HIV-1 integrase variants shift viral integration and associate with disease progression in a Chronic Infection Cohort

Jonas Demeulemeester[1,2], Sofie Vets[1], Rik Schrijvers[1], Paradise Madlala[1,3], Marc De Maeyer[2], Frauke Christ[1], Jan De Rijck[1], Thumbi Ndung'u[3–6], Zeger Debyser[1,*], Rik Gijsbers[1,*]

[1]Molecular Virology and Gene Therapy, Department of Pharmaceutical and Pharmacological Sciences; [2]Biomolecular Modeling, Department of Chemistry, KU Leuven – University of Leuven, 3000 Leuven, Belgium
[3]HIV Pathogenesis, Doris Duke Medical Research Institute; [4]KwaZulu-Natal Research Institute for Tuberculosis and HIV Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, 4013 Durban, South Africa
[5]Ragon Institute of MGH, MIT and Harvard, Cambridge, MA 02139, USA
[6]Max Planck Institute for Infection Biology, Chariteplatz, 10117 Berlin, Germany
*Shared senior author: correspondence: zeger.debyser@med.kuleuven.be (Z.D.) or rik.gijsbers@med.kuleuven.be (R.G.)

Background: Distinct integration patterns of different retroviruses, including HIV-1, have puzzled virologists for over 20 years. While we now know that these patterns are largely due to alternative tethering factors being co-opted by retroviral subfamilies (e.g. LEDGF/p75 by lentiviruses and Bromodomain and extra-terminal domain –BET– proteins by γ-retroviruses), the virally encoded integrase (IN) has the final say.
A tetramer of the viral integrase (IN) assembles on the two viral cDNA ends, docks onto the target DNA (tDNA) to form the target capture complex (TCC) and catalyzes viral genome insertion into the host chromatin.
Methods: We combined structural information on the Prototype Foamy Virus TCC with conservation in retroviral IN protein alignments to determine aa-tDNA base contacts (**Fig 1A**). We generated HIV-1 variants based on the observed variability at these positions (**Fig 1B, 4A–B**), assessed replication capacities (**Fig 2, 4C–E**) and performed integration site sequencing to reveal their local (**Fig 3, 5**) as well as their global (**Fig 6**) integration preferences. Finally, we examined their effect on disease progression in a chronic HIV-1 subtype C infection cohort (**Fig 7**).
Results: We identified retroviral IN amino acids affecting molecular recognition in the TCC (**Fig 1**). Altered molecular recognition in the TCC underlies distinct local tDNA nucleotide biases (**Fig 3, 5**). These residues also determine the propensity of the virus to integrate into flexible tDNA sequences (**Fig 6A**). Remarkably, natural polymorphisms IN_{S119G} and IN_{R231G} retarget viral integration away from gene dense regions (**Fig 6B**). Precisely these variants were associated with rapid disease progression in a chronic HIV-1 subtype C infection cohort (**Fig 7**).
Conclusions: Our findings reveal how polymorphisms at positions corresponding to HIV IN_{119} and IN_{231} affect local as well as global integration site targeting. Intriguingly, 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_{119} is under selection by HLA alleles and integrase inhibitors.

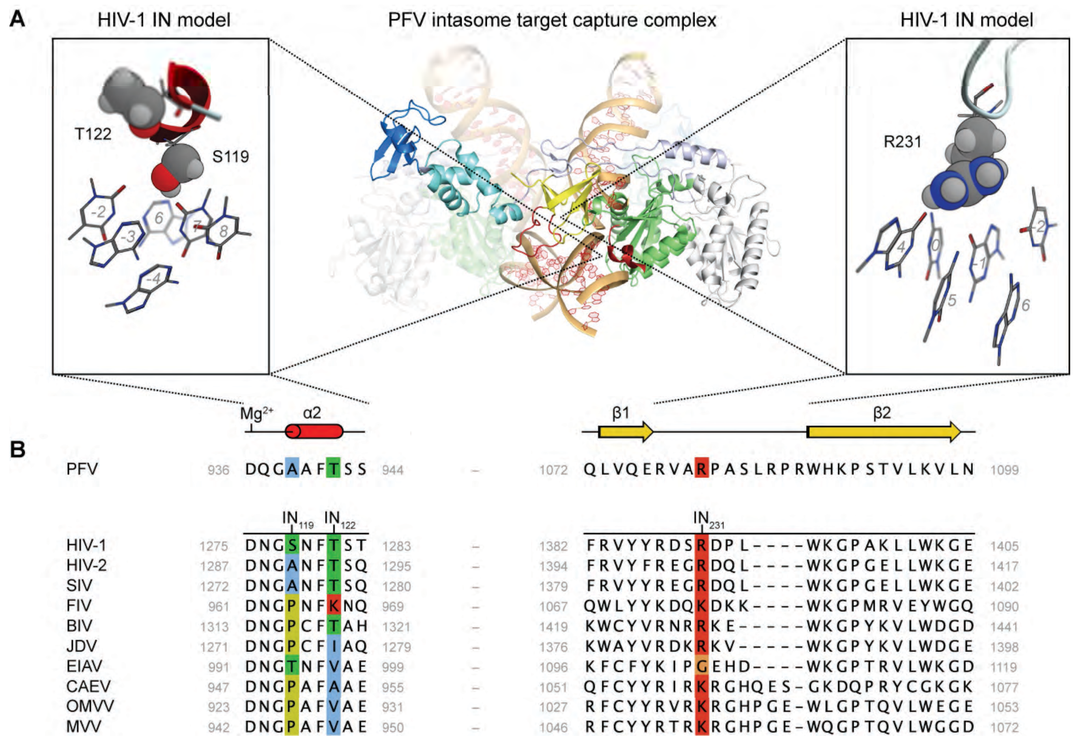


Figure 1. IN Residues Contacting tDNA Bases
(A) Cartoon of the PFV intasome target capture complex. Boxes on either side zoom in on the aa-tDNA base contacts in the HIV-1 intasome model. Only the IN CCD α2 helix containing residues S119 and T122, and part of the CTD β1/β2 loop harboring R231, are shown. Proximal tDNA bases are depicted and numbered starting from the sites of strand transfer on both strands.
(B) The IN CCD α2 helix and CTD β1/β2 regions from a lentiviral sequence alignment are matched with the PFV IN sequence. Numbering refers to the (Gag)-Pol polyprotein context. Secondary structures and Mg2+ cofactor binding are indicated above, and HIV-1 positions 119, 122, and 231 below, the PFV sequence. Accession: PFV (P14350); HIV-1 (P03367); HIV-2 (P04584); SIV, simian immunodeficiency virus (Q8AII1); FIV, feline immunodeficiency virus (P31822); BIV, bovine immunodeficiency virus (P19560); JDV, Jembrana disease virus (Q82851); EIAV, equine infectious anemia virus (P32542); CAEV, caprine arthritis encephalitis virus (P33459); OMVV, ovine maedi visna-related virus (P16901); MVV, maedi visna virus (P03370).

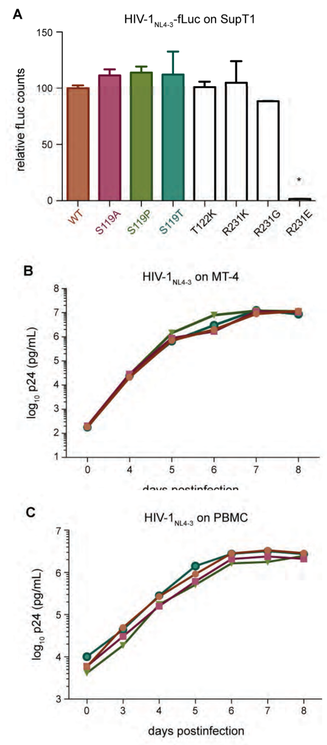


Figure 2. Replication of tDNA Contact Variant Viruses
(A) Transduction efficiency as measured by fLuc reporter activity (relative light units, RLUs/mg protein) 3 days postinfection of SupT1 cells for the indicated HIV-1_{NL4-3}-fLuc carrying IN mutants. Activity is reported as percentage of HIV-1_{NL4-3}-fLuc 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_{NL4-3} WT and $IN_{S119A/P/T}$ viruses (colored as in A) in MT-4 cells (B) and PBMCs (C).

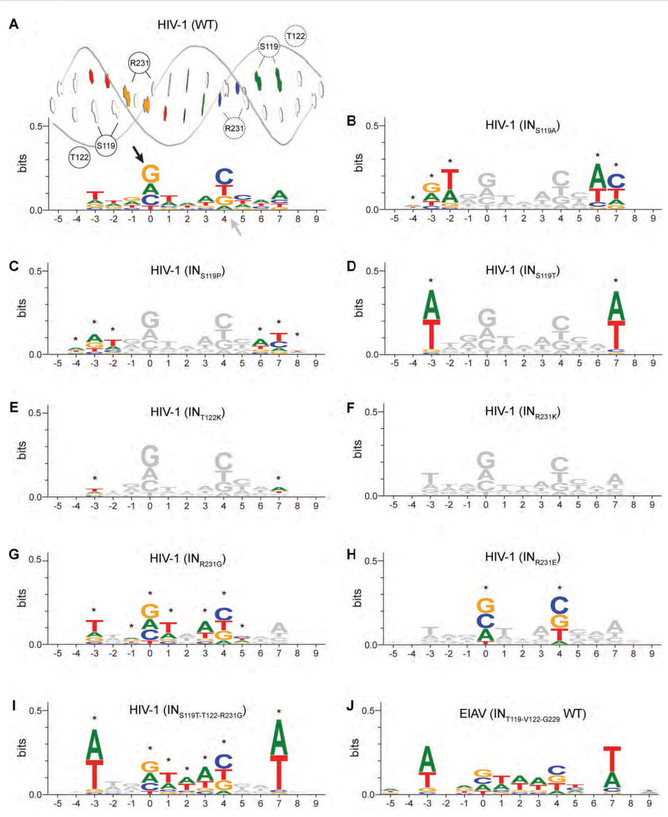


Figure 3. Integration Site Sequence Logos of HIV-1_{NL4-3} IN Variants
Sequence conservation is indicated as the total height of each stack (measured in bits), while the relative height of bases in a stack reflects base frequencies at that position. Arrows on the HIV-1_{NL4-3} WT logo (A) indicate the 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_{NL4-3} variants. Positions showing an altered base distribution compared to WT HIV-1_{NL4-3} are colored (*$p < 10$ 4, $\chi 2$ test).
(J) Integration site sequence logo for EIAV.

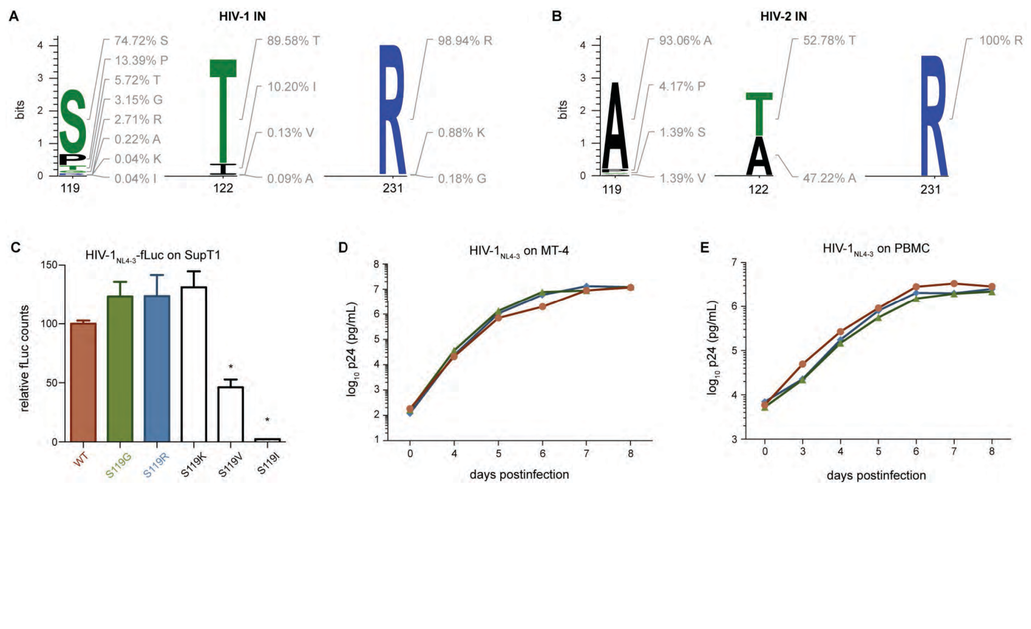


Figure 4. HIV IN tDNA Contacts Variants and Their Replication
(A and B) Sequence conservation (indicated in bits) and aa frequencies of (A) 2,276 aligned HIV-1 and (B) 72 HIV-2 sequences at positions IN_{119}, IN_{122}, and IN_{231}. (C) Transduction efficiency as measured by fLuc reporter activity (relative light units, RLUs/mg protein) 3 days postinfection of SupT1 cells for the indicated HIV-1_{NL4-3}-fLuc single-round viruses. fLuc reporter activity is reported as percentage of HIV-1_{NL4-3}-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_{NL4-3} WT and $IN_{S119G/R}$ viruses, colored as in (C), on MT-4 cells (D) and PBMCs (E).

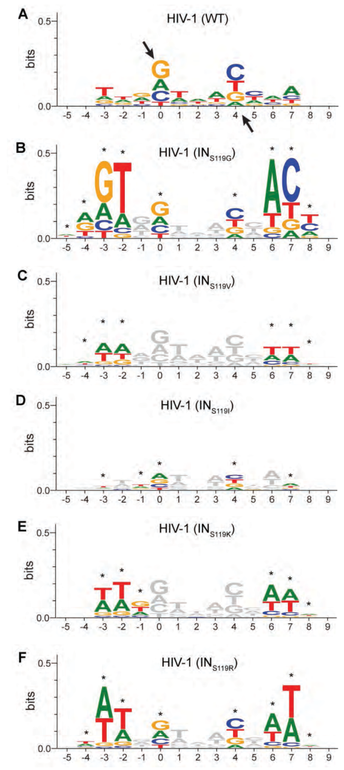


Figure 5. Integration Site Sequence Logos of HIV-1_{NL4-3} IN Variants
(A–F) Local integration sites obtained for (A) HIV-1_{NL4-3} WT and variants (B) IN_{S119G}, (C) IN_{S119V}, (D) IN_{S119I}, (E) IN_{S119K}, and (F) IN_{S119R}. Positions showing an altered base distribution compared to WT HIV-1_{NL4-3} (A) are colored (*$p < 10$ 4, $\chi 2$ test).

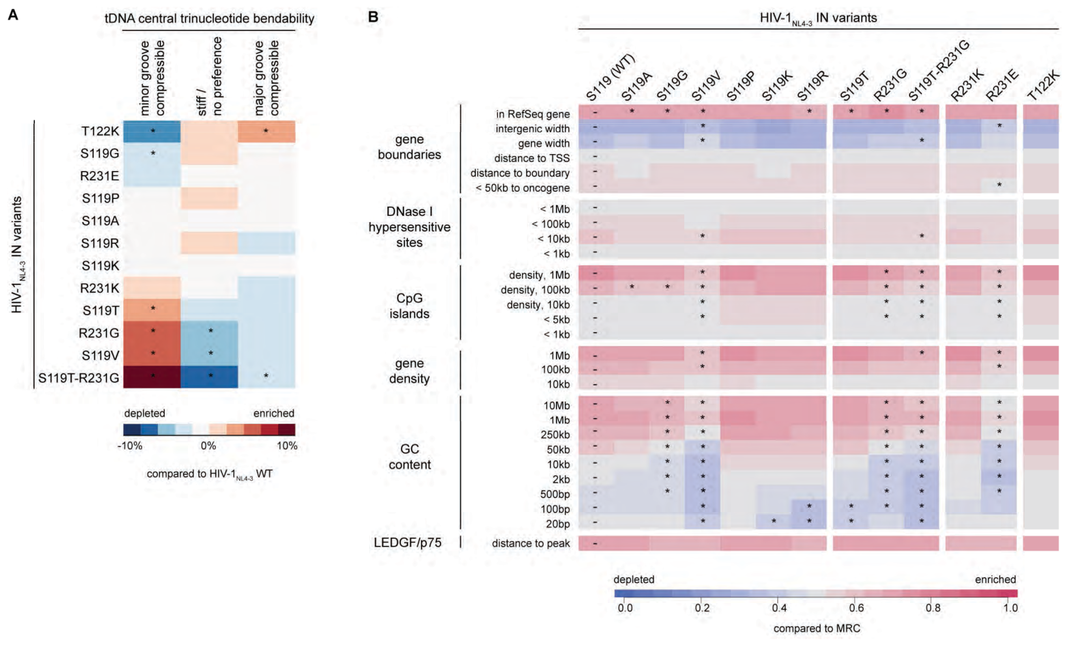


Figure 6. tDNA Bendability and Genomic Integration Preferences
(A) Integration site central trinucleotide (positions 1–3) bendability heat map. Depicted are the changes in frequency of occurrence relative to WT HIV-1_{NL4-3} for each of three groups of trinucleotides: those preferring minor groove compression, those that are stiff/show no preference, or those favoring major groove compression. Statistically significant deviations from the WT HIV-1_{NL4-3} 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_{NL4-3} WT integration pattern.

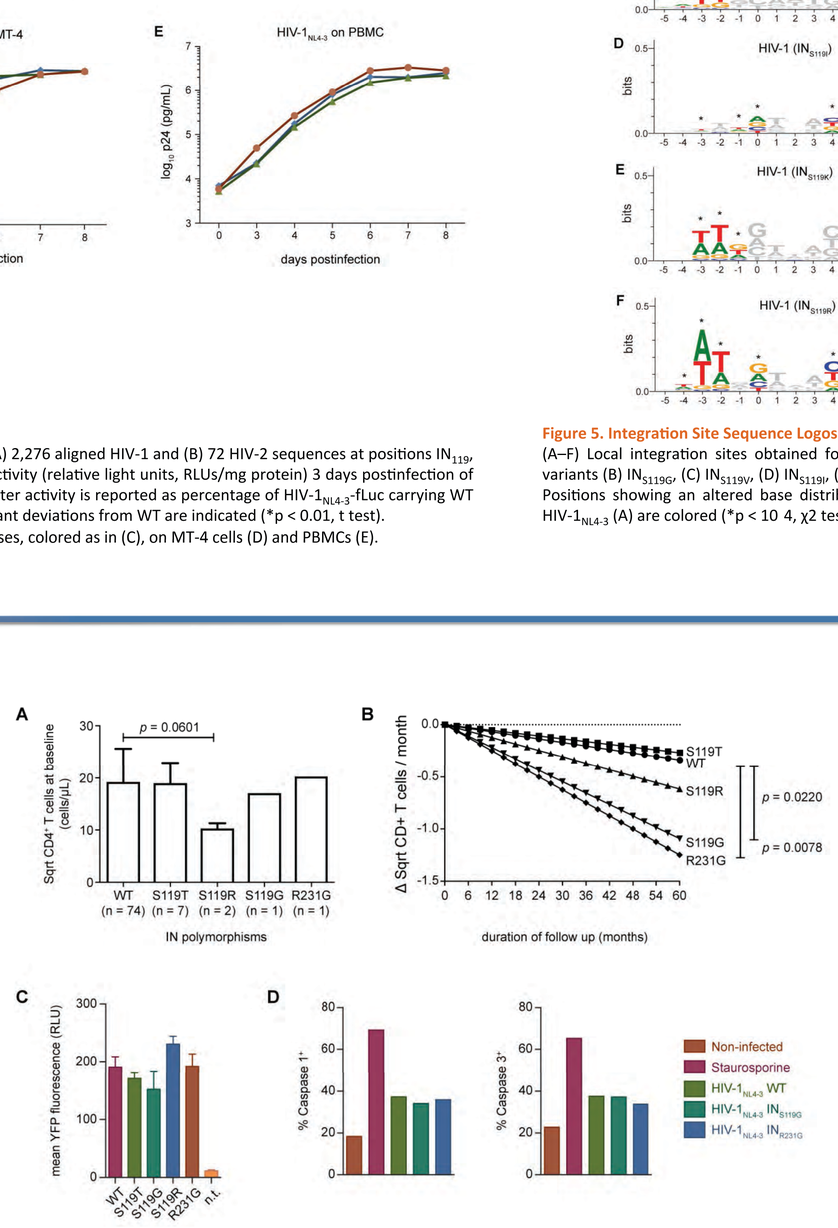


Figure 7. Influence of Retargeting Polymorphisms on Disease Characteristics
(A and B) Influence of IN_{119} and IN_{231} polymorphisms in a chronic HIV-1 subtype C infection cohort on (A) CD4+ T cell counts at baseline and (B) the rate of CD4+ T cell decline. p values are indicated when significant ($p < 0.05$) or almost significant.
(C) Proviral transcriptional activity measured as mean YFP fluorescence intensity of SupT1 cells 5 days postinfection with the indicated HIV-1_{NL4-3} 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_{NL4-3} variants. Uninfected and staurosporine-treated PBMCs are included as controls.

Ciuffi, A. et al. (2005). A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med*, *11*(12), 1287–1289.
De Rijck, J., de Kogel, C. et al. (2013). The BET family of proteins targets moloney murine leukemia virus integration near transcription start sites. *Cell Reports*, *5*(4), 886–894.
Sharma, A. et al. (2013). BET proteins promote efficient murine leukemia virus integration at transcription start sites. *Proc Natl Acad Sci USA, 110*(29), 12036–12041.
Gupta, S. S. et al. (2013). Bromo- and extraterminal domain chromatin regulators serve as cofactors for murine leukemia virus integration. *J Virol*, *87*(23), 12721–12736.
Demeulemeester, J. et al. (2014). HIV-1 integrase variants retarget viral integration and are associated with disease progression in a chronic infection cohort. *Cell Host & Microbe*, *16*(5), 651–662.

FWO Vlaanderen – Fonds Wetenschappelijk Onderzoek – Research Foundation – Flanders
CHAARM – Combined Highly Active Anti-Retroviral Microbicides
CNHR
NIH – National Institutes of Health
BELVIR
KU LEUVEN