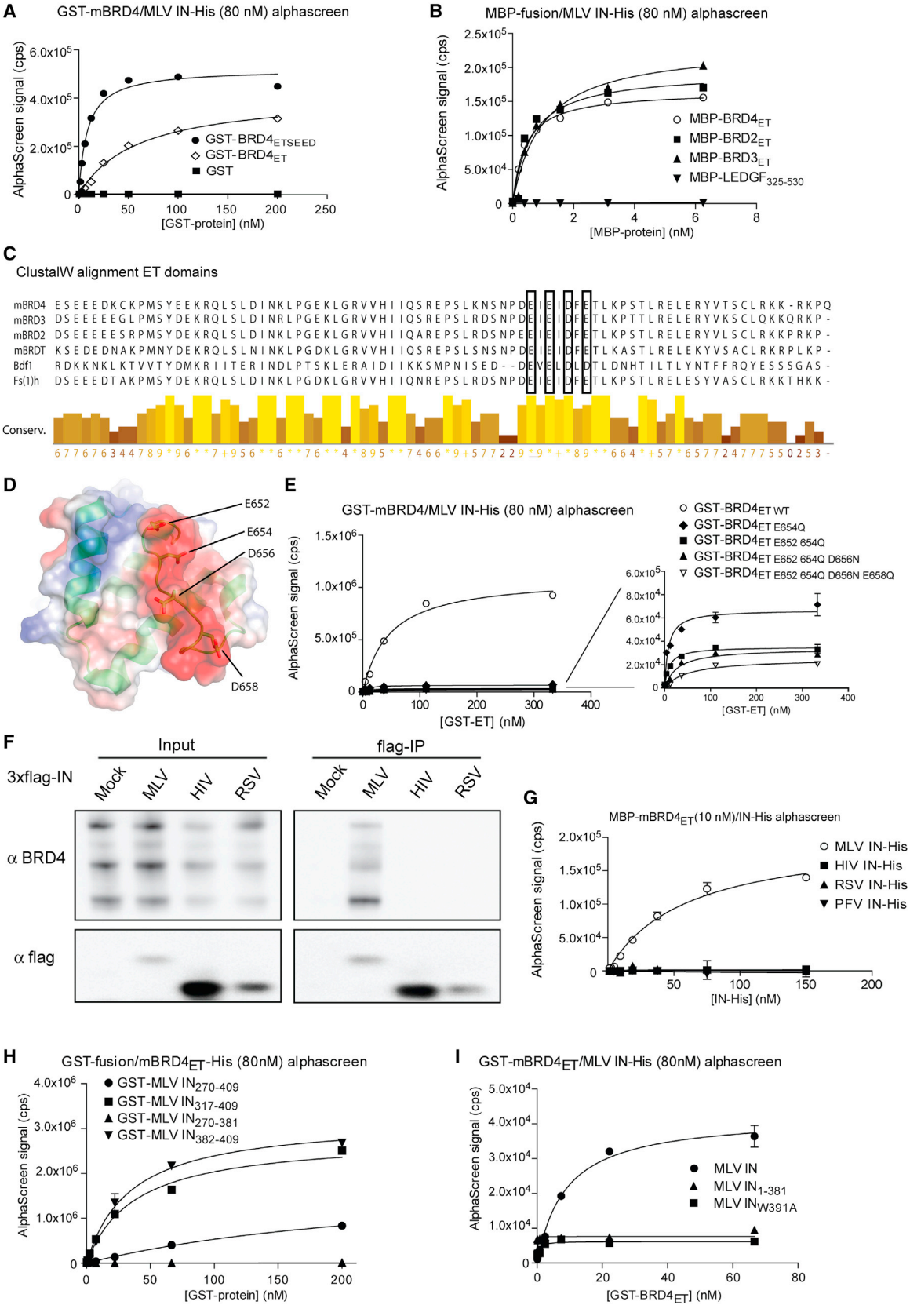
RESULTS

BET Proteins Bind MLV Integrase

We singled out MLV integrase (IN)-interacting proteins from 293T cell extracts expressing triple flag-tagged MLV IN via coimmunoprecipitation (coIP) of IN using flag-affinity matrix. Eluted proteins were identified by mass spectrometry (MS). Wild-type 293T cells were analyzed in parallel as control. Bromodomain containing protein 4 (BRD4) was represented with the largest set of peptides (data not shown), but we also identified BRD3. Of note, an earlier study picked up BRD2 as an MLV IN-interacting protein (Studamire and Goff, 2008). All these proteins are members of the bromodomain and extraterminal (BET) protein family (Figure 1A) and share two chromatin-interacting bromodomains, recognizing acetylated histone tails, and a protein-interacting extraterminal (ET) domain (for a review, see Devaiah and Singer, 2013). To confirm the interaction of endogenous BRD4 with MLV IN, transiently expressed flag-tagged MLV IN was immunoprecipitated from 293T nuclear extracts. BRD4 was readily detected in the pull-down fraction by western blot (Figure 1B). In addition, 3xflag MLV IN interacted with both human and mouse BRD4 (hBRD4 and mBRD4) (coIP, data not shown). A similar coIP experiment in cells transiently expressing eGFP-tagged BET proteins and flag-tagged MLV IN showed that, next to eGFP-mBRD4, MLV IN interacts with eGFP-tagged mBRD2 and mBRD3 as well (Figure 1C). When expressed alone, eGFP-tagged BET proteins located in the nucleus of NIH 3T3 cells, whereas MLV IN fused to the monomeric red fluorescent protein (mRFP-MLV IN) predominantly located to the cytoplasm with only trace amounts in the nucleus (Figure S1A). However, coexpression of mBRD2, -3, or -4 with mRFP-MLV IN relocated MLV IN to the nucleus of the cell, colocalizing with the respective BET proteins (Figures S1A and S1B). Similar data were obtained in HeLa cells (data not shown).

Bromodomain Inhibitors Reduce MLV Replication

To evaluate the role of BET proteins in MLV replication, we exploited the recently identified BET protein bromodomain inhibitors JQ1(+) and I-BET (Filippakopoulos et al., 2010; Nicodeme et al., 2010). JQ1(−) (the inactive R-enantiomer) and DMSO served as negative controls. Based on 50% cytotoxicity (CC_{50}) and inhibitory concentrations (IC_{50}) (Table S1), we used JQ1(+) and I-BET concentrations of 200 and 500 nM, respectively. NIH 3T3 cells were either transduced with a retroviral vector



(legend on next page)

encoding firefly luciferase (MLV-Fluc vector) (Figure 1D) or infected with a viral clone expressing eGFP (MLV-eGFP virus) (Figure 1E). Both vector transduction and virus infection were inhibited 5- to 10-fold compared to JQ1(−) or DMSO control, respectively, whereas transduction with an HIV-derived vector (HIV-Fluc) was not inhibited (Figure S1C). Similar results were obtained in HeLa cells using the MLV-Fluc vector (data not shown). Reporter gene expression following transfection of the MLV-eGFP molecular clone was not affected by the presence of JQ1(+) or I-BET excluding transcriptional effects (Figure S1D). To determine the step in the viral replication cycle where bromodomain inhibitors inhibit MLV replication, viral DNA intermediates were measured via quantitative PCR (qPCR). Quantification of the integrated proviral copies at 10 days post-transduction/infection in the presence of BET inhibitors revealed that the integrated MLV (virus or vector) copies were reduced 2- to 3-fold, whereas HIV-Fluc integration was not inhibited (Figures 1F, S1E, and S1F, for MLV-eGFP virus, MLV-Fluc vector, and HIV-Fluc, respectively). Because the amount of total DNA at early time points after infection, which is a measure of reverse transcription, was not reduced in the presence of BET inhibitors (Figure S1G), we conclude that BET proteins act at a step between reverse transcription and integration.

The MLV IN C Terminus Binds the BET Extraterminal Domain

Pull-down experiments using a panel of eGFP-tagged mBRD4 truncation mutants pinpointed the ET domain ($BRD4_{ET}$, aa 601–685) as the minimal IN binding domain (Figures S2A and S2B). Confocal microscopy experiments corroborated that $BRD4_{ET}$ is the minimal domain required for colocalization with mRFP-MLV IN (Figure S2C; data not shown). Recombinant glutathione-S-transferase (GST)-tagged $BRD4_{ET}$ and $BRD4_{ETSEED}$ (aa 601–721) and His_6-tagged MLV IN were shown to interact in an AlphaScreen protein-protein interaction assay (apparent K_d of 58.70 ± 8.05 nM and 8.56 ± 1.55 nM respectively), confirming a direct interaction (Figure 2A). As expected, recombinant $BRD2_{ET}$ and $BRD3_{ET}$ interacted with MLV IN as well, with K_d values in the same range as $BRD4_{ET}$ (Figure 2B). Considering the conservation in the ET domain among BET proteins (Figure 2C) and the $BRD4_{ET}$ NMR structure (Lin et al., 2008), we introduced E652Q, E654Q, D656N, and E658Q mutations in the BRD4 ET domain (Figure 2D). Maximal loss of binding was obtained with the quadruple mutant (referred to as $BRD4_{ETmut}$ in further experiments) (Figure 2E). The observed interaction with BET proteins was specific for MLV IN, because IN proteins from other retroviral families did not interact with BET proteins as revealed by coIP using lysates of 293T cells transiently expressing flag-tagged MLV, HIV, or RSV (Rous Sarcoma Virus) IN (Figure 2F). Similar data were obtained in an AlphaScreen assay measuring the interaction of maltose-binding protein (MBP)-tagged $BRD4_{ET}$ with His-tagged MLV, HIV, RSV, or PFV (Prototype Foamy Virus) IN (Figure 2G). To pinpoint the MLV IN-BET interacting domain, different flag-tagged MLV IN deletion constructs were generated, transiently expressed in 293T cells, and tested by coIP (Figure S2D). Only MLV IN fragments containing the C-terminal domain (MLV $IN_{270-409}$) pulled down endogenous BRD4 from 293T cell lysates. Specific interaction in cells was corroborated by colocalization of the mRFP-MLV $IN_{270-409}$ and eGFP-tagged mBRD4 using confocal microscopy (Figure S2E). AlphaScreen analysis using smaller truncation mutants of the MLV IN C-terminal domain revealed that the last 27 residues of MLV IN (aa 382–409) were sufficient to interact with $BRD4_{ET}$ (Figure 2H). Finally, alanine scanning of the latter domain revealed W391 to be critically important for the interaction (Figure 2I; data not shown). Similar results were obtained for $BRD2_{ET}$ and $BRD3_{ET}$ (data not shown). Taking into account that BET proteins are known to associate with promoter regions through their bromodomains (Leroy et al., 2012) and interact with MLV IN through their ET domain in a gammaretrovirus-specific manner, we considered BET proteins as good candidate MLV targeting factors.

MLV Vectors Integrate near BET Protein Hot Spots

To correlate MLV integration with BET-chromatin binding sites, we determined 10,514 unique MLV vector integration sites in 293T cells and computationally generated matched random control (MRC) sites. Integration sites were compared with the BRD2, -3, and -4 chromatin binding profile (Table S2) (Leroy et al., 2012). *HOX* gene clusters are enriched for BRD2–4-bound nucleosomes (Leroy et al., 2012). Indeed, we detected 11, 10, and 23 MLV integration sites in the *HOXA*, *HOXB* (data not shown), and *HOXC* (Figure 3A) clusters, respectively, versus 1,

Figure 2. The BET Protein ET Domain Interacts with the MLV IN C Terminus

(A) Direct interaction of His_6-tagged MLV IN (80 nM) with an increasing amount of GST-tagged $mBRD4_{ET}$, GST-tagged $mBRD4_{ETSEED}$ or GST alone as measured by AlphaScreen. Apparent K_ds of 8.56 ± 1.50 nM (ETSEED) and 58.70 ± 8.05 nM (ET) were determined using a nonlinear regression curve fit for specific binding.

(B) Comparable affinity of different BET proteins for MLV IN. His_6-tagged MLV IN (40 nM) was titrated against the indicated MBP-fused ET domains. Apparent K_ds of 0.60 ± 0.17, 0.97 ± 0.18, and 0.41 ± 0.03 nM were determined for MBP-BRD2, -3, and -4, respectively. MBP-$LEDGF_{325-530}$ was used as a negative control.

(C) Sequence alignment (ClustalW) of the ET domain of mBRD4 (Q9ESU6, aa 601–683), mBRD3 (Q8K2F0, aa 563–645), mBRD2 (Q7JJ13, aa 630–712), mBRDT (Q91Y44, aa 496–578), *Saccharomyces* cerevisiae (Sc) Bromodomain factor (Bdf1) (P35817, aa 518–598), and *Drosophila melanogaster* (Dm) Female Sterile Homeotic (Fs[1]h) (P13709, aa 942–1024).

(D) NMR structure of the ET domain of BRD4 (Lin et al., 2008).

(E) Interaction of His_6-tagged MLV IN (80 nM) with GST-tagged $mBRD4_{601-685}$ mutants as measured by AlphaScreen.

(F) Coimmunoprecipitation of endogenous BRD4 with flag-tagged MLV, HIV, or RSV IN analyzed by western blot.

(G) Interaction of 10 nM MBP-tagged $mBRD4_{ET}$ with His_6-tagged MLV, HIV, RSV, or PFV IN measured by AlphaScreen. The K_d for MLV IN (56.7 ± 23.08 nM) was determined using a nonlinear regression curve fit for specific binding. The other INs showed no binding to $BRD4_{601-721}$.

(H) Interaction of 80 nM His-tagged $mBRD4_{ET}$ with increasing amounts of the GST-tagged MLV IN C-terminal domain (aa 270–409) and truncation mutants thereof as measured by AlphaScreen.

(I) Interaction of 80 nM His-tagged $mBRD4_{ET}$ with GST-tagged MLV IN, the indicated C-terminal MLV-IN deletion mutant (aa 1–381) or MLV IN with a W391A point mutation as measured by AlphaScreen. All AlphaScreen experiments were performed three times.

Representative experiments are shown. Error bars indicate the SDs of triplicate data points.

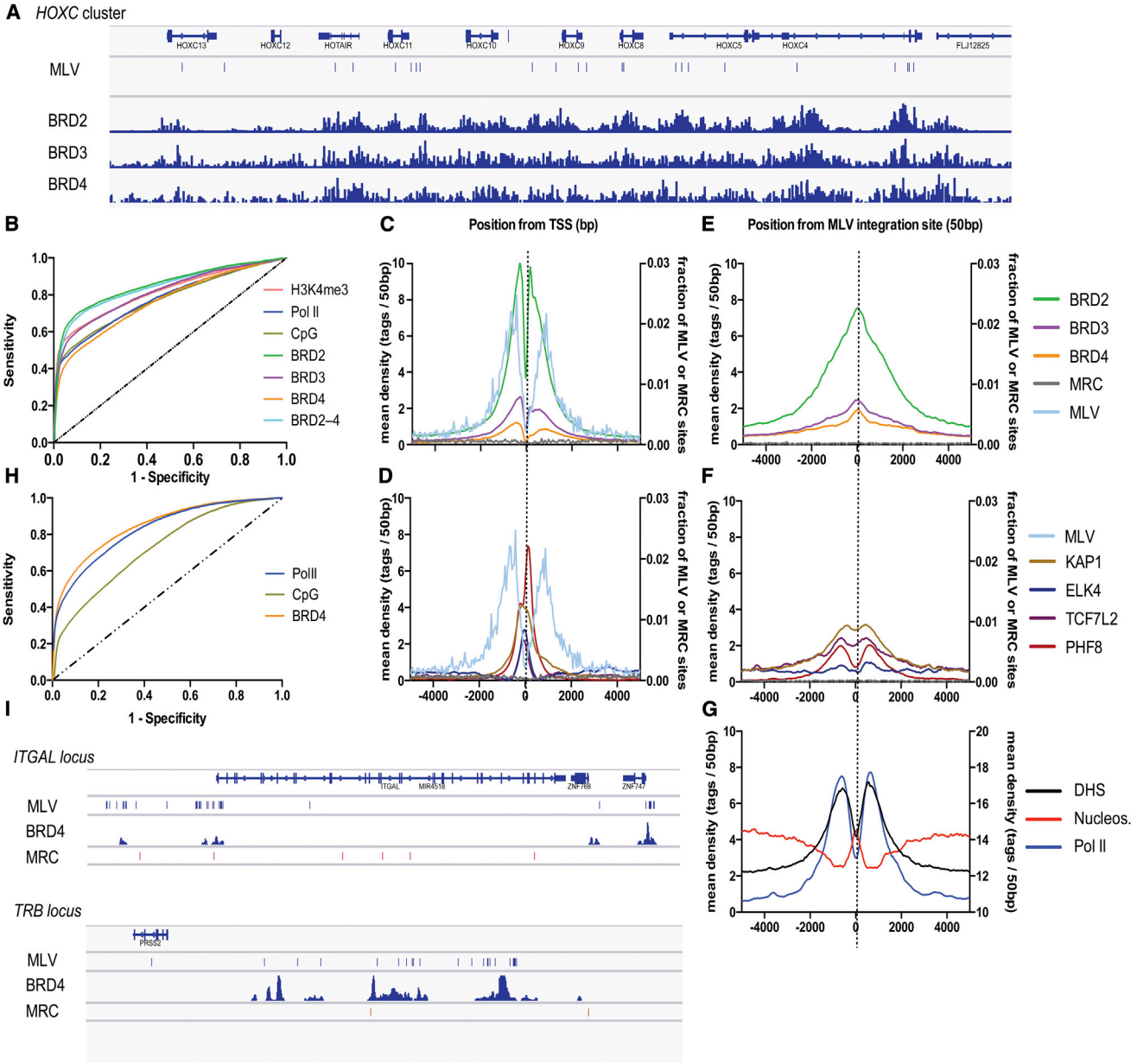


Figure 3. MLV Vectors Integrate in BET Protein Hot Spots

(A) Schematic representation of the BET chromatin binding profile (Leroy et al., 2012) and MLV integration sites (MLV) in the *HOXC* cluster in 293T cells. Matched random control (MRC) sites were absent in this region.

(B) Receiver-operator characteristic (ROC) analysis of BRD2–4, Pol II, H3K4me3, and CpG islands (Berry et al., 2006). The area under the curve (AUC) is calculated for the different markers and shown in Table S3.

(C–G) Mean background-subtracted sequencing read density in 50 bp bins in a 10 kb window around (C and D) TSSs or (E–G) MLV integration sites. ChIP-seq read density for (C and E) BRD2–4 and (D and F) the unrelated transcription factors PHF8, ELK4, KAP1, and TCF7L2 are plotted on the left y axis. (G) Open chromatin (DNase I hypersensitivity [DHS]) and Pol II (ChIP-seq) read densities are plotted on the left axis, whereas those from micrococcal nuclease sequencing (MN-seq) revealing nucleosome positions are shown on the right axis. In (C–F), the number of MLV and MRC sites is plotted on the right y axis as a fraction of the total number of respective sites.

(H) ROC analysis of BRD4, Pol II, and CpG islands in primary human $CD4^+$ T cells. Corresponding AUC values are given in Table S3.

(I) Schematic representation of the *ITGAL* (Integrin alpha-L) locus and the 3′ end of the *TRB* (T cell receptor beta) locus highlighting MLV integration sites, BRD4 peaks, and MRC sites.

0, and 0 MRC sites. Genome-wide MLV integration was significantly ($p < 0.0001$, Mann-Whitney U test) enriched in regions bound by either of the three BET proteins: 42.7%, 23.5%, and 15% of integration sites were situated in BRD2, -3, or -4 islands respectively, compared to 1.9%, 0.9%, and 0.6% of MRC (Table S3). In addition, we observed stronger correlation with BET

protein binding than with previously described markers associated with MLV integration such as Pol II binding, H3K4me3, or CpG islands (9.1, 6.9, 4.4%, respectively) (Table S3) (Cavazza et al., 2013; Santoni et al., 2010). BRD2 proved to be the best predictor for MLV integration (Figure 3B), with over 50% of MLV integration sites locating within 149 bp of a BRD2 binding site (Table S3). Moreover, both MLV integration sites and BRD2–4 chromatin immunoprecipitation sequencing (ChIP-seq) tags concentrate around RefGene TSSs with a similar bimodal distribution (Figure 3C) (Cattoglio et al., 2010), which differs from the pattern of other transcription factors defined in 293T cells (KAP1, ELK4, TCF7L2, and PHF8; Table S2; Figure 3D). Analysis of the distribution of BRD2–4 around MLV integration sites revealed that the BRD2–4 occupancy is highest at the integration site itself (Figure 3E), whereas the maximal tag density for other transcription factors is adjacent to the integration site (Figure 3F). In addition, we observed a clear peak of nucleosome occupancy at the site of integration (Figure 3G), supporting the notion that MLV preferentially integrates into nucleosomal DNA in vivo (Roth et al., 2011). Indeed, open chromatin (Kundaje et al., 2012; Natarajan et al., 2012) and RNA Pol II are favored at either side of the integration site, in accordance with the preference for nucleosomal targeting (Figure 3G). Similar results were obtained when MLV integration sites were compared to ChIP-seq data for BRD4 in primary human $CD4^+$ T cells (Roth et al., 2011; Zhang et al., 2012). 18.8% of MLV integrations were located in BRD4 islands compared to 0.7% of MRC sites, 17.0% of Pol II peaks, and 2.3% CpG islands (Table S3). Receiver-operator characteristic (ROC) analysis confirmed that MLV integration sites correlate best with the BRD4 chromatin binding profile (Figure 3H; Table S3). Analysis of $CD4^+$ T cell-specific loci that are active and bound by BRD4, such as the *ITGAL* and the *TRB* locus, underscored the link between BET proteins (BRD4) and MLV integration (Figure 3I). The *ITGAL* locus (16p11.2), for instance, encodes integrin alpha-L (CD11A) and contains four BRD4 islands all of which are associated with MLV integration sites (MLV n = 52; MRC n = 6). A similar pattern was observed at the 3′ end of the T cell receptor beta locus (*TRB*, location 7q34) in the joining and constant segment coding region (MLV n = 33; MRC n = 2).

BRD4 Hybrids Retarget MLV Integration

Taken together, our data are consistent with a role of BET proteins in gammaretroviral integration site targeting. If this is the case, fusions of the MLV IN-interacting domain with another chromatin binding domain should redirect integration away from the MLV-like pattern. To unambiguously prove this hypothesis, we generated NIH 3T3 and SupT1 cell lines stably expressing a chimeric fusion protein linking $mBRD4_{ETSEED}$ with the chromatin binding domain of the lentiviral targeting factor, LEDGF/p75 (aa 1–324) ($LEDGF_{1-324}mBRD4_{ETSEED}$). As a control, we also generated cell lines expressing the quadruple-interaction-defective mutant ($LEDGF_{1-324}mBRD4_{ETSEEDmut}$). Protein expression, verified by western blotting, did not affect cell growth (data not shown). Following transduction with a MLV-based vector, we amplified MLV integration sites and analyzed their distribution. In line with an earlier report, MLV integration in wild-type SupT1 cells was enriched within a 2 kb window near TSSs (22.0%), CpG islands (23.1%), and DNase I-hypersensitive sites (DHSs) (47.7%) ($p < 0.001$ compared to MRC) (Figure 4A). Expression of $LEDGF_{1-324}mBRD4_{ETSEED}$ shifted integration away from these features and toward RefSeq genes (from 49.3% to 58.9%), whereas an intermediate phenotype was observed for the interaction-deficient mutant. Even more pronounced results were obtained in NIH 3T3 cells (Figure 4A). When integration sites were binned based on their distance to TSSs and CpG or DHS island midpoints (Figures 4B–4D, respectively), $LEDGF_{1-324}mBRD4_{ETSEED}$ overexpression targeted integration away from the TSS, CpG island, and DHS midpoints (compare red and green bars), a pattern reminiscent of that of lentiviral vector integration (purple bars) (Wu et al., 2003). In addition, when analyzing integration preferences relative to a wide range of genomic and epigenetic features (Figures 4E and S3), expression of the $LEDGF_{1-324}mBRD4_{ETSEED}$ shifted integration from a MLV to an HIV-like phenotype for all investigated markers, in contrast to overexpression of the ETSEED quadruple mutant protein. Together these data show that overexpression of a $LEDGF_{1-324}mBRD4_{ETSEED}$ fusion protein efficiently shifts the MLV integration profile, corroborating that BET proteins function as integration targeting factors for MLV.

DISCUSSION

Although HIV integration site targeting is mediated by LEDGF/p75 (Cherepanov et al., 2003; Ciuffi et al., 2005; Gijsbers et al., 2010; Llano et al., 2006; Schrijvers et al., 2012; Shun et al., 2007), the cellular cofactor driving MLV integration site targeting remained unknown. Here, we describe BET proteins as the MLV targeting factors. In agreement with Sharma et al. (2013), we show that BRD2, BRD3, and BRD4 specifically interact with MLV IN and that bromodomain inhibitors can block MLV replication at the integration step. Further, it was shown that a recombinant BRD4 deletion mutant containing the bromodomains and the ET domain stimulated MLV concerted integration in vitro (Sharma et al., 2013). BET protein knockdown or treatment with JQ1(+) decreases integration around TSSs and CpG islands (Sharma et al., 2013). In addition, we demonstrate that, even in the presence of endogenous BET proteins, MLV integration efficiently shifts toward an HIV phenotype upon expression of a $LEDGF_{1-324}mBRD4_{ETSEED}$ fusion, underscoring the role of BET proteins in MLV targeting. Retroviruses tend to direct integration into outward-facing major grooves on nucleosome-wrapped DNA (Roth et al., 2011). TSSs of expressed genes are nucleosome depleted, whereas the TSSs of the same genes when not expressed are nucleosome bound (Struhl and Segal, 2013). Because MLV integrates at TSSs of actively transcribed genes, the bimodal MLV integration pattern naturally follows. However, the exact role of BET proteins, known to bind polyacetylated histone tails found around TSSs, remains to be investigated.

The specificity of the interaction of BET proteins for gammaretroviral IN explains MLV integration site distribution. Lentiviral (HIV), alpharetroviral (RSV), nor spumaviral (PFV) IN interact with BET proteins. When and why different retroviral families evolved to interact with distinct targeting factors and how this relates to replication kinetics and pathogenesis remains poorly understood.

A

Virus-Cell line	type	Total sites	TSS within 2Kb	TSS within 4Kb	CpG within 2Kb	CpG within 4Kb	DNAse within 2Kb	DNAse within 4Kb	inRef gene
MLV-SupT1	insertion	696	22.0%	27.0%	23.1%	28.2%	47.7%	62.9%	49.3%
MLV-SupT1 LEDGF$_{1-324}$mBRD4$_{ETSEED}$	insertion	868	7.14% ***|***	13% ***|***	7.37% ***|***	12.6% ***|***	29.5% ***|*	45.4% ***|-	58.9% ***|***
MLV-SupT1 LEDGF$_{1-324}$mBRD4$_{ETSEED\ mut}$	insertion	903	14.7% ***|***	20.8% **|***	14.1% ***|***	20.9% ***|***	39.4% **|***	55.9% **|***	52.0% -|***
HIV-SupT1	insertion	4875	2.36%	6.56%	2.05%	7.67%	25.1%	43.9%	76.6%
MLV-SupT1	MRC	2082	2.64%	4.66%	2.4%	4.37%	14.6%	26.6%	39.6%
HIV-SupT1	MRC	14619	2.22%	4.51%	2.27%	4.77%	14.7%	26.6%	40.0%
MLV-NIH3T3	insertion	1281	27.5%	33,0%	40.4%	26.3%	31.4%	38.3%	42.2%
MLV-NIH3T3 LEDGF$_{1-324}$mBRD4$_{ETSEED}$	insertion	732	10.2% ***	16.3% ***	28.4% ***	9.29% ***	15.3% ***	26.1% ***	68.3% ***
MLV-NIH3T3 LEDGF$_{1-324}$mBRD4$_{ETSEED\ mut}$	insertion	927	16.4% ***	22.8% ***	31.3% ***	14.7% ***	20.9% ***	29.6% ***	57.7% ***
MLV-NIH3T3-WT	MRC	3845	2.39%	5.05%	10.5%	1.69%	3.43%	6.97%	36.6%

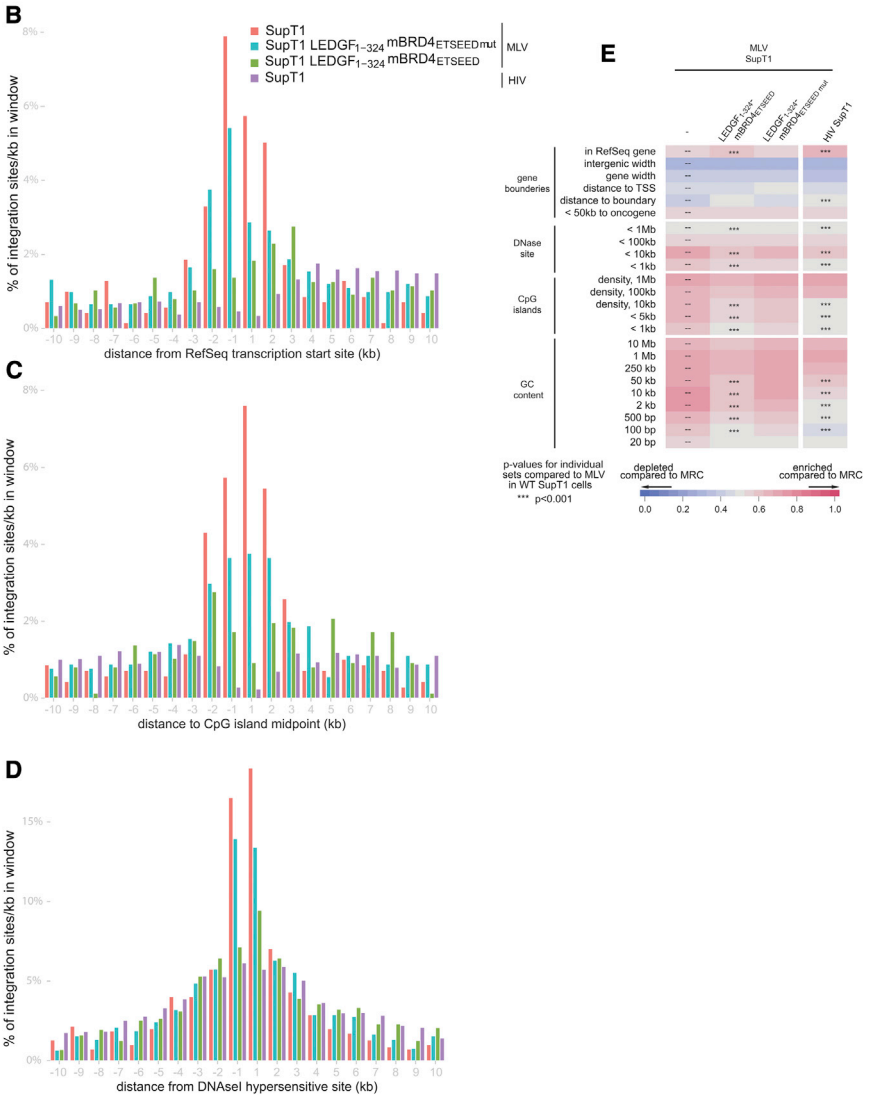


Figure 4. A LEDGF-BRD4 Chimeric Protein Retargets MLV Integration toward an HIV-like Pattern

(A) MLV or HIV integration sites obtained from SupT1 or NIH 3T3 cells and their genomic distribution. The percentage of integrations in RefSeq genes and around TSS, CpG islands, and DNase I-hypersensitive sites (2 and 4 kb window) is shown. Matched random control (MRC) sites or MLV and HIV integration sites in wild-type cells are shown. TSS, transcription start site; CpG, CpG-rich island; DNase, DNase I-hypersensitive site. Asterisks depict pairwise Fisher's test compared to MLV-SupT1|HIV-SupT1 *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. All data reach significance, $p < 0.001$, compared to MRC.

(B–D) Integration frequencies surrounding RefSeq TSSs, CpG islands, and DNase I-hypersensitive sites in SupT1 cells.

(E) Heatmap of integration frequency relative to genomic features in SupT1 cells, summarizing the relation between proviral integration sites and genomic features. Integration data sets are indicated above the columns. Genomic features analyzed are shown to the left of the corresponding row of the heatmap. Tile color indicates whether a particular feature is favored (red) or disfavored (blue) 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 (asterisks) show significance of departures from the MLV integration sites in WT SupT1 cells (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$, Wald statistics referred to χ^2 distribution). The naming of the genomic features is described in Brady et al. (2009).

Our in vitro analysis revealed that the evolutionary conserved BRD4 ET domain amino acids E652, E654, D656, and E658 are pivotal for interaction with MLV IN. Still, no interaction was detected between MLV IN and the yeast Bdf1 ET domain, suggesting the existence of other important interaction points (data not shown). Although LEDGF/p75 binds HIV IN across the catalytic core domain dimer interface, BET proteins interact with the 27 C-terminal aa of MLV IN. A single point mutant (W391A) is sufficient to abolish this interaction. Future research will show whether it is possible to replace the BET binding region by alternative chromatin interaction domains, possibly resulting in a safer retroviral integration site profile.

In conclusion, we propose a model for MLV integration targeting incorporating previous insights on the function of MLV p12 and the MLV IN-BET interaction (Figure S4) (Elis et al., 2012; Schneider et al., 2013).

EXPERIMENTAL PROCEDURES

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) and dissolved in DMSO.

Retroviral Vector Transduction

NIH 3T3 (2×10^4) or SupT1 (8×10^4) cells were seeded in 96-well plates. After 24 hr, cells were transduced with MLV- or HIV-derived vector particles. After 48 hr, cells were washed and cultured for another 24 hr in normal growth medium. Subsequently, cells were split, and 50% was reseeded for luciferase assays or flow cytometry analysis, whereas the remaining 50% was kept in culture to determine integrated copies and/or to perform integration site analysis.

Virus Infection

To monitor early MLV-eGFP replication in the presence/absence of BET inhibitors, 2×10^6 NIH 3T3 cells were seeded in a 6-well plate. Twenty-four hours later, cells were infected with MLV-eGFP at a multiplicity of infection (moi) of 5,000 reverse transcriptase units per well with or without the indicated compounds. Four hours after infection, cells were washed twice with phosphate-buffered saline (PBS) and fresh medium containing the indicated compounds. Cells were trypsinized and pelleted at 4, 8, 10, 12, and 24 hr after transfection. To measure integrated copies, cells were passaged over 10 days in the presence of 1 μM raltegravir to block viral replication.

Analysis of Next-Generation Sequencing Data

Data were obtained from the Gene Expression Omnibus (GEO) (Edgar et al., 2002) or the Sequence Read Archive (SRA) (Leinonen et al., 2011) as detailed in Table S2. Raw sequencing reads were mapped to the GRCh37/hg19 human genome assembly using the Bowtie2 short read aligner (Langmead and Salzberg, 2012). To delineate regions significantly enriched in BRD2–4, we used a shape-based peak calling approach (Hower et al., 2011) considering an average fragment length of 150 bp, the size of nucleosomal DNA, and a p value cutoff of 0.001. Distances from MLV integration or MRC sites to BRD2, -3, -4, Pol II, H3K4me3, or CpG islands were determined and analyzed (Berry et al., 2006) using Prism 5.0 (GraphPad software).

Sequence read densities were determined in 10 kb windows around MLV integration, MRC, or refGene TSSs by counting fragment-length-extended sequence tags in 50 bp bins for the sample and (when available) control libraries. The signal density was calculated as the difference between these two with negative values set to 0, after normalization by total sequencing depth. MLV and MRC sites were counted according to the same procedure in 50 bp bins in the studied regions and normalized to the total number of sites.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2013.09.040.

ACKNOWLEDGMENTS

We thank Nam Joo Van der Veken, Martine Michiels, and Paulien Van de Velde for excellent technical assistance. We acknowledge SyBioMa (formerly known as Prometa, KU Leuven) for technical assistance with LC-MS analysis and S. Aerts and A. Verfaillie for help with the BRD ChIP-seq analysis. Work at KU Leuven was supported by grants and fellowships from the FWO, IWT, and the KU Leuven research council (OT, IDO).

Received: July 18, 2013
Revised: July 28, 2013
Accepted: September 25, 2013
Published: October 31, 2013

REFERENCES

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.

Cattoglio, C., Pellin, D., Rizzi, E., Maruggi, G., Corti, G., Miselli, F., Sartori, D., Guffanti, A., Di Serio, C., Ambrosi, A., et al. (2010). High-definition mapping of retroviral integration sites identifies active regulatory elements in human multipotent hematopoietic progenitors. Blood *116*, 5507–5517.

Cavazza, A., Moiani, A., and Mavilio, F. (2013). Mechanisms of retroviral integration and mutagenesis. Hum. Gene Ther. *24*, 119–131.

Cavazzana-Calvo, M., Hacein-Bey, S., de Saint Basile, G., Gross, F., Yvon, E., Nusbaum, P., Selz, F., Hue, C., Certain, S., Casanova, J.L., et al. (2000). Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. Science *288*, 669–672.

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.

Ciuffi, A., Llano, M., Poeschla, E., Hoffmann, C., Leipzig, J., Shinn, P., Ecker, J.R., and Bushman, F. (2005). A role for LEDGF/p75 in targeting HIV DNA integration. Nat. Med. *11*, 1287–1289.

Deichmann, A., Hacein-Bey-Abina, S., Schmidt, M., Garrigue, A., Brugman, M.H., Hu, J., Glimm, H., Gyapay, G., Prum, B., Fraser, C.C., et al. (2007). Vector integration is nonrandom and clustered and influences the fate of lymphopoiesis in SCID-X1 gene therapy. J. Clin. Invest. *117*, 2225–2232.

Devaiah, B.N., and Singer, D.S. (2013). Two faces of brd4: mitotic bookmark and transcriptional lynchpin. Transcription *4*, 13–17.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. *30*, 207–210.

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.

Felice, B., Cattoglio, C., Cittaro, D., Testa, A., Miccio, A., Ferrari, G., Luzi, L., Recchia, A., and Mavilio, F. (2009). Transcription factor binding sites are genetic determinants of retroviral integration in the human genome. PLoS ONE *4*, e4571.

Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W.B., Fedorov, O., Morse, E.M., Keates, T., Hickman, T.T., Felletar, I., et al. (2010). Selective inhibition of BET bromodomains. Nature *468*, 1067–1073.

Gaspar, H.B., Parsley, K.L., Howe, S., King, D., Gilmour, K.C., Sinclair, J., Brouns, G., Schmidt, M., Von Kalle, C., Barington, T., et al. (2004). Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. Lancet *364*, 2181–2187.

Gijsbers, R., Ronen, K., Vets, S., Malani, N., De Rijck, J., McNeely, M., Bushman, F.D., and Debyser, Z. (2010). LEDGF hybrids efficiently retarget lentiviral integration into heterochromatin. Mol. Ther. *18*, 552–560.

Hacein-Bey-Abina, S., Le Deist, F., Carlier, F., Bouneaud, C., Hue, C., De Villartay, J.-P., Thrasher, A.J., Wulffraat, N., Sorensen, R., Dupuis-Girod, S., et al. (2002). Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy. N. Engl. J. Med. *346*, 1185–1193.

Hower, V., Evans, S.N., and Pachter, L. (2011). Shape-based peak identification for ChIP-Seq. BMC Bioinformatics *12*, 15.

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.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Leinonen, R., Sugawara, H., and Shumway, M.; International Nucleotide Sequence Database Collaboration. (2011). The sequence read archive. Nucleic Acids Res. *39*(Database issue), D19–D21.

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.

Lewinski, M.K., Yamashita, M., Emerman, M., Ciuffi, A., Marshall, H., Crawford, G., Collins, F., Shinn, P., Leipzig, J., Hannenhalli, S., et al. (2006). Retroviral DNA integration: viral and cellular determinants of target-site selection. PLoS Pathog. *2*, e60.

Lin, Y.-J., Umehara, T., Inoue, M., Saito, K., Kigawa, T., Jang, M.-K., Ozato, K., Yokoyama, S., Padmanabhan, B., and Güntert, P. (2008). Solution structure of the extraterminal domain of the bromodomain-containing protein BRD4. Protein Sci. *17*, 2174–2179.

Llano, M., Saenz, D.T., Meehan, A., Wongthida, P., Peretz, M., Walker, W.H., Teo, W., and Poeschla, E.M. (2006). An essential role for LEDGF/p75 in HIV integration. Science *314*, 461–464.

Mitchell, R.S., Beitzel, B.F., Schroder, A.R.W., Shinn, P., Chen, H., Berry, C.C., Ecker, J.R., and Bushman, F.D. (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol. *2*, E234.

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.

Nicodeme, E., Jeffrey, K.L., Schaefer, U., Beinke, S., Dewell, S., Chung, C.-W., Chandwani, R., Marazzi, I., Wilson, P., Coste, H., et al. (2010). Suppression of inflammation by a synthetic histone mimic. Nature *468*, 1119–1123.

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

Santoni, F.A., Hartley, O., and Luban, J. (2010). Deciphering the code for retroviral integration target site selection. PLoS Comput. Biol. *6*, e1001008.

Schneider, W.M., Brzezinski, J.D., Aiyer, S., Malani, N., Gyuricza, M., Bushman, F.D., and Roth, M.J. (2013). Viral DNA tethering domains complement replication-defective mutations in the p12 protein of MuLV Gag. Proc. Natl. Acad. Sci. USA *110*, 9487–9492.

Schrijvers, R., De Rijck, J., Demeulemeester, J., Adachi, N., Vets, S., Ronen, K., Christ, F., Bushman, F.D., Debyser, Z., and Gijsbers, R. (2012). LEDGF/p75-independent HIV-1 replication demonstrates a role for HRP-2 and remains sensitive to inhibition by LEDGINs. PLoS Pathog. *8*, e1002558.

Sharma, A., Larue, R.C., Plumb, M.R., Malani, N., Male, F., Slaughter, A., Kessl, J.J., Shkriabai, N., Coward, E., Aiyer, S.S., et al. (2013). BET proteins promote efficient murine leukemia virus integration at transcription start sites. Proc. Natl. Acad. Sci. USA *110*, 12036–12041.

Shun, M.-C., Raghavendra, N.K., Vandegraaff, N., Daigle, J.E., Hughes, S., Kellam, P., Cherepanov, P., and Engelman, A. (2007). LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. Genes Dev. *21*, 1767–1778.

Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. Nat. Struct. Mol. Biol. *20*, 267–273.

Studamire, B., and Goff, S.P. (2008). Host proteins interacting with the Moloney murine leukemia virus integrase: multiple transcriptional regulators and chromatin binding factors. Retrovirology *5*, 48.

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, G.P., Levine, B.L., Binder, G.K., Berry, C.C., Malani, N., McGarrity, G., Tebas, P., June, C.H., and Bushman, F.D. (2009). Analysis of lentiviral vector integration in HIV+ study subjects receiving autologous infusions of gene modified CD4+ T cells. Mol. Ther. *17*, 844–850.

Wu, X., Li, Y., Crise, B., and Burgess, S.M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. Science *300*, 1749–1751.

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.