

# HIV-1 Integrase Variants Retarget Viral Integration and Are Associated with Disease Progression in a Chronic Infection Cohort

Jonas Demeulemeester,<sup>1,2</sup> Sofie Vets,<sup>1</sup> Rik Schrijvers,<sup>1</sup> Paradise Madlala,<sup>1,3</sup> Marc De Maeyer,<sup>2</sup> Jan De Rijck,<sup>1</sup> Thumbi Ndung'u,<sup>3,4,5,6</sup> Zeger Debyser,<sup>1,7,\*</sup> and Rik Gijsbers<sup>1,7,\*</sup>

<sup>1</sup>Laboratory for Molecular Virology and Gene Therapy, Department of Pharmaceutical and Pharmacological Sciences

<sup>2</sup>Laboratory for Biomolecular Modeling, Department of Chemistry

KU Leuven–University of Leuven, 3000 Leuven, Belgium

<sup>3</sup>HIV Pathogenesis Programme, Doris Duke Medical Research Institute

<sup>4</sup>KwaZulu-Natal Research Institute for Tuberculosis and HIV

Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, 4013 Durban, South Africa

<sup>5</sup>Ragon Institute of MGH, MIT and Harvard, Cambridge, MA 02139, USA

<sup>6</sup>Max Planck Institute for Infection Biology, Chariteplatz, 10117 Berlin, Germany

<sup>7</sup>Co-senior author

\*Correspondence: [zegeer.debyser@med.kuleuven.be](mailto:zegeer.debyser@med.kuleuven.be) (Z.D.), [rik.gijsbers@med.kuleuven.be](mailto:rik.gijsbers@med.kuleuven.be) (R.G.)

<http://dx.doi.org/10.1016/j.chom.2014.09.016>

## SUMMARY

Distinct integration patterns of different retroviruses, including HIV-1, have puzzled virologists for over 20 years. A tetramer of the viral integrase (IN) assembles on the two viral cDNA ends, docks onto the target DNA (tDNA), and catalyzes viral genome insertion into the host chromatin. We identified the amino acids in HIV-1 IN that directly contact tDNA bases and affect local integration site sequence selection. These residues also determine the propensity of the virus to integrate into flexible tDNA sequences. Remarkably, natural polymorphisms IN<sub>S119G</sub> and IN<sub>R231G</sub> retarget viral integration away from gene-dense regions. Precisely these variants were associated with rapid disease progression in a chronic HIV-1 subtype C infection cohort. These findings link integration site selection to virulence and viral evolution, but also to the host immune response and antiretroviral therapy, since HIV-1 IN<sub>119</sub> is under selection by HLA alleles and integrase inhibitors.

## INTRODUCTION

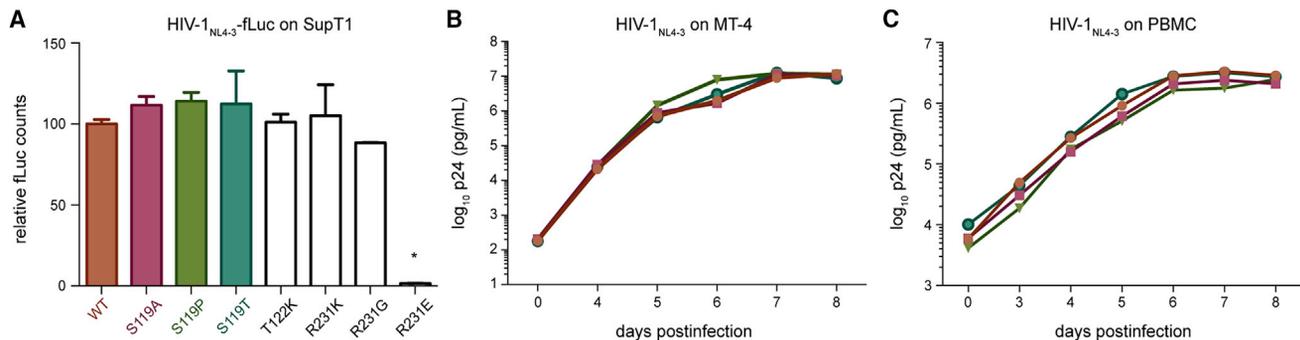
The retroviral life cycle is characterized by a key integration step in which a DNA copy of the viral RNA genome is inserted into the host chromatin. In the past decade, significant progress has been made in understanding this process. Massive integration site sequencing revealed that different retroviruses show distinct preferences with regard to the chromatin environment. Lentiviral integrase (IN) co-opts the cellular chromatin reader lens epithelium-derived growth factor/p75 (LEDGF/p75; *PSIP1*) (Ciuffi et al., 2005) to target integration into active genes. Gammaretroviral integration, on the other hand, is enriched near strong enhancers and transcription start sites by an interaction between

IN and bromodomain and extraterminal domain (BET) proteins (BRD2, BRD3, and BRD4; *BRD2–4*) (De Rijck et al., 2013; Gupta et al., 2013; Sharma et al., 2013). It is, however, the viral intasome, a tetramer of IN assembled on the two viral cDNA ends, that catalyzes the final DNA cutting and joining reactions and thus determines the precise position of integration. Before strand transfer can proceed, the intasome has to dock onto target (nucleosomal) DNA (tDNA), forming the target capture complex (TCC) (Maertens et al., 2010). Molecular recognition between intasome and tDNA imposes specific constraints on the implicated sequence.

Already in 2001, patient-derived viral INs were employed to propose that residue IN<sub>119</sub> plays a role in target site selection on naked DNA templates in vitro (Harper et al., 2001). However, a mechanistic basis for these observations was lacking. The tools for this would only become available in 2010, when crystal structure of the prototype foamy virus (PFV) intasome TCC was reported (Maertens et al., 2010). This model provides insight into the retroviral integration machinery and suggests how the intasome recognizes and docks onto tDNA. Recently, this structure was used to show that central tDNA distortion in the HIV-1 TCC is spread out over two overlapping dinucleotides surrounding the integration site palindrome dyad axis (Serrao et al., 2014). Additionally, the authors demonstrated that residue S119 in HIV-1 IN is indeed analogous to A188 in PFV IN, and that IN<sub>S119A</sub> and IN<sub>S119T</sub> substitutions can alter target site selection of recombinant IN protein into naked DNA substrates in vitro. The effect of tDNA contacting amino acids (aa) on integration site selection in a viral context, however, remains largely unstudied.

In this work, we combined structural information on the PFV TCC with conservation in retroviral IN protein alignments to determine aa-tDNA base contacts. We generated HIV-1 variants based on the observed variability at these positions, assessed replication capacities, and performed integration site sequencing to reveal their integration preferences. Altered molecular recognition in the TCC underlying distinct local tDNA nucleotide biases was extrapolated to other retroviral genera. Finally, we examined the global integration profile of these





**Figure 2. Replication of tDNA Contact Variant Viruses**

(A) Transduction efficiency as measured by fLuc reporter activity (relative light units, RLU/μg protein) 3 days postinfection of SupT1 cells for the indicated HIV-1<sub>NL4-3</sub>-fLuc carrying IN mutants. Activity is reported as percentage of HIV-1<sub>NL4-3</sub>-fLuc carrying WT IN. Data represent averages and SDs from triplicate measurements. Significant deviations from WT are indicated (\**p* < 0.01, *t* test).

(B and C) Viral breakthrough experiments of HIV-1<sub>NL4-3</sub> WT and IN<sub>S119A/P/T</sub> viruses (colored as in A) in MT-4 cells (B) and PBMCs (C). See also Figure S2.

equivalent to HIV-1 IN<sub>S119</sub> are occupied by small aa (Pro, Ala, Ser, Thr). Positions corresponding to HIV-1 IN<sub>122</sub> also harbor substantial variation. Interestingly, feline immunodeficiency virus (FIV) encodes a Lys residue at this position, which may directly contact tDNA bases.

More divergence is present in the CTD β1/β2-loop; next to substitutions, insertions and deletions are also present, complicating the retroviral alignment (Figures 1B and S1). In accordance with a possible role in tDNA binding, conservation is higher again at positions corresponding to HIV-1 IN<sub>231</sub>. Only equine infectious anemia virus (EIAV) does not encode a basic Lys or Arg tether, but a Gly instead.

### Amino Acids Contacting tDNA Bases Can Be Varied without Compromising Viral Fitness

We next wondered how introduction of the observed aa into HIV-1 would affect the virus. Therefore, substitutions at IN positions 119, 122, and 231 were introduced into multiple-round HIV-1<sub>NL4-3</sub> and in the single-round firefly luciferase reporter version, HIV-1<sub>NL4-3</sub>-fLuc. At position IN<sub>119</sub> we generated S119A/P/T substitutions (Table S1). As previously observed, a long and flexible aa at position IN<sub>122</sub> may directly contact tDNA bases. As Lys was observed at the equivalent position in FIV IN (FIV IN<sub>K124</sub>), we introduced a T122K substitution (Figures 1B and S1; Table S1). At the third tDNA base contact point, HIV-1 IN<sub>231</sub>, we investigated R231K and R231G substitutions (Table S1). As a control for our model and to corroborate its tethering role in HIV-1, we also included an IN<sub>R231E</sub> substitution, in analogy to the previously reported recombinant PFV IN<sub>R329E</sub> (Maertens et al., 2010).

Single-round HIV-1<sub>NL4-3</sub>-fLuc viruses carrying the different IN variants were produced, and their transduction efficiency was assessed in SupT1 and HeLa P4 cells (Figures 2A and S2A, respectively). Viral variants IN<sub>S119A/P/T</sub> and IN<sub>T122K</sub> maintained wild-type (WT) transduction efficiency. Alterations at position IN<sub>231</sub> were tolerated as well; a conservative IN<sub>R231K</sub> substitution resulted in WT activity, while removal of this basic tether in the IN<sub>R231G</sub> virus resulted in a trend toward lower transduction efficiency (88.3% ± 0.3% of WT). However, inversion of the charge at this position (IN<sub>R231E</sub>) severely hampered transduction (3.1% ± 0.7% of WT, *p* < 0.01, *t* test). For comparison, we included HIV-

1<sub>NL4-3</sub>-fLuc virus containing the enzymatically dead IN catalytic triad mutant D64N/D116N/E152Q. This mutant exhibited a transduction efficiency 3–4 log lower than WT virus (data not shown).

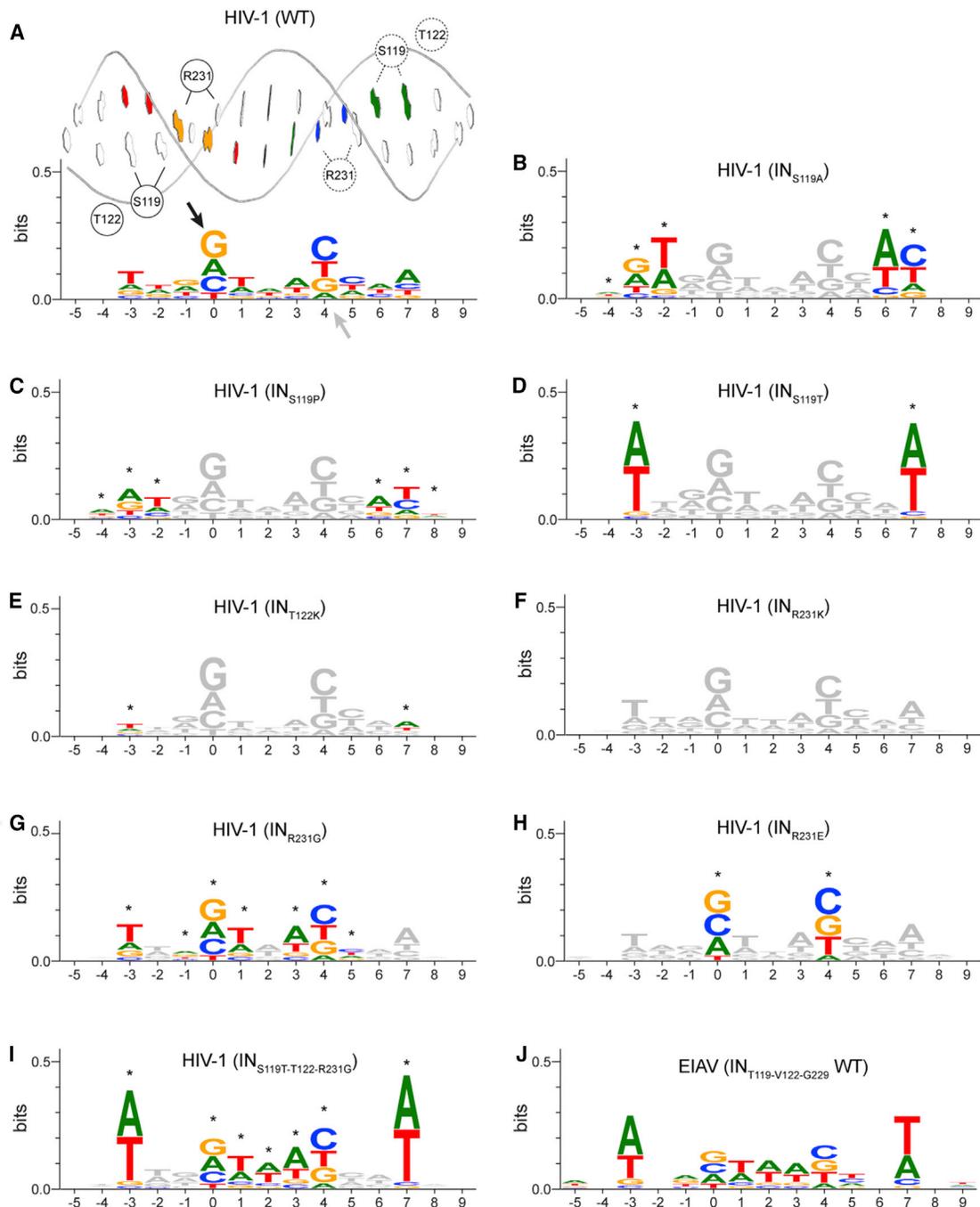
To confirm their activity, we determined the replication capacity of the HIV-1<sub>NL4-3</sub> IN variants in a multiple-round viral breakthrough assay in MT-4 cells and peripheral blood mononuclear cells (PBMCs). Multiple-round replication was monitored by measuring viral p24 production in the supernatant. All three IN<sub>S119</sub> mutant viruses exhibited growth kinetics comparable to WT HIV-1<sub>NL4-3</sub> corroborating the results of the fLuc assay (Figures 2B and 2C). The IN<sub>R231G/K</sub> variants also replicated similarly to WT, while IN<sub>T122K</sub> virus lagged behind (Figures S2B and S2C). As expected, IN<sub>R231E</sub> virus did not replicate on either MT-4 cells or PBMCs.

These results indicate that substitutions of tDNA base-contacting aa in IN do not necessarily affect the viral transduction efficiency or fitness.

### Altering HIV-1 Intasome tDNA Contacts Locally Retargets Viral Integration

Retroviral INs show weak, but discernable, target sequence specificity at the site of integration. To assess whether the above-mentioned substitutions alter the local integration site preferences, we infected SupT1 and HeLa P4 cells with the respective HIV-1<sub>NL4-3</sub> viruses and employed 454 technology to sequence proviral integration sites. A total of 36,264 distinct sites were retrieved, 31,615 in SupT1 (Table S2) and 4,649 in HeLa P4 cells (data not shown). The 15 base pair (bp) genomic DNA sequence surrounding the integration site roughly corresponds to the intasome footprint. Sequence conservation and relative base frequencies in this region were evaluated using sequence logos.

Figure 3A shows the local, palindromic integration site sequence logo for WT HIV-1<sub>NL4-3</sub> (IN<sub>S119-T122-R231</sub>). Most notable is the bias against thymine at the site of strand transfer (position 0) and a corresponding symmetrical bias against adenine at position 4 (the point of strand transfer on the complementary strand). The intasome structure suggests that a thymine C5-methyl is incompatible with the transesterification reaction (Maertens et al., 2010). Further away from the palindrome dyad



**Figure 3. Integration Site Sequence Logos of HIV-1<sub>NL4-3</sub> IN Variants**

Sequence conservation is indicated as the total height of each stack (measured in bits), while the relative height of bases in a stack reflects base frequencies at that position. Arrows on the HIV<sub>NL4-3</sub> WT logo (A) indicate the sites of strand transfer (position 0) on the plus (black) and minus (gray) target DNA strands. On top is a B-DNA cartoon of the tDNA and aa-base contacts where positions directly correspond to the sequence logo below.

(B–I) Integration site sequence logos for the indicated HIV-1<sub>NL4-3</sub> variants. Positions showing an altered base distribution compared to WT HIV-1<sub>NL4-3</sub> are colored (\* $p < 10^{-4}$ ,  $\chi^2$  test).

(J) Integration site sequence logo for EIAV. See also Figures S3 and S4.

axis, at positions  $-2$ ,  $-3$  on the plus and  $6$ ,  $7$  on the minus strand, are the bases contacting IN<sub>119</sub> (Figures 3A and S3A). WT HIV-1<sub>NL4-3</sub> IN codes for Ser at this position, inducing a small predilection for thymine at position  $-3$  through a weak hydrogen bonding

interaction between the S119 hydroxyl and the complementary adenine N3 at position  $7$  (Figure S3A). G:C bp are disfavored at position  $-3$ , as the presence of a guanine C2 amino group in the minor groove would sterically hinder the Ser side chain.

IN<sub>S119A</sub> substitution in HIV-1<sub>NL4-3</sub>, as observed in SIV and HIV-2 IN, resulted in stronger nucleotide biases (Figure 3B).  $\chi^2$  analyses comparing base distributions at each position to those observed for WT HIV-1<sub>NL4-3</sub> indicate significantly different preferences at positions -4, -3, -2, and symmetrically 6 and 7 ( $p < 10^{-4}$ ). The smaller Ala allows the CCD  $\alpha 2$  helix to approach the tDNA bases more closely. Van der Waals interactions with the methyl side chain are established, preferably involving a guanine base at position -3 and a complementary cytosine at 7 (Figure S3B). Modeling also suggests a polar contact between the Ala backbone NH and the neighboring -2 thymine C2 carbonyl, or adenine N3. This interaction is not possible with a G:C bp due to its bulky guanine C2 amino group in the minor groove, explaining the bias for A:T bp at this position.

IN<sub>S119P</sub> substitution gives rise to small additional sequence biases at positions -4 and complementary 8, suggesting that contacts may be established with these bases as well (Figures 3C and S3C). Preferences distinct from those of WT HIV-1<sub>NL4-3</sub> are also observed at positions -2, -3 and 6, 7. Pro residues are frequently observed to make extensive ring-stacking interactions with adenine, and to a lesser extent, thymine (Luscombe et al., 2001), rationalizing the novel predilections. Due to the more optimal van der Waals interactions, IN<sub>S119P</sub> virus generally prefers -3 R:Y 7 bp to -3 Y:R 7 bp (International Union of Biochemistry and Molecular Biology base notation).

Increasing the volume of the IN<sub>S119</sub> side chain by substitution with Thr, as present in EIAV, results in a strong preference for A:T bp at positions -3 and 7, while leaving the remainder of the logo unaffected (Figure 3D). Our model suggests that the IN<sub>S119T</sub> side chain can engage either a thymine C2 carbonyl or adenine N3 in a hydrogen bonding interaction (Figure S3D), whereas the additional minor groove bulk associated with a G:C bp is sterically disfavored.

Sequence logo analysis of HIV-1<sub>NL4-3</sub> IN<sub>T122K</sub> integration sites shows a decreased bias at position -3 (and 7) compared to WT (Figure 3E), suggesting that additional tDNA contacts can be introduced at IN<sub>122</sub>. While we cannot exclude the possibility that the IN<sub>T122K</sub> substitution causes a slight shift of the CCD  $\alpha 2$  helix, altering interactions with IN<sub>S119</sub>, the observed loss of bias likely reflects the broad hydrogen bonding capacities of Lys with minor groove bases (Figure S3E) (Luscombe et al., 2001).

As expected, the conservative IN<sub>R231K</sub> substitution, occurring in several lentiviral INs, does not alter the local integration site preferences (Figure 3F). Both the WT Arg (Figure S3F) and the variant Lys side chain can contact tDNA bases at positions -1, 0 and 4, 5 to balance the energetic penalty accompanying the strong central tDNA bending (Figure S3G). Both aa are relatively promiscuous in establishing hydrogen bonding interactions with major groove bp and do not impose overly specific sequence requirements (Luscombe et al., 2001; Maertens et al., 2010).

Analysis of IN<sub>R231G</sub> virus integration sites reveals that sequence biases at positions -1, 0 and complementary 4, 5 are reduced (Figure 3G), with significant decreases (> five SDs) in the proportion of guanine at the former two positions and of cytosine at the latter. Both the WT IN<sub>R231</sub> and the variant IN<sub>R231K</sub> side chains will optimally engage a position 0 guanine base in a bidentate hydrogen bonding interaction (Figures S3F and S3G) (Luscombe et al., 2001). The IN<sub>R231G</sub> virus, however, is stripped of this interaction (Figure S3H). Simultaneously,

increased preferences for thymine and adenine are observed at positions 1 and 3, respectively, hinting at an increased frequency of flexible TAA, AAA, AAT trinucleotides at the integration site center (positions 1-3; Figure 3G) (Gabrieli and Pongor, 1996; Satchwell et al., 1986). To balance the energetic penalty for tDNA bending, integration into more flexible sequences may be required in the absence of a central Arg or Lys tether.

The IN<sub>R231E</sub> virus integration site sequence logo showed an increased bias toward cytosine at the site of strand transfer (Figure 3H). In this case a novel hydrogen bonding interaction can be established between the Glu side chain and a cytosine C4-amino group (Figure S3I) (Luscombe et al., 2001). An increased bias for cytosine was observed for PFV IN<sub>R329E</sub>-mediated integration in vitro (Maertens et al., 2010). However, in this context, the bias appeared at the base just 5' to the site of strand transfer.

In conclusion, substitutions at positions IN<sub>119</sub>, IN<sub>122</sub>, and IN<sub>231</sub> result in local retargeting of HIV-1 integration.

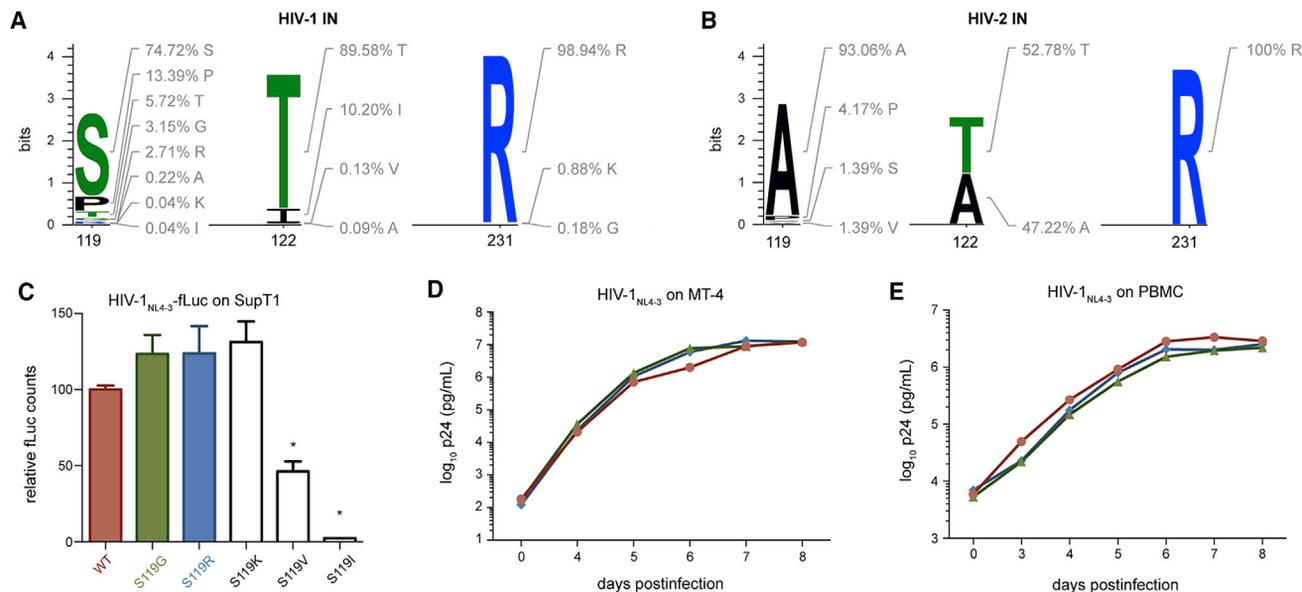
### HIV-1<sub>NL4-3</sub> IN<sub>S119T-R231G</sub> Mimics EIAV Local Integration Site Preferences

Different retroviruses exhibit distinct nucleotide preferences at their local integration site. Introduction of single aa substitutions in HIV-1<sub>NL4-3</sub> IN allows mimicry of local preferences of the respective viruses. Comparing positions -4 to -2 in HIV-1<sub>NL4-3</sub> IN<sub>S119P</sub> and FIV (Figures 3C and S4B) reveals preferences for the same nucleotides (W, R, and T at -4, -3, and -2, respectively). Likewise, biases toward R and W at respective positions -3 and -2 for HIV-1<sub>NL4-3</sub> IN<sub>S119A</sub> are also observed for HIV-2 and SIV (Figures 3B, S4C, and S4D). To assess whether the same molecular interaction patterns govern local integration site selection for lentiviruses, and by extension, retroviruses, we attempted to shift the HIV-1 local integration site preference toward that of EIAV (Figure 3J). EIAV IN contains T119, V122, and G229, matching HIV-1 IN positions 119, 122, and 231. Introducing either single aa substitution IN<sub>S119T</sub> or IN<sub>R231G</sub> altered the integration site nucleotide preferences of HIV-1<sub>NL4-3</sub>. As IN<sub>V122</sub> is unlikely to interact directly with tDNA, we created an HIV-1<sub>NL4-3</sub> IN<sub>S119T-R231G</sub> double mutant. Sequence logo analysis of viral integration sites in SupT1 cells reflects a merger between the effects observed for the single IN substitutions: a strong bias toward A:T bp at positions -3 and 7 and increases in their presence around the palindrome dyad axis (Figure 3I). Of note, the central base itself (position 2) also exhibits a significantly heightened bias toward A:T bp, which was not observed for the IN<sub>S119T</sub> or IN<sub>R231G</sub> viruses (Figures 3D and 3G). Overall, integration site base preferences of the HIV-1<sub>NL4-3</sub> IN<sub>S119T-R231G</sub> virus are similar to those of EIAV, indicating close mimicry of the tDNA base interactions in the EIAV intasome.

### Patient-Derived HIV Sequences Reveal Further Polymorphisms at IN<sub>119</sub>

Complementary to our breadth-first approach evaluating IN sequences of different lenti- and retroviruses, we exploited the large number of patient-derived HIV IN sequences contained in the Los Alamos HIV Sequence Database to obtain an in-depth view on the variability at tDNA base contact positions.

Figure 4A summarizes the sequence conservation and aa frequencies at positions IN<sub>119</sub>, IN<sub>122</sub>, and IN<sub>231</sub> for 2,276 aligned HIV-1 sequences. Interestingly, all aa observed in other



**Figure 4. HIV IN tDNA Contacts Variants and Their Replication**

(A and B) Sequence conservation (indicated in bits) and aa frequencies of (A) 2,276 aligned HIV-1 and (B) 72 HIV-2 sequences at positions IN<sub>119</sub>, IN<sub>122</sub>, and IN<sub>231</sub>. (C) Transduction efficiency as measured by fLuc reporter activity (relative light units, RLUs/ $\mu$ g protein) 3 days postinfection of SupT1 cells for the indicated HIV-1<sub>NL4-3</sub>-fLuc single-round viruses. fLuc reporter activity is reported as percentage of HIV-1<sub>NL4-3</sub>-fLuc carrying WT IN. Data represent averages and SDs from triplicate measurements. Significant deviations from WT are indicated (\*p < 0.01, t test).

(D and E) Viral breakthrough experiments of HIV-1<sub>NL4-3</sub> WT and IN<sub>S119G/R</sub> viruses, colored as in (C), on MT-4 cells (D) and PBMCs (E). See also Figures S2 and S5.

lentiviruses occur as natural polymorphisms in HIV-1. Additional substitutions to Gly or Arg and less frequently to Lys or Ile are found at IN<sub>119</sub>. HIV-1 IN<sub>122</sub> allows less diversity, with Thr occurring most frequently, followed by Ile, Val, and Ala. The strongest conservation is found at position IN<sub>231</sub>, where the entropy approaches four bits; Arg is preferred, followed by Lys, and occasionally Gly.

Of note, we also split up the results with respect to the different HIV-1 groups and subtypes. Distinct aa frequencies exist at IN<sub>119</sub> when comparing viral groups, subtypes, and circulating recombinant forms (Figure S5). Regardless of the higher conservation, similar observations could be made for IN positions 122 and 231. In particular, IN<sub>R231K</sub> was found among subtype B only, while IN<sub>R231G</sub> substitution was unique for subtype C, although the limited number of sequences may preclude finding of these variants among other subtypes.

Analysis of 72 HIV-2 sequences is shown in Figure 4B. Aside from the highly frequent Ala at position IN<sub>119</sub>, Pro, Ser, and Val are also encountered. At position IN<sub>122</sub>, Thr and Ala are found with nearly equal probabilities. Arg is the only aa found at HIV-2 IN<sub>231</sub>. Indubitably, the smaller sample size plays a role in this finding. Nevertheless, it further underscores the contribution of IN<sub>R231</sub> to tDNA binding.

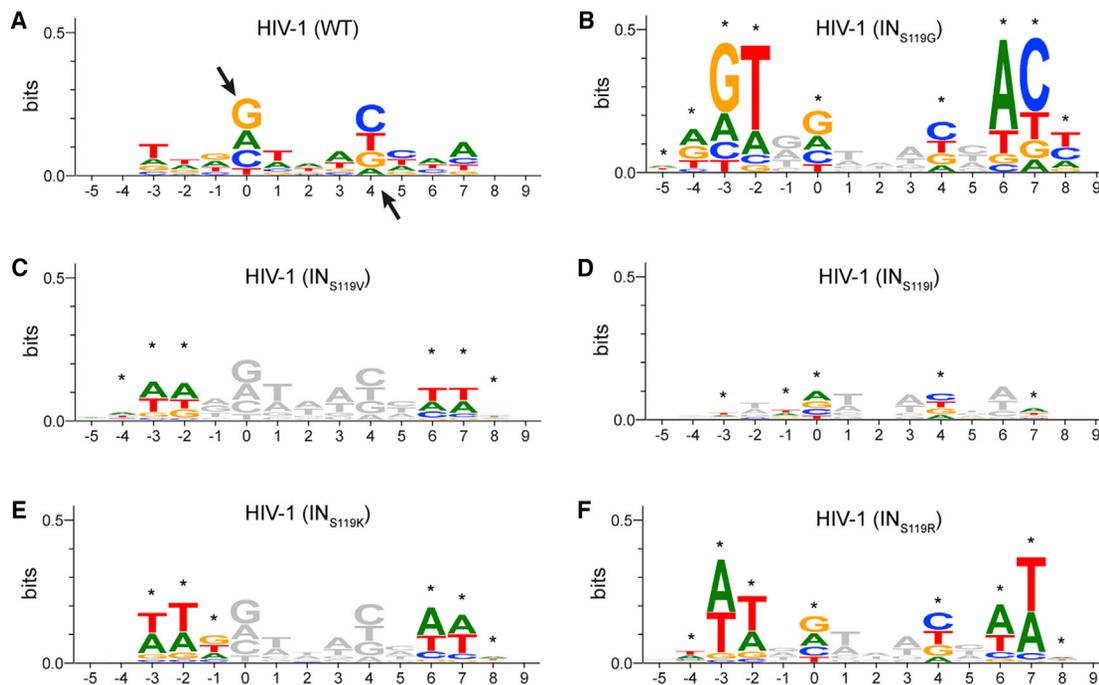
Based on these results, we generated and characterized additional single-round HIV-1<sub>NL4-3</sub>-fLuc and multiple-round HIV-1<sub>NL4-3</sub> IN<sub>S119G/R/K/V/I</sub> variants. Transduction efficiency in SupT1 and HeLa P4 cells revealed that IN<sub>S119G/R/K</sub> viruses maintained WT transduction efficiency (Figures 4C and S2D, respectively). Substitution by the more inflexible, bulky Val and Ile at this position caused a 2- and 30-fold reduction in fLuc signal, respectively (p < 0.01, t test). Clinically relevant IN<sub>S119G/R</sub> variants

exhibited growth kinetics similar to WT virus in a multiple-round infection assay in MT-4 cells and PBMCs (Figures 4D and 4E). IN<sub>S119K</sub> virus also replicated similarly to WT, while the IN<sub>S119I</sub> variant did not break through, and IN<sub>S119V</sub> virus exhibited an intermediate phenotype (Figures S2E and S2F).

#### Naturally Occurring HIV-1 IN<sub>119</sub> Variants Exhibit Distinct Local Integration Preferences

If a close approach between the CCD  $\alpha$ 2 helix and tDNA allows the formation of additional polar contacts as suggested for HIV-1<sub>NL4-3</sub> IN<sub>S119A</sub>, then the IN<sub>S119G</sub> variant should demonstrate this more clearly. Indeed, integration site analysis shows an increased preference for thymine at position -2 (Figure 5B). Additionally, a second weak electrostatic interaction is established in this case, between the polar Gly C<sub>2</sub>-H and the C2 carbonyl group of a pyrimidine base (Figure S3J). Together with increased van der Waals contacts with a -3 G:C 7 bp, this likely results in the observed preferences at nucleotide -3. The close approach allows contacts to be established at positions -4 and 8 as well, giving rise to clear nucleotide biases. Base distributions at the sites of strand transfer are distinct from WT. Biases against thymine at position 0 and the complementary adenine at 4 are alleviated (> five SDs). Slight changes in the overall tDNA binding mode permitted by the close fit between the IN<sub>S119G</sub> CCD  $\alpha$ 2 helix and the tDNA minor groove may be enough to reduce clashes with the thymine C5-methyl during transesterification. A trend toward this effect was also observed for the IN<sub>S119A</sub> virus (p values of 0.0017 and 0.024 for positions 0 and 4, respectively).

Comparing sequence logos for the IN<sub>S119V</sub> (Figure 5C) and IN<sub>S119T</sub> viruses underscores the value of the hydrogen bond



**Figure 5. Integration Site Sequence Logos of HIV-1<sub>NL4-3</sub> IN Variants**

(A–F) Local integration sites obtained for (A) HIV-1<sub>NL4-3</sub> WT and variants (B) IN<sub>S119G</sub>, (C) IN<sub>S119V</sub>, (D) IN<sub>S119I</sub>, (E) IN<sub>S119K</sub>, and (F) IN<sub>S119R</sub>. Positions showing an altered base distribution compared to WT HIV-1<sub>NL4-3</sub> (A) are colored (\* $p < 10^{-4}$ ,  $\chi^2$  test). See also Figure S3.

with Thr as well as the limitations for steric bulk at this site. HIV-1<sub>NL4-3</sub>-fLuc IN<sub>S119V</sub> virus indeed showed a 2-fold reduction in infectivity. With the loss of the hydrogen bonding potential, the selectivity for A:T bp at positions  $-3$  and  $7$  is weakened. However, the slightly larger Val side chain also establishes contacts around this position (Figure S3K). Increased preferences for the smaller A:T bp are apparent at positions  $-4$  to  $-2$  (and  $6$ – $8$ ) at the cost of the more bulky G:C.

Further increase in the aa bulk in the IN<sub>S119I</sub> mutant produced a considerably crippled virus. As a result, we retrieved insufficient canonical integration sites ( $n = 114$ ) to draw solid conclusions on its nucleotide biases (Figure 5D).

The IN<sub>S119K</sub> virus biases against G:C bp at positions  $-2$ ,  $-3$  (and  $6$ ,  $7$ ) of the viral integration site (Figure 5E), highlighting the space factor, but also suggesting the establishment of hydrogen bonding between the Lys  $\epsilon$ -amino group and a thymine C2 carbonyl or adenine N3 (Figure S3L). The IN<sub>S119R</sub> variant also exhibits a bias against G:C bp at the same two positions (Figure 5F). Additionally, preferences for thymine at position  $-2$  and complementary adenine at  $6$  are apparent. Interactions are also established at positions  $-4$  or  $8$ , as a significant preference for A:T bp appears (Figure S3M). Base distributions at the sites of strand transfer are also affected. It is possible that the size or the interactions of the Arg side chain induce a subtle shift in the conformation of the tDNA, which is propagated to the active sites.

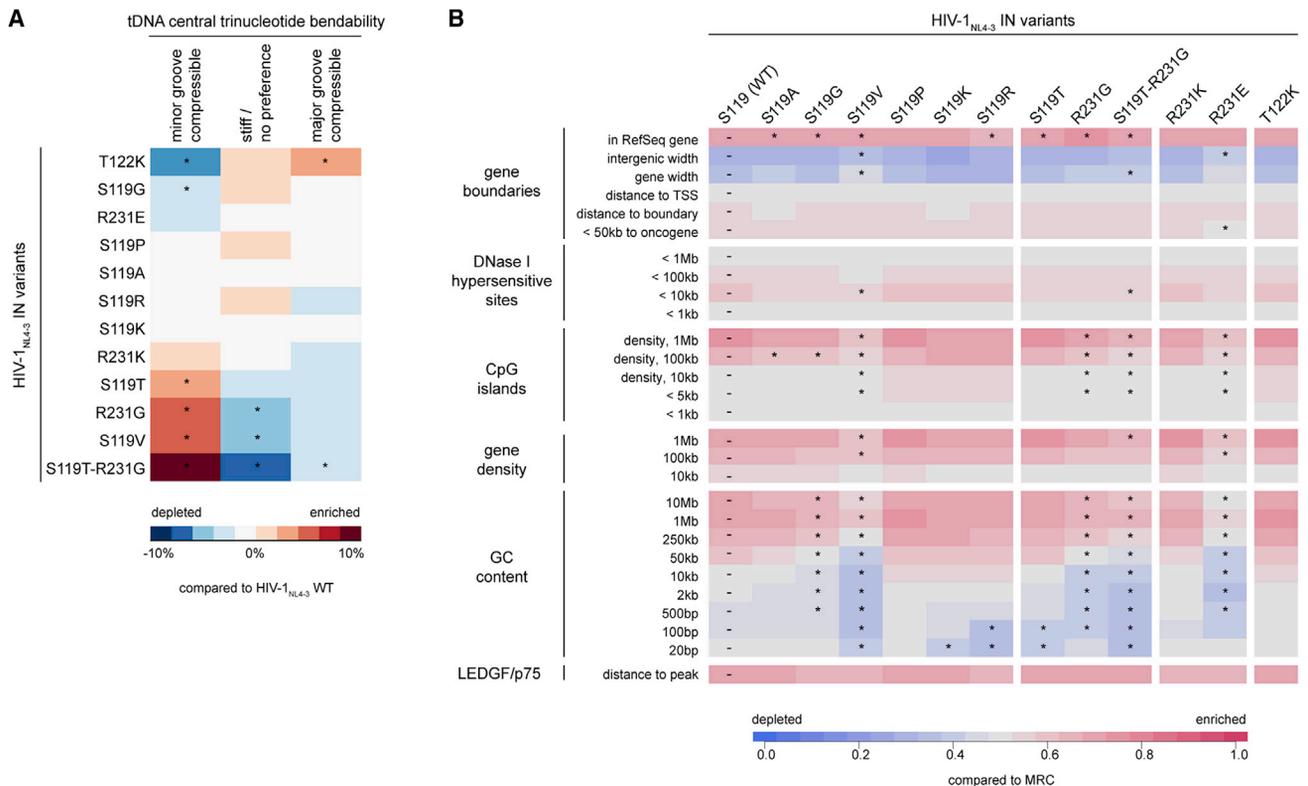
#### IN Amino Acid–tDNA Base Contacts Affect tDNA Flexibility Requirements

In the intasome, the tDNA major groove is pried open at the center of the integration site while the minor groove is com-

pressed, resulting in a partial destacking of the central bp. In the case of HIV-1, which catalyzes strand transfer with a 5 bp instead of a 4 bp stagger as in PFV, this severe bending can extend over a central trinucleotide instead of a dinucleotide. In order to investigate the effect of tDNA flexibility on our HIV-1<sub>NL4-3</sub> variants, we grouped the 64 possible trinucleotides (32 considering base complementarity) into three classes (Gabriellian and Pongor, 1996; Satchwell et al., 1986): (i) those that prefer minor groove compression, (ii) those that are stiff/show no preference, and (iii) those inclined toward major groove compression. Next, the frequencies of these classes were determined and compared to those observed for WT HIV-1<sub>NL4-3</sub> (Figure 6A). The IN<sub>S119I</sub> virus was excluded due to its limited data set.

As expected, the IN<sub>R231G</sub> variant shows an increased tendency to integrate into minor groove-compressible sequences when compared to WT. This is accompanied by biases against stiff and major groove-compressible trinucleotides. Both the IN<sub>R231K</sub> and IN<sub>R231E</sub> viruses did not show altered preferences regarding central tDNA flexibility. The novel interactions of these viruses with tDNA bases may balance the energetic penalty associated with bending.

A remarkable correlation emerges between the bias toward minor groove-compressible trinucleotides and steric bulk at position IN<sub>119</sub>. While the IN<sub>S119G</sub> variant shows a preference for more rigid sequences, the IN<sub>S119A/P</sub> viruses behave more like WT, and the bulkier HIV-1<sub>NL4-3</sub> IN<sub>S119T/V</sub> show the strongest bias toward minor groove-compressible trinucleotides. The IN<sub>S119K/R</sub> viruses, due to their flexible side chains, do not alter the tDNA flexibility requirements for integration. The EIAV mimic, HIV-1<sub>NL4-3</sub> IN<sub>S119T-R231G</sub>, shows the combined effect of the



**Figure 6. tDNA Bendability and Genomic Integration Preferences**

(A) Integration site central trinucleotide (positions 1–3) bendability heat map. Depicted are the changes in frequency of occurrence relative to WT HIV-1<sub>NL4-3</sub> for each of three groups of trinucleotides: those preferring minor groove compression, those that are stiff/show no preference, or those favoring major groove compression. Statistically significant deviations from the WT HIV-1<sub>NL4-3</sub> distribution are indicated (\* $p < 0.05$ ,  $\chi^2$  test).

(B) Heat map of the integration frequency relative to genomic features in SupT1 cells. Each column represents a single viral variant, while each row corresponds to the genomic feature under analysis. Tile color indicates whether a particular feature is favored (red) or disfavored (blue) for integration relative to their matched random control (MRC). Asterisks indicate deviations (\* $p < 0.001$ , Wald statistic, compared to  $\chi^2$  distribution) from the HIV-1<sub>NL4-3</sub> WT integration pattern. See also Figure S6.

IN<sub>S119T</sub> and IN<sub>R231G</sub> substitutions with an even larger bias toward minor groove-compressible trinucleotides around the integration site dyad axis.

The IN<sub>T122K</sub> virus is most tolerant of stiff or major groove-compressible sequences, possibly due to the additional interactions offsetting the energetic penalty of tDNA bending.

### HIV-1 IN Variants Display Distinct Global Integration Profiles

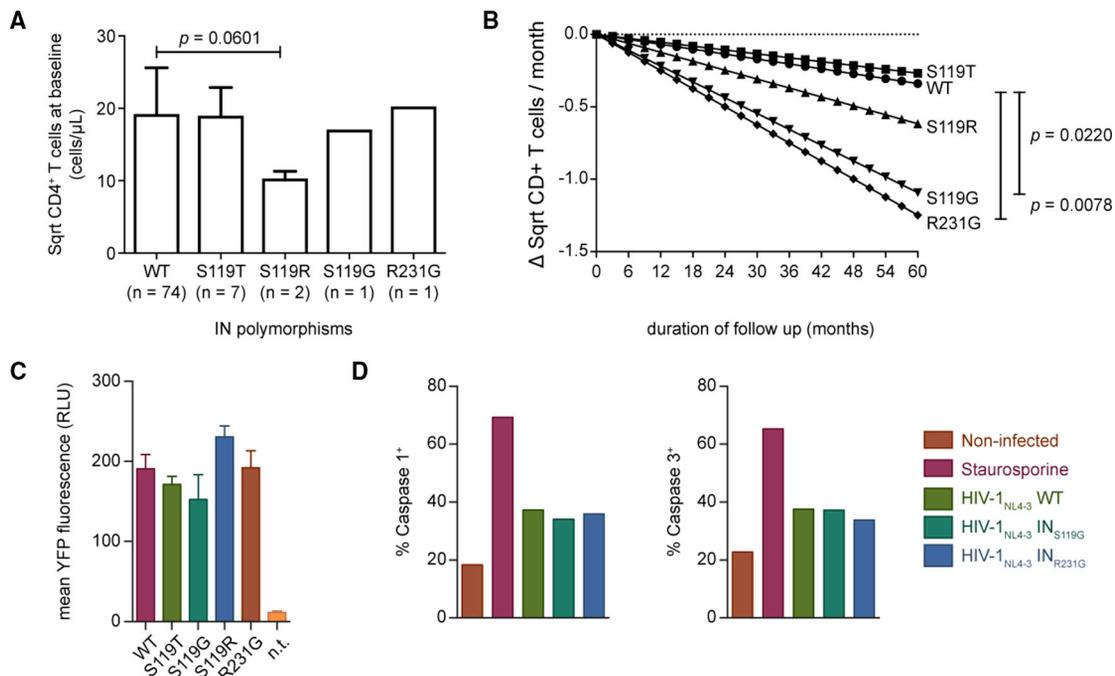
While IN-tDNA base contacts are expected to influence local integration site biases, this is less evident for the global preferences. We hence zoomed out and investigated whether our HIV-1<sub>NL4-3</sub> variants exhibit distinct global integration site profiles. Figure 6B shows heat maps summarizing viral integration preferences relative to diverse genomic features. The IN<sub>S119T/V/K/R</sub> viruses all show a reduced tendency to integrate into GC-rich regions of the genome when small windows (20–100 bp) are considered. This finding is in agreement with their sequence logos, all of which indicated relatively strong biases toward adenine or thymine at positions –3 or –2. Looking at window sizes > 100 bp, this predilection fades, except for the IN<sub>S119V</sub> virus, which shows an integration profile shifting away from

gene-dense regions. IN<sub>S119T/K/R</sub> viruses show global integration profiles comparable to WT. Like IN<sub>S119V</sub>, IN<sub>S119G</sub> virus integrates less into globally gene-dense regions, albeit to a more limited extent, as evidenced, for example, by the smaller reduction in bias for GC content over large window sizes. The integration site profile of the IN<sub>S119A</sub> variant appears intermediate between that of IN<sub>S119G</sub> and WT, while IN<sub>S119P</sub> virus integrates similarly to WT.

Turning to the IN<sub>231</sub> variants, also with regard to the global integration site distribution the IN<sub>R231K</sub> virus is indistinguishable from WT. IN<sub>R231G/E</sub> viruses shift integration away from gene-dense regions as well, with the latter showing the most potent effect. The IN<sub>T122K</sub> virus has a global profile indistinguishable from that of WT.

Finally, HIV-1<sub>NL4-3</sub> IN<sub>S119T-R231G</sub> virus combines the integration profile shifts observed for the separate IN<sub>S119T</sub> and IN<sub>R231G</sub> variants and integrates markedly less in gene-dense regions.

Analysis of integration sites relative to epigenetic features for these viruses (Figure S6A) revealed shifts and preferences corresponding to those observed in the genomic heat map. HIV-1<sub>NL4-3</sub> variants IN<sub>S119T/P/K/R</sub>, IN<sub>T122K</sub>, and IN<sub>R231K</sub> are virtually indistinguishable from WT when looking at their global integration



**Figure 7. Influence of Retargeting Polymorphisms on Disease Characteristics**

(A and B) Influence of IN<sub>119</sub> and IN<sub>231</sub> polymorphisms in a chronic HIV-1 subtype C infection cohort on (A) CD4<sup>+</sup> T cell counts at baseline and (B) the rate of CD4<sup>+</sup> T cell decline. p values are indicated when significant ( $p < 0.05$ ) or almost significant.

(C) Proviral transcriptional activity measured as mean YFP fluorescence intensity of SupT1 cells 5 days postinfection with the indicated HIV-1<sub>NL4-3</sub> IRES-YFP single-round viruses. Data represent averages and SDs from triplicate measurements. n.t., nontransduced.

(D) Percentage caspase-1- or caspase-3-positive PBMCs after breakthrough of the indicated HIV-1<sub>NL4-3</sub> variants. Uninfected and staurosporine-treated PBMCs are included as controls. See also Figure S7.

profiles. On the other hand, IN<sub>S119G/V</sub>, IN<sub>R231G/E</sub>, as well as IN<sub>S119T-R231G</sub> viruses exhibit retargeting to regions with a reduced activation status. IN<sub>S119A</sub> virus again shows a behavior in between that of WT and IN<sub>S119G</sub>. Although we obtained a more limited number of integration sites for the different sets in HeLa P4 cells, the genomic distributions paralleled those observed in SupT1 cells (data not shown).

HIV IN positions 119, 122, and 231 are not expected to alter LEDGF/p75 binding. To confirm that integration of these variants is still globally targeted by LEDGF/p75, we mapped chromosomal binding of LEDGF/p75 by ChIP-seq and determined the enrichment of LEDGF/p75 peaks in the vicinity of integration sites compared to MRC sites. The bottom row in Figure 6B indicates that all viral variants are similarly inclined to integrate into LEDGF/p75-bound regions. To further exclude an effect of the globally retargeting IN substitutions on LEDGF/p75 binding, we performed AlphaScreen protein-protein interaction assays. Titration of recombinant His<sub>6</sub>-tagged IN<sub>WT</sub>, IN<sub>S119G</sub>, or IN<sub>R231G</sub> against FLAG-tagged LEDGF/p75 resulted in comparable binding curves, confirming that the recognition of this cofactor is unaffected by the substitutions (Figure S6B).

In addition, we also evaluated whether S119G or R231G substitutions influence IN multimerization, a key process during viral replication. Employing a previously reported IN multimerization AlphaScreen assay (Demeulemeester et al., 2012), we titrated recombinant His<sub>6</sub>-IN WT or variants against the corresponding

glutathione S-transferase (GST)-tagged IN versions. Both IN<sub>S119G</sub> and IN<sub>R231G</sub> displayed a similar multimerization behavior as WT (Figure S6C).

### IN<sub>S119G</sub> and IN<sub>R231G</sub> Are Associated with Rapid Disease Progression in a Chronic Infection Cohort

We next wondered whether the altered integration site selection of our variants affects disease characteristics. To address this, we looked at the Sinikithemba cohort, which comprises antiretroviral-naive, HIV-1 subtype C chronically infected adults enrolled at the McCord Hospital (Durban, South Africa) from August 2003 to December 2008. We considered a set of 85 participants from the cohort for which IN sequences and longitudinal follow-up data were available. Of these, 74 encoded Ser, 7 Thr, 2 Arg, and only 1 Gly at position IN<sub>119</sub> against a WT IN<sub>R231</sub> background. One viral isolate had an IN<sub>R231G</sub> substitution in a WT IN<sub>S119</sub> context.

Analysis of CD4<sup>+</sup> T cell counts at baseline did not reveal any significant differences among the viral variants, although a trend toward lower CD4<sup>+</sup> T cell counts was observed for IN<sub>S119R</sub> viruses (Figure 7A). Interestingly, both globally retargeting viral variants, IN<sub>S119G</sub> and IN<sub>R231G</sub>, were associated with rapid disease progression as reflected by the higher rate of CD4<sup>+</sup> T cell decline compared to WT virus (IN<sub>S119-R231</sub>, Figure 7B, mixed linear model,  $p = 0.0220$  and  $p = 0.0078$ , respectively).

Considering that the viral fitness and transduction efficiency of these variants are comparable to WT, altered transcription of the

retargeted provirus may be the link with disease progression. To test this hypothesis, we generated HIV-1<sub>NL4-3</sub> single-round viruses carrying a yellow fluorescent protein (YFP) reporter between the *env* and *nef* genes using an internal ribosome entry site (IRES; HIV-1<sub>NL4-3</sub> IRES-YFP). SupT1 and HeLa P4 cells were infected with serial dilutions of HIV-1<sub>NL4-3</sub> IRES-YFP carrying the IN<sub>S119G/R/T</sub> or IN<sub>R231G</sub> variants, and harvested for fluorescence-activated cell sorting (FACS) analysis 5 days post-infection. The mean fluorescence intensity obtained with virus dilutions giving rise to a low percentage of YFP<sup>+</sup> cells (< 20%, single integrated copy) was used as a measure of provirus transcriptional activity. Expression levels did not differ significantly between IN variant viruses (Figures 7C and S7A). Comparable data were obtained after culturing the infected cells for 10 days (data not shown).

Lastly, we asked whether we could detect increased cell death following infection with HIV-1<sub>NL4-3</sub> IN<sub>S119G</sub> or IN<sub>R231G</sub>, which may explain the higher rate of CD4<sup>+</sup> T cell decline in patients. HIV infection of CD4<sup>+</sup> T cells has been reported to induce both caspase-3-mediated apoptosis as well as inflammatory caspase-1-mediated pyroptosis (Doitsh et al., 2014). While apoptosis occurs in a minority of productively infected cells, the majority of quiescent CD4<sup>+</sup> T cells die through pyroptosis triggered by abortive infection. We used fluorescently labeled inhibitor of caspase (FLICA) probes to monitor caspase activation during a spreading infection in PBMCs. At day 7, coinciding with full viral breakthrough, we observed increased activation of both caspase-1 and -3 (Figures 7D and S7B). Neither IN<sub>S119G</sub> nor IN<sub>R231G</sub> virus induced more activation than WT. It is, however, likely that the limited timeframe as well as the in vitro setting are unable to capture the differences in the slow decay of CD4<sup>+</sup> T cells observed in patients.

Despite the limited number of patients with rare IN polymorphisms, these data suggest that globally retargeting IN polymorphisms, such as IN<sub>S119G</sub> and IN<sub>R231G</sub>, could be linked to HIV pathogenesis. Although similar levels of transcriptional activity and cell death were detected for both variants in cell culture, we cannot exclude that these factors contribute to the increased rate of CD4<sup>+</sup> T cell decline in patients over time.

## DISCUSSION

While it is clear how IN-tDNA base contacts can affect local integration site preferences, this is far less the case for the global retargeting effects. Several hypotheses can be formulated. First, there may exist other, as yet unknown, cofactors aside from LEDGF/p75 (or HRP-2; Schrijvers et al., 2012) that contribute to integration site selection. We show that LEDGF/p75 binding is unaffected, but cannot exclude that the globally retargeting IN variants affect binding to other cofactors. Second, reduced IN strand transfer activity may increase the lifetime of the PIC as it will not integrate as quickly. The additional time could allow the PIC to travel further into the nucleus to regions not normally targeted. Such an effect could play a role for the partially defective IN<sub>S119V</sub> and IN<sub>R231E</sub> variants. Third, based on the intasome model and the tDNA flexibility effects, we hypothesize that substitutions IN<sub>S119G/V</sub> and IN<sub>R231G/E</sub> reduce the shape and/or electrostatic compatibility of the intasome with nucleosomal tDNA. As intergenic regions are believed to exhibit

lower nucleosome occupancy than coding regions, this could explain the observed shift toward globally less gene-dense regions.

The prevalence of retargeting IN variants differs among HIV-1 subtypes (Figure S5). Overall, we hence expect slightly different integration site preferences for one subtype when compared to another. Notably, based on the available viral sequences, the globally retargeting IN<sub>S119G</sub> and IN<sub>R231G</sub> variants appear to occur more frequently in subtypes B and C, respectively. As different subtypes are associated with different rates of disease progression (Vasan et al., 2006), it would be interesting to test whether part of this variability can be explained by differences in integration site targeting. Subtype D viruses, for instance, show an overall higher prevalence of globally retargeting variants and are associated with faster progression toward AIDS when compared to subtype A or C (Vasan et al., 2006).

Our findings reveal a number of intriguing interplays. First, HIV integration into gene-dense regions of the genome is hypothesized to afford the provirus better access to transcription factors (Schröder et al., 2002). Targeting integration away from these regions, such as in the IN<sub>119G</sub> and IN<sub>231G</sub> variants, could thus result in altered transcriptional regulation of the provirus. In turn, this may directly affect latency and disease progression, suggesting interaction between the integration neighborhood and virulence or viral evolution. Although we did not detect gross transcriptional or cytotoxicity differences in cell culture, subtle effects may exist in the patient, which accumulate over time.

Second, the IN<sub>S119R</sub> variant was recently described as an uncommon polymorphism selected by the HLA-C\*05 allele (Brockman et al., 2012). Immune escape was suggested to result in a slightly reduced viral replication capacity, which could be rescued by a single compensatory IN<sub>A91E</sub> mutation. Covariation between IN positions 91 and 119 or between 119 and 122 in some HIV-1 subtypes was reported previously (Rhee et al., 2008). Structurally these aa are in relatively close proximity and may indeed influence one another. In the TCC, however, the A91 side chain is directed toward the tDNA phosphodiester backbone and is unlikely to directly alter tDNA base contacts. In any case, IN<sub>119</sub> represents a site of close interaction between HIV-1 immune evasion and integration site targeting.

Third, not only do IN<sub>S119R/G/T</sub> variants differ in frequency among HIV-1 subtypes, they also change prevalence in untreated versus IN strand transfer inhibitor (INSTI)-treated patients, and they have been described as secondary resistance mutations (Ceccherini-Silberstein et al., 2010). Due to the polymorphic nature of this position, however, it is still unclear to which extent IN<sub>119</sub> substitutions contribute to INSTI resistance (Armenia et al., 2012). INSTIs are competitive inhibitors of tDNA binding, and it can be expected that energetically more favorable tDNA contacts at position IN<sub>119</sub> and/or slight changes in the tDNA conformation (as suggested for the IN<sub>S119R/G</sub> variants) can indirectly decrease INSTI inhibition by increasing tDNA affinity. Treatment with suboptimal doses of raltegravir can lead to aberrant HIV-1 integration where a single or both end(s) of the provirus are associated with duplications, inversions, deletions, and insertions (Varadarajan et al., 2013). Our results suggest that INSTI treatment, by selecting for certain IN<sub>119</sub> polymorphisms, can influence the viral integration site indirectly as well.

It is important to note that while we propose IN<sub>119</sub> and IN<sub>231</sub> to be the sole direct tDNA base contacts in HIV-1, other positions may exist which indirectly affect target site selection. The identification and interpretation of aa with such subtle effects is considerably more involved, however. For example, a mechanistic explanation has yet to be put forward for variation in the size of the insertion stagger. In theory, the aa involved could equally well be polymorphic or under immune selection and may hence represent further links with integration site selection. The highly polymorphic A124 and T125 in HIV-1 IN, for example, are part of the LEDGF/p75 binding site. These residues could alter target site selection indirectly by affecting cofactor binding.

In conclusion, our findings reveal how polymorphisms at positions corresponding to HIV IN<sub>119</sub> and IN<sub>231</sub> affect local as well as global integration site targeting. Intriguingly, these findings provide a number of links between integration site selection on the one hand and immune evasion, antiretroviral therapy, virulence, and disease progression on the other.

## EXPERIMENTAL PROCEDURES

### Plasmids and Mutagenesis

The viral molecular clone pNL4-3 and reporter construct pNL4-3.Luc.R<sup>+</sup>E<sup>-</sup> were obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH. Details on the mutagenesis are included in the [Supplemental Information](#).

### Recombinant Protein Purification

N-terminally GST-tagged and C-terminally His<sub>6</sub>-tagged IN WT and mutants were produced and purified as described previously ([Busschots et al., 2005](#); [Demeulemeester et al., 2012](#)). Purification of triple-FLAG-tagged LEDGF/p75 is outlined in the [Supplemental Information](#).

### AlphaScreen Protein-Protein Interaction Assay

The IN multimerization assay was performed essentially as reported before ([Demeulemeester et al., 2012](#)). Details on the IN-LEDGF/p75 AlphaScreen are presented in the [Supplemental Information](#).

### Integration Site Sequencing

Linkers were ligated to MseI-digested genomic DNA isolated from infected SupT1 and HeLa P4 cells, and virus-host DNA junctions were amplified by nested PCR. Samples were barcoded through the second pair of PCR primers in order to generate 454 libraries. PCR products were purified and sequenced using 454/Roche pyrosequencing (Titanium Technology). Reads were quality filtered by requiring perfect matches to the long terminal repeat linker and barcode and subsequently mapped to the human genome. Details are included in the [Supplemental Information](#).

### Virus Production and Transduction

Virus production is described in the [Supplemental Information](#). Lentiviral transduction (HIV-1<sub>NL4-3</sub> fLuc, HIV-1<sub>NL4-3</sub> IRES-YFP) and infection (HIV-1<sub>NL4-3</sub>) experiments were performed as reported previously ([Schrijvers et al., 2012](#)).

### Viral Replication and Caspase Assay

For the breakthrough experiments 10<sup>5</sup> MT-4 cells or 10<sup>6</sup> PBMCs were infected with virus amounting to 0.2 or 20 pg p24/mL, respectively. Starting 2 or 3 days postinfection, supernatants were sampled daily, and viral replication was assessed using p24 ELISA.

Caspase-1/3 activation was monitored at selected time points using FLICA probes (ImmunoChemistry Technologies) followed by FACS analysis. Staurosporine (1 μM) was used as a positive control.

### Sinikithemba Cohort Analysis

The Sinikithemba cohort comprises 450 antiretroviral naive, HIV-1 subtype C chronically infected adults. Sociodemographic characteristics, plasma viral

load, and CD4<sup>+</sup> T cell count measurements were obtained at baseline. CD4<sup>+</sup> T cell counts and viral loads were measured every 3 and 6 months, respectively, starting from enrollment. Viral loads were determined using the automated Cobas Amplicor HIV-1 Monitor test (v1.5, Roche Diagnostics). CD4<sup>+</sup> T cells were enumerated using the Multitest Kit (CD4/CD3/CD8/CD45) on a FACSCalibur flow cytometer (Becton Dickinson).

Mann-Whitney *U* tests were used to assess the association of polymorphisms with baseline CD4<sup>+</sup> counts. Linear mixed models were employed to estimate the rate of CD4<sup>+</sup> T cell decline. Statistical analyses were done using SAS v9.3 (SAS Institute) and Prism 5 (GraphPad). Results where *p* < 0.05 were considered significant.

## Homology Modeling and Bioinformatics

Structural work on the PFV intasome and homology modeling of the HIV-1 intasome were performed with Chemical Computing Group's Molecular Operating Environment 2012.10. Sequence alignments were generated using MSAProbs and visualized and manually realigned using Jalview. Sequence logos were created using WebLogo 3.3 with compositional adjustment for the human genome base background distribution. Statistical analyses on base and trinucleotide distributions were performed using R-3.0.1.

## ACCESSION NUMBERS

The GEO accession number for the LEDGF/p75 ChIP-Seq reported in this paper is GSE61003.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, two tables, and Supplemental Experimental Procedures and can be found with this article online at <http://dx.doi.org/10.1016/j.chom.2014.09.016>.

## AUTHOR CONTRIBUTIONS

J.D., S.V., and R.S. performed the experiments. J.D., J.D.R., Z.D., and R.G. designed experiments and analyzed the data. J.D. and R.G. wrote the manuscript. T.N. and P.M. were responsible for patient isolate analysis and M.D.M. for the protein modeling. All authors read, corrected, and approved the final manuscript.

## ACKNOWLEDGMENTS

We thank Paulien Van De Velde, Barbara van Remoortel, and Nam Joo Van der Veken for excellent technical assistance. We are also grateful to Nirav Malani and Frederic D. Bushman for the use of the Integration Site Pipeline and Database, to Nonhlanhla Yende for the statistical work on the clinical data, to Madapura M. Pradeepa for the ChIP, and to Jonathan Crowther for critical reading of the manuscript. We acknowledge the participants and their clinicians who took part in the Sinikithemba cohort. J.D. and R.S. were doctoral fellows for the Research Foundation - Flanders (FWO). Research was funded by grants from the KU Leuven Research Fund, FWO, NIH-CNIHR, IAP BelVir (BelSpo), FP7 CHAARM, and the Eranet EUREKA. The Sinikithemba cohort was supported by the NIH (grant R01-AI067073, contract N01-AI-15422).

Received: May 16, 2014

Revised: July 31, 2014

Accepted: September 15, 2014

Published: November 12, 2014

## REFERENCES

Armenia, D., Vandenbroucke, I., Fabeni, L., Van Marck, H., Cento, V., D'Arrigo, R., Van Wesenbeeck, L., Scopelliti, F., Micheli, V., Bruzzone, B., et al. (2012). Study of genotypic and phenotypic HIV-1 dynamics of integrase mutations during raltegravir treatment: a refined analysis by ultra-deep 454 pyrosequencing. *J. Infect. Dis.* 205, 557–567.

- Brockman, M.A., Chopera, D.R., Olvera, A., Brumme, C.J., Sela, J., Markle, T.J., Martin, E., Carlson, J.M., Le, A.Q., McGovern, R., et al. (2012). Uncommon pathways of immune escape attenuate HIV-1 integrase replication capacity. *J. Virol.* *86*, 6913–6923.
- 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.
- Ceccherini-Silberstein, F., Malet, I., Fabeni, L., Dimonte, S., Svicher, V., D'Arrigo, R., Artese, A., Costa, G., Bono, S., Alcaro, S., et al. (2010). Specific HIV-1 integrase polymorphisms change their prevalence in untreated versus antiretroviral-treated HIV-1-infected patients, all naive to integrase inhibitors. *J. Antimicrob. Chemother.* *65*, 2305–2318.
- 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.
- De Rijck, J., de Kogel, C., Demeulemeester, J., Vets, S., El Ashkar, S., Malani, N., Bushman, F.D., Landuyt, B., Husson, S.J., Busschots, K., et al. (2013). The BET family of proteins targets moloney murine leukemia virus integration near transcription start sites. *Cell Rep.* *5*, 886–894.
- Demeulemeester, J., Tintori, C., Botta, M., Debyser, Z., and Christ, F. (2012). Development of an AlphaScreen-based HIV-1 integrase dimerization assay for discovery of novel allosteric inhibitors. *J. Biomol. Screen.* *17*, 618–628.
- Doitsh, G., Galloway, N.L.K., Geng, X., Yang, Z., Monroe, K.M., Zepeda, O., Hunt, P.W., Hatano, H., Sowinski, S., Muñoz-Arias, I., and Greene, W.C. (2014). Cell death by pyroptosis drives CD4 T-cell depletion in HIV-1 infection. *Nature* *505*, 509–514.
- Gabrielian, A., and Pongor, S. (1996). Correlation of intrinsic DNA curvature with DNA property periodicity. *FEBS Lett.* *393*, 65–68.
- Gupta, S.S., Maetzig, T., Maertens, G.N., Sharif, A., Rothe, M., Weidner-Glunde, M., Galla, M., Schambach, A., Cherepanov, P., and Schulz, T.F. (2013). Bromo- and extraterminal domain chromatin regulators serve as cofactors for murine leukemia virus integration. *J. Virol.* *87*, 12721–12736.
- Harper, A.L., Skinner, L.M., Sudol, M., and Katzman, M. (2001). Use of patient-derived human immunodeficiency virus type 1 integrases to identify a protein residue that affects target site selection. *J. Virol.* *75*, 7756–7762.
- Luscombe, N.M., Laskowski, R.A., and Thornton, J.M. (2001). Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.* *29*, 2860–2874.
- Maertens, G.N., Hare, S., and Cherepanov, P. (2010). The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* *468*, 326–329.
- Rhee, S.-Y., Liu, T.F., Kiuchi, M., Zioni, R., Gifford, R.J., Holmes, S.P., and Shafer, R.W. (2008). Natural variation of HIV-1 group M integrase: implications for a new class of antiretroviral inhibitors. *Retrovirology* *5*, 74.
- Satchwell, S.C., Drew, H.R., and Travers, A.A. (1986). Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* *191*, 659–675.
- Schrijvers, R., De Rijck, J., Demeulemeester, J., Adachi, N., Vets, S., Ronen, K., Christ, F., Bushman, F.D., Debyser, Z., and Gijssbers, R. (2012). LEDGF/p75-independent HIV-1 replication demonstrates a role for HRP-2 and remains sensitive to inhibition by LEDGfIns. *PLoS Pathog.* *8*, e1002558.
- Schröder, A.R.W., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* *110*, 521–529.
- Serrao, E., Krishnan, L., Shun, M.-C., Li, X., Cherepanov, P., Engelman, A., and Maertens, G.N. (2014). Integrase residues that determine nucleotide preferences at sites of HIV-1 integration: implications for the mechanism of target DNA binding. *Nucleic Acids Res.* *42*, 5164–5176.
- 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.
- Varadarajan, J., McWilliams, M.J., and Hughes, S.H. (2013). Treatment with suboptimal doses of raltegravir leads to aberrant HIV-1 integrations. *Proc. Natl. Acad. Sci. USA* *110*, 14747–14752.
- Vasan, A., Renjifo, B., Hertzmark, E., Chaplin, B., Msamanga, G., Essex, M., Fawzi, W., and Hunter, D. (2006). Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype. *Clin. Infect. Dis.* *42*, 843–852.