Title

Integrase-independent HIV-1 infection is augmented under conditions of DNA damage and produces a viral reservoir

Hirotaka Ebina*, Yuka Kanemura, Yasutsugu Suzuki, Kozue Urata, Naoko Misawa and Yoshio Koyanagi

Laboratory of Viral Pathogenesis, Institute for Virus Research, Kyoto University, 53 Shogoin-kawara-cho, Sakyo-ku, Kyoto 606-8507, Japan

* Corresponding author

Laboratory of Viral Pathogenesis, Institute for Virus Research, Kyoto University, 53 Shogoin-kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan

Phone: +81-75-751-4813

Fax: +81-75-751-4812

E-mail: hebina@virus.kyoto-u.ac.jp

17 Abstract

18 HIV-1 possesses a viral protein, integrase (IN), which is necessary
19 for its efficient integration in target cells. However, it has been reported
20 that an IN-defective HIV strain is still capable of integration. Here, we
21 assessed the ability of wild type (WT) HIV-1 to establish infection in the
22 presence of IN inhibitors. We observed a low, yet clear infection of
23 inhibitor-incubated cells infected with WT HIV which was identical to cells
24 infected with IN-deficient HIV, D64A. Furthermore, the IN-independent
25 integration could be enhanced by the pretreatment of cells with
26 DNA-damaging agents suggesting that integration is mediated by a DNA
27 repair system. Moreover, significantly faster viral replication kinetics with
28 augmented viral DNA integration was observed after infection in irradiated
29 cells treated with IN inhibitor compared to nonirradiated cells. Altogether,
30 our results suggest that HIV DNA has integration potential in the presence
31 of an IN inhibitor and may serve as a virus reservoir.

32

33 Keywords

34 HIV-1, integration, integrase inhibitor, provirus, HIV reservoir, DNA repair

35

Introduction

Retroelements, such as long terminal repeat (LTR)-retrotoransposons, non-LTR-retrotransposons, and retroviruses, insert their reverse transcribed cDNA into the host chromosome during viral replication. To carry out efficient replication, the retrovirus family has robust integration machinery consisting of the retrovirus integrase (IN), which executes the insertion of viral cDNA into the genome of host cells. The IN protein consists of three distinct domains (Craigie, 2001): the N-terminal domain contains an HHCC motif; the catalytic core domain in the center of INs possesses a DDE motif that mediates catalysis; and the C-terminal domain of IN has little sequence conservation yet possesses nonspecific DNA binding activity.

It has been reported that human immunodeficiency virus type 1 (HIV-1) cDNA is preferentially inserted into the gene coding region of the host genome (Schroder et al., 2002). The host lens epithelium-derived growth

52 factor/p75 (LEDGF/p75) protein directly binds IN (Llano et al., 2004;
53 Maertens et al., 2003) and recruits a pre-integration complex consisting of
54 viral cDNA, host proteins and viral proteins to the gene coding region
55 (Cherepanov et al., 2005; Engelman and Cherepanov, 2008; Shun et al.,
56 2007). The genome position of inserted HIV-1 provirus is thought to
57 determine the magnitude of viral gene expression. In fact, transcription of
58 provirus integrated into gene coding regions is extremely high and produces
59 large amounts of viral particles (Wang et al., 2007). In contrast, the
60 transcription of proviruses inserted outside of the gene coding region is
61 relatively dormant and has a potential for persistent and latent infection
62 (Brady et al., 2009; Skupsky et al., 2010).

63 INs dramatically increase the efficiency of viral nucleic acid
64 insertion into the host DNA, however, other mechanisms for exogenous
65 nucleic acid incorporation may also exist. Accumulating lines of evidence
66 indicate that foreign nucleic acids, not related retroelements, can be inserted
67 into the genome of host cells. For instance, transfected plasmid DNA (Suzuki

et al., 2010), a DNA virus genome (Dall et al., 2008), mitochondorial DNA fragments (Nitz et al., 2004) and borna virus cDNA (Horie et al., 2010) have been reported to integrate into host chromatin DNA despite the absence of an IN enzyme. The host DNA repair system which involves homologous recombination or non-homologous end joining appears to be involved in the insertion (Horie and Tomonaga, 2011).

In this study, we first tested the integration of HIV with a mutated IN lacking catalytic activity. We could detect a low but significant amount of integrated HIV cDNA. Because integration is essential for HIV-1 replication, IN inhibitors have been developed and used as an antiviral therapy for HIV (Summa et al., 2008). We tested the illegitimate integration of WT HIV in the presence of IN inhibitors. A low frequency of integration was observed in the presence of an IN inhibitor which capable of producing infectious virus particles, and the level of integration was clearly enhanced under DNA-damaged condition. Our results suggest that retroviral cDNA is inserted into the host chromosome through host DNA repair machinery via

84 an IN-independent pathway and serves as a virus reservoir.

Results

Generation of provirus and stable expression of HIV-1 in the absence of IN activity

We initially used a vesicular stomatitis virus (VSV)-G pseudotyped EGFP-expressing HIV vector packaged with a catalytically inactive HIV-1 IN, which contains a D64V mutation. Jurkat cells were infected with varying amounts of the D64V mutant virus corresponding to the amount of p24CA used at a multiplicity of infection (MOI) of 1, 5 and 10 of wt virus. Cells were also infected with IN-proficient WT virus at an MOI ranging from 1-10 in the absence or presence of IN inhibitors, Elvitegravir (Elv) or Raltegravir (Ral). To ensure the removal of unintegrated HIV-1 DNA associated with cell division, the cells were further cultured for 2 weeks and the levels of unintegrated, 2LTR, and integrated HIV-1 DNA were assessed. As shown Fig.1A, the increased integrated but not unintegrated viral DNA was

101 detected with increasing MOI in WT+Elv- and D64V-infected cells. Based on
102 this result, we calculated the efficiency of integration during an IN-deficient
103 condition to be 0.1-0.2% of WT integration (Fig.1B). We further assessed
104 green fluorescent protein (GFP) transduction efficiency by flow cytometry
105 analysis and observed a dose dependent GFP transduction under
106 IN-deficient conditions, in WT+Elv or WT+Ral, and D64V-infected cells
107 (Fig.1C). Up to 2.3% GFP positive cells were detected in D64V mutant
108 infected cells. The copy number of integrated viral DNA highly correlated
109 with the transduction efficiency ($R=0.9019 > 0.590$; $\alpha=0.01$) (Fig.1D). These
110 results indicated that the transduction was from the integrated form of
111 HIV-1 DNA. Therefore, we assumed that the level of GFP expression 2 weeks
112 after infection using this HIV vector was representative of the level of
113 integrated DNA.

114 It has been shown that the introduction of a
115 DNA double-stranded break in a target gene can stimulate
116 retrotransposition of LINE-1, gene targeting and genome rearrangement

117 (Francis and Richardson, 2007; Morrish et al., 2002; Richardson and Jasin,
118 2000). To test whether DNA damage augments integration activity of HIV-1
119 lacking IN activity, cells were exposed to various doses of gamma irradiation
120 in order to induce DNA double-strand breaks (DSB) before virus infection
121 and then flow cytometry analysis was performed two weeks post infection
122 (Fig. 1E). The percentage of GFP positive cells was clearly augmented and
123 correlated with increasing doses of gamma irradiation. Furthermore, we also
124 examined the effect of chemical-induced DNA damage by hydrogen peroxide
125 in IN-deficient HIV-1 integration. Significant enhancement of GFP
126 transduction of IN-deficient HIV was observed in cells treated with hydrogen
127 peroxide in a dose dependent manner (Fig. 1F). Finally, we directly analyzed
128 the copy number of integrated HIV-1 DNA in the cells used in Fig. 1E and F.
129 (Fig. 1G and H, respectively). The efficiency of integration under IN-deficient
130 conditions was augmented from 0.1% to 1.7% depending on the dose of DNA
131 damage. These results suggested that the induction of DNA damage in target
132 cells enhanced the efficiency of retroviral IN-independent integration.

133 Moreover these findings suggest that HIV uses a host DNA repair system for
134 the IN-independent integration.
135
136 *Attenuation of HIV-1 gene expression from proviruses established through*
137 *an IN-independent pathway*
138
139 We showed that HIV-1 DNA was inserted into the host chromosome
140 without IN activity. To examine the LTR promoter activity and level of gene
141 expression from IN-independent proviruses, VSV-G pseudotyped LTIG
142 (LTR-Tat-IRES-GFP) vector was used. The HIV tat protein, an accessory
143 protein responsible for regulating HIV transcription, is expressed under the
144 regulation of an LTR promoter and the transcriptional level can be
145 monitored by GFP expression. As expected, a parallel transduction of Jurkat
146 cells with GFP-expressing WT virus in the presence of Elv and D64V mutant
147 virus in increasing doses was observed (Fig. 2A and B). However, the mean
148 fluorescence intensity (MFI) was significantly lower in cells infected with

149 LTIG virus in the presence of Elv and D64V mutant compared with that of
150 WT virus without Elv (Fig. 2A and C). Furthermore, we isolated 17, 10 and
151 20 cell clones from WT virus infected cells without IN inhibitor, WT virus
152 infected cells with inhibitor, and D64V virus infected cells, respectively. The
153 level of GFP expression in each clone was analyzed by flow cytometry as
154 shown in Fig.2D. The isolated cell clones were divided into two groups,
155 IN-dependent (WT) and IN-indendent (WT+Elv and D64V) transduction.
156 Then, the frequency of distribution based on the MFI of GFP expression is
157 shown in Fig. 2E. Chi-square distribution was assessed and statistically
158 significant difference between IN-dependent and independent groups was
159 observed ($\chi 2=10.927>9.488$; four-degree-of-freedom, $P=0.05$).

160 Given that the chromatin environment near the provirus is known
161 to affect the level of viral gene expression, we attempted to address this
162 possibility by analyzing the integration sites of proviruses (Table 1). In
163 WT-infected cells, 84% of the integration sites were detected in gene coding
164 regions as shown RefSeq, while only 68% of the events were detected in gene

coding regions under IN-deficient condition. Although, we could not find statistically significant differences (P=0.098) under these parameters, if we analyzed the frequency of viral DNA insertion in respect to the presence of repeat sequences, then significant differences were observed (2% vs. 19%; P=0.0048). Furthermore, the deletion and insertion of nucleotides in the junction of LTR-genomic DNA was a frequent feature in the IN-deficient mutant as shown in a previous report (6% vs. 27%; P=0.0039) (Gaur and Leavitt, 1998). These results suggest that the HIV integration pattern is modified under IN-deficient conditions and it may cause reduced promoter activity.

Formation of an intact HIV-1 reservoir under IN-deficient conditions

Here, we showed a lower level of viral gene expression from the provirus generated through an IN-independent pathway. To test the ability of replication-competent HIV-1 to produce infectious virus under

181 IN-suppressed conditions, we performed the experiment depicted in Fig. 3A.
182 Irradiated or untreated Jurkat cells were infected with replication
183 competent HIV-1 NL4-3 in the presence of an IN inhibitor. Three days after
184 infection, the IN inhibitor and free viral particles were removed with
185 extensive washing, followed by the addition of fresh Jurkat cells (as
186 indicated day 0). p24CA viral antigen in culture supernatant was monitored.
187 Under nonirradiated culture conditions, the peak of p24CA was observed at
188 days 9 and 11 after the removal of IN inhibitor. In contrast, faster viral
189 replication was observed in the irradiated culture. The peak of viral
190 replication was observed at 5 days post infection (dpi) (Fig. 3B). The amount
191 of integrated and 2LTR proviral DNA at day 0 was performed by
192 quantitative PCR and we found that the irradiated culture contained a
193 fivefold higher amount of integrated provirus than the nonirradiated culture.
194 On the other hand, there was no significant difference in the copy number of
195 2LTR cDNA between irradiated and non-irradiated cultures, suggesting that
196 DNA damage before viral infection increased the HIV cDNA insertion under

197 IN-deficient condition and it promoted faster viral replication.

198

Discussion

Previously, Gaur *et al.* reported that mutant HIV lacking IN activity due to mutations in a highly conserved DDE motif of IN is able to integrate into the host chromosome (Gaur and Leavitt, 1998). Here, we showed that not only an IN-deficient virus, but also WT virus in the presence of an IN inhibitor is integrated into the host chromosome. The efficiency of HIV integration when using an IN antagonist was only 0.1-0.2% of IN mediated integration, indicating that the IN-independent integration pathway may be only a minor pathway or accidental event *in vivo*. However, we showed that stress inducing DNA damage enhances IN-independent infection of cells by HIV and has the potential to serve as a virus reservoir, thus suggesting that it may play a role in disease progression. For instance, although combination antiretroviral therapy (cART) has reduced the pathogenesis of AIDS-related malignancies, there has been an increase of HIV- positive patients with non-AIDS-defining malignancies such as Hodgkin's lymphoma, invasive anal

carcinoma, lung cancer, skin cancers, and hepatocellular carcinoma (Spano et al., 2008). Radiotherapy is a standard treatment procedure for many individuals with cancer and HIV, even though it may adversely affect HIV disease status and CD4 counts (Housri et al., 2010). Our results suggest that patients receiving simultaneous medical treatment for cancer in the form of radiation and anticancer drug therapy may be at a higher risk for DNA repair mediated integration of proviral cDNA. In addition, it is conceivable that even mental stress-induced radical oxidents may augment IN-independent infection (Adachi et al., 1993; Morimoto et al., 2008). Another interesting point of our study is that the level of gene expression under IN-deficient conditions is reduced when compared with functional IN infection. Altogether, the IN-independent integration enhanced by the stress may lead to latent infection *in vivo*.

The mechanism of IN-independent integration remains to be elucidated. Previously, Gaur *et al.* sequenced the host-virus junction and showed that IN-independently integrated provirus do not have a duplicate

231 5-bp repeat of host cell DNA which is characteristically generated by the
232 staggered cleavage of host DNA during the strand transfer reaction of HIV
233 (Gaur and Leavitt, 1998). They also demonstrated that the integrated DNA
234 of IN defective virus includes the deletion of host DNA, LTR, and the
235 insertion of unknown sequences between the virus-host DNA junction. Based
236 on these results, they suggested that the integration of HIV lacking IN
237 activity may be catalyzed by the host DNA repair system (Gaur and Leavitt,
238 1998). In line with their findings, our results also demonstrate a significant
239 increase in the deletion and insertion of nucleotides at virus-host DNA
240 junctions in cells infected with an IN-deficient virus (WT: 6% vs.
241 IN-deficient: 27%, P=0.0039). Furthermore, induced DNA damage in target
242 cells before virus infection increased the level of integration (Fig. 1G and H).
243 These findings support the idea that HIV DNA is inserted into a DNA break
244 point by the host DNA repair system, possibly by non-homologous end
245 joining and/or homologous recombination.

246 It has been reported that the provirus integration site determines

247 the level of viral gene expression (Wang et al., 2007). The host protein
248 LEDGF/p75 promotes HIV integration into active gene coding regions (Shun
249 et al., 2007). However, silencing of the active provirus integrated at gene
250 coding regions is challenging, making it unclear how latently infected cells
251 are established. In this paper, we demonstrated that the level of gene
252 expression is reduced depending on distinct integration pathways. To
253 understand the mechanism of gene silencing, we analyzed integration sites
254 of HIV-1 proviruses. Unexpectedly, the integration site analysis focused on
255 gene coding sequences did not show significant differences between the
256 targeted frequency of integration under IN-deficient and conventional
257 conditions. However, HIV-1 cDNAs were frequently inserted into
258 minisatellite-like repeat sequences of genomic DNA in IN-deficient
259 conditions (P=0.0048). Minisatellites consist of 10-100 bp repeat sequences
260 and is observed near cis-acting meiotic double-strand break hotspots
261 (Richard and Paques, 2000). If the IN-independent integration is carried out
262 by the DNA repair pathway, then double-strand break hotspots may be the

263 target for the insertion. Moreover, an unknown determinant for these
264 hotspots, such as the chromatin environment, may cause reduced gene
265 expression.

266 In our study, the integration frequency of provirus into gene coding
267 regions was 68% under an IN-deficient condition and it was much higher
268 than predicted random integration (33%). Hence, this result suggests that
269 there is an integration preference into gene coding regions even during
270 IN-independent integration. One plausible explanation for this may be
271 attributed to the sensitivity of the host chromosome to DNA breaks at gene
272 coding regions. If the retroviral DNA is inserted at a DNA break point under
273 IN-independent conditions, our results might reflect a high frequency of
274 DNA breaks that occur at gene coding regions in the host chromosome.
275 Another possibility is that LEDGF/p75 may promote viral cDNA tethering to
276 the gene coding region even under IN-deficient conditions the same as it
277 would if functional IN was present. We used a lentivirus vector that has a
278 mutation in the catalytic domain of IN. It has been reported that the

279 mutation at D64 to alanine in IN does not inhibit the protein binding with
280 LEDGF/p75 (Cherepanov et al., 2005). In addition, the binding interface of
281 IN and LEDGF/p75 does not overlap with the active site of IN inhibitors
282 (Hare et al., 2010; Michel et al., 2009). Therefore, it is not surprising that
283 retroviral DNA preferentially integrated into gene coding regions even when
284 IN is devoid of function.

285 In sum, our findings serve as a caveat for an alternative infection
286 route that HIV can take during DNA damage to bypass a drug therapy
287 involving IN inhibitors. The analysis of integrated provirus obtained from
288 HIV-positive patients after radiotherapy will shed light on the practical risk
289 of the alternative infection pathway of HIV *in vivo*.
290

Materials and Methods

Virus preparation

Human embryonic kidney (HEK) 293T cells were used for virus preparation. HEK293T cells were transfected by the calcium phosphate method as described previously (Kawano et al., 2004). The culture supernatants at 48 hr post-transfection were centrifuged and filtrated. To prepare HIV-1 NL4-3, cells were transfected pNL4-3 and the 50% tissue culture infective dose ($TCID_{50}$) was calculated (Kawano et al., 1997; Sato et al., 2008). To prepare the EGFP-expressing HIV vector, pCS-CDF-EG-Pre was used. pCS-CDF-EG-Pre was constructed by exchanging the CMV promoter of pCS-CDF-CG-Pre (Miyoshi et al., 1998) with human elongation factor 1α (EF) promoter. The pCS-CDF-EG-Pre was transfected together with packaging plasmids pMD.G, pMDLg/pRRE and pRSV Rev (Kawano et al., 2004). To prepare VSV-G pseudotyped WT LTIG vector, pEV731, kindly

307 provided by Dr. Eric Verdin (Jordan et al., 2003), pMD.G, pMDLg/pRRE and
308 pRSV Rev were cotransfected. To prepare IN-deficient D64V mutant virus,
309 pMDKg/pRRE/D64V, kindly provided by Dr. Ikawa (Okada et al., 2009), were
310 used instead of pMDLg/pRRE. The infectious dose of GFP expressing vectors
311 was assessed as follows. Two hundreds thousand of Jurkat cells were
312 infected with serial volumes of stock virus in 1 ml culture. At 3 days after
313 infection, GFP positive cells were analyzed by flow cytometry and the
314 infectious dose were calculated based on the input volume of virus counted
315 approximate 25% of GFP positive.
316
317 *ELISA*
318
319 To quantify the viral antigen p24CA in virus solutions, an HIV-1 p24
320 antigen enzyme linked immunosorbent assay (ELISA) kit (ZetroMetrix
321 Buffalo, NY) was used.
322

323 *Cell culture*

324

325 HEK293T cells were maintained in Dulbecco's Modified Eagle
326 Medium (DMEM) containing 10% fetal calf serum (FCS), 100 U/ml penicillin
327 and 100 g/ml streptomycin. Jurkat cells were maintained in RPMI 1640
328 medium containing 10% FCS, 100 U/ml penicillin and 100 g/ml
329 streptomycin.

330

331 *Pseudotyped virus infection and flow cytometry analaysis*

332

333 To infect with pseudotyped EGFP-expressing HIV vector or LTIG
334 virus, Jurkat cells (2 x 10^5 cells) were infected with a pseudotyped virus
335 solution as corresponded as the infectious dose of WT virus at indicated MOI.
336 Under IN-deficient conditions, 1 μM Ral (NIH, Bethesda, MD) or 100 nM Elv
337 (Selleck Chemicals, Houston, TX) were added in the culture medium. Two
338 weeks after infection, the percentage of GFP positive cells were measured by

339 flow cytometry. Cells were suspended in phosphate-buffered saline (PBS)
340 containing 1% formamide. Flow cytometry was performed with a
341 FACSCalibur (BD Biosciences), and data were analyzed using CellQuest
342 software (BD Biosciences).
343
344 *Cell cloning*
345
346 GFP+ Jurkat cells were sorted from the bulk culture 2 weeks after
347 pseudotyped LTIG virus infection by FACSAria (BD Biosciences) and cell
348 cloning was carried out by limiting dilution method.
349
350 *Induction of DNA damage in target cells*
351 To induce DNA damages with ionizing radiation, Jurkat cells (2 x
352 10^5 cells/ml) were exposed to appropriate doses (5-10 Gy) of gamma
353 irradiation at room temperature. Gamma irradiation was performed using a
354 Faxitron RX-650 (Faxitron bioptcs, Lincolnshire, IL). To induce DNA damage

355 with reactive oxygen species, Jurkat cells (2 x 10^5 cells/ml) were incubated
356 with medium containing 10, 50 or 100 μM of hydrogen peroxide at 37˚C for 6
357 hr. The DNA damage-induced cells were immediately used for virus infection
358 after washing with medium once.
359
360 *Analyses of HIV-1 replication kinetics of IN-independent integrated*
361 *proviruses*
362
363 The irradiated or non-treated Jurkat cells (4 x 10^4 cells) were
364 incubated for 2 hr in HIV-1 NL4-3 solution containing 40 ng of p24CA
365 ($TCID_{50}$, 984419/ml) at MOI of 1 with 100 nM Elv or 1 μM Ral. After
366 extensive washing (twice with PBS, once with 5% tripsin/EDTA at 37 ˚C for 5
367 min, and twice with 10% FCS RPMI), the cells were resuspended and
368 cultured in 200 μl of 10% RPMI containing an appropriate concentration of
369 IN inhibitor. After 3 dpi, the cells were extensively washed as mentioned
370 above. Then, the cells were co-cultured with fresh Jurkat cells (4 x 10^4 cells)

371 in the absence of IN inhibitor. The culture medium was harvested at the
372 indicated time points and the level of p24CA antigen was measured by
373 ELISA.
374
375 *Quantitative analysis of retroviral DNA*
376
377 The amount of proviral, 2LTR, and full-length forms of HIV-1 DNA
378 was quantified by real-time PCR and the copy number of HIV-1 DNA was
379 normalized by ß-actin as previously described (Suzuki et al., 2003).
380 Unintegrated linear DNA was calculated by subtracting the copy number of
381 provirus and 2LTR DNA from that of full-length viral DNA.
382
383 *Integration site analyses*
384
385 The integration sites were determined by the linker ligation method
386 described previously (Ciuffi et al., 2009). Briefly, genomic DNA was extracted

387 from the GFP positive population using a DNeasy column (QIAGEN) and
388 digested with AvrII, NheI and XbaI. The digested genomic DNA was ligated
389 with a linker adapter and then used to perform nested PCR amplification.
390 The PCR products were cloned into a pGEM-T (Promega, Madison, WI)
391 vector and sequenced using an M13 primer.
392
393 *Statistical analysis*
394
395 The student's *t* test was used to determine statistical significance. *P*
396 values of <0.05, <0.01 and <0.001 were considered significant. The
397 chi-square test was also used to determine statistical significance from the
398 frequency of distribution.
399

400 **Acknowledgments**

401

402 We thank Eric Verdin (Gladstone Institute of Virology and

403 Immunology) for providing pEV731, and Masahito Ikawa (Research Institute

404 for Microbial Diseases, Osaka University) for providing pMDKg/pRRE/D64V.

405 We are also grateful to Peter Gee (Institute for Virus Research, Kyoto

406 University) for proofreading and comments on the manuscript, and Yuri

407 Nakamura (Kyoto University) for support of experimental techniques and

408 helpful discussions. This work was supported in part by Grant-in-Aid for

409 Scientific Research, Grant-in-Aid for Exploratory Research, the Uehara

410 memorial foundation, and the Fujiwara memorial foundation.

411

412

413 **References**

414

415 Adachi, S., Kawamura, K., Takemoto, K., 1993. Oxidative damage of nuclear
416 DNA in liver of rats exposed to psychological stress. Cancer Res. 53,
417 4153-4155.
418 Brady, T., Agosto, L.M., Malani, N., Berry, C.C., O'Doherty, U., Bushman, F.,
419 2009. HIV integration site distributions in resting and activated CD4+ T
420 cells infected in culture. AIDS 23, 1461-1471.
421 Cherepanov, P., Sun, Z.Y., Rahman, S., Maertens, G., Wagner, G., Engelman,
422 A., 2005. Solution structure of the HIV-1 integrase-binding domain in
423 LEDGF/p75. Nat Struct Mol Biol 12, 526-532.
424 Ciuffi, A., Ronen, K., Brady, T., Malani, N., Wang, G., Berry, C.C., Bushman,
425 F.D., 2009. Methods for integration site distribution analyses in animal cell
426 genomes. Methods 47, 261-268.
427 Craigie, R., 2001. HIV integrase, a brief overview from chemistry to
428 therapeutics. J. Biol. Chem. 276, 23213-23216.
429 Dall, K.L., Scarpini, C.G., Roberts, I., Winder, D.M., Stanley, M.A.,
430 Muralidhar, B., Herdman, M.T., Pett, M.R., Coleman, N., 2008.
431 Characterization of naturally occurring HPV16 integration sites isolated
432 from cervical keratinocytes under noncompetitive conditions. Cancer Res. 68,
433 8249-8259.
434 Engelman, A., Cherepanov, P., 2008. The lentiviral integrase binding protein
435 LEDGF/p75 and HIV-1 replication. PLoS Pathog 4, e1000046.
436 Francis, R., Richardson, C., 2007. Multipotent hematopoietic cells
437 susceptible to alternative double-strand break repair pathways that promote
438 genome rearrangements. Genes Dev. 21, 1064-1074.
439 Gaur, M., Leavitt, A.D., 1998. Mutations in the human immunodeficiency
440 virus type 1 integrase D,D(35)E motif do not eliminate provirus formation. J.
441 Virol. 72, 4678-4685.
442 Hare, S., Gupta, S.S., Valkov, E., Engelman, A., Cherepanov, P., 2010.
443 Retroviral intasome assembly and inhibition of DNA strand transfer. Nature

444 464, 232-236.
445 Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., Oshida, T., Ikuta,
446 K., Jern, P., Gojobori, T., Coffin, J.M., Tomonaga, K., 2010. Endogenous
447 non-retroviral RNA virus elements in mammalian genomes. Nature 463,
448 84-87.
449 Horie, M., Tomonaga, K., 2011. Non-retroviral fossils in vertebrate genomes.
450 Viruses 3, 1836-1848.
451 Housri, N., Yarchoan, R., Kaushal, A., 2010. Radiotherapy for patients with
452 the human immunodeficiency virus: are special precautions necessary?
453 Cancer 116, 273-283.
454 Jordan, A., Bisgrove, D., Verdin, E., 2003. HIV reproducibly establishes a
455 latent infection after acute infection of T cells in vitro. EMBO J. 22,
456 1868-1877.
457 Kawano, Y., Tanaka, Y., Misawa, N., Tanaka, R., Kira, J.I., Kimura, T.,
458 Fukushi, M., Sano, K., Goto, T., Nakai, M., Kobayashi, T., Yamamoto, N.,
459 Koyanagi, Y., 1997. Mutational analysis of human immunodeficiency virus
460 type 1 (HIV-1) accessory genes: requirement of a site in the nef gene for
461 HIV-1 replication in activated CD4+ T cells in vitro and in vivo. J. Virol. 71,
462 8456-8466.
463 Kawano, Y., Yoshida, T., Hieda, K., Aoki, J., Miyoshi, H., Koyanagi, Y., 2004.
464 A lentiviral cDNA library employing lambda recombination used to clone an
465 inhibitor of human immunodeficiency virus type 1-induced cell death. J.
466 Virol. 78, 11352-11359.
467 Llano, M., Vanegas, M., Fregoso, O., Saenz, D., Chung, S., Peretz, M.,
468 Poeschla, E.M., 2004. LEDGF/p75 determines cellular trafficking of diverse
469 lentiviral but not murine oncoretroviral integrase proteins and is a
470 component of functional lentiviral preintegration complexes. J. Virol. 78,
471 9524-9537.
472 Maertens, G., Cherepanov, P., Pluymers, W., Busschots, K., De Clercq, E.,
473 Debyser, Z., Engelborghs, Y., 2003. LEDGF/p75 is essential for nuclear and
474 chromosomal targeting of HIV-1 integrase in human cells. J. Biol. Chem. 278,
475 33528-33539.
476 Michel, F., Crucifix, C., Granger, F., Eiler, S., Mouscadet, J.F., Korolev, S.,

Agapkina, J., Ziganshin, R., Gottikh, M., Nazabal, A., Emiliani, S., Benarous, R., Moras, D., Schultz, P., Ruff, M., 2009. Structural basis for HIV-1 DNA integration in the human genome, role of the LEDGF/P75 cofactor. EMBO J. 28, 980-991.

Miyoshi, H., Blomer, U., Takahashi, M., Gage, F.H., Verma, I.M., 1998. Development of a self-inactivating lentivirus vector. J. Virol. 72, 8150-8157.

Morimoto, K., Morikawa, M., Kimura, H., Ishii, N., Takamata, A., Hara, Y., Uji, M., Yoshida, K., 2008. Mental stress induces sustained elevation of blood pressure and lipid peroxidation in postmenopausal women. Life Sci. 82, 99-107.

Morrish, T.A., Gilbert, N., Myers, J.S., Vincent, B.J., Stamato, T.D., Taccioli, G.E., Batzer, M.A., Moran, J.V., 2002. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat. Genet. 31, 159-165.

Nitz, N., Gomes, C., de Cassia Rosa, A., D'Souza-Ault, M.R., Moreno, F., Lauria-Pires, L., Nascimento, R.J., Teixeira, A.R., 2004. Heritable integration of kDNA minicircle sequences from Trypanosoma cruzi into the avian genome: insights into human Chagas disease. Cell 118, 175-186.

Okada, Y., Ueshin, Y., Hasuwa, H., Takumi, K., Okabe, M., Ikawa, M., 2009. Targeted gene modification in mouse ES cells using integrase-defective lentiviral vectors. Genesis 47, 217-223.

Richard, G.F., Paques, F., 2000. Mini- and microsatellite expansions: the recombination connection. EMBO reports 1, 122-126.

Richardson, C., Jasin, M., 2000. Coupled homologous and nonhomologous repair of a double-strand break preserves genomic integrity in mammalian cells. Mol. Cell. Biol. 20, 9068-9075.

Sato, K., Aoki, J., Misawa, N., Daikoku, E., Sano, K., Tanaka, Y., Koyanagi, Y., 2008. Modulation of human immunodeficiency virus type 1 infectivity through incorporation of tetraspanin proteins. J. Virol. 82, 1021-1033.

Schroder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., Bushman, F., 2002. HIV-1 integration in the human genome favors active genes and local hotspots. Cell 110, 521-529.

Shun, M.C., Raghavendra, N.K., Vandegraaff, N., Daigle, J.E., Hughes, S.,

510 Kellam, P., Cherepanov, P., Engelman, A., 2007. LEDGF/p75 functions
511 downstream from preintegration complex formation to effect gene-specific
512 HIV-1 integration. Genes Dev. 21, 1767-1778.
513 Skupsky, R., Burnett, J.C., Foley, J.E., Schaffer, D.V., Arkin, A.P., 2010. HIV
514 promoter integration site primarily modulates transcriptional burst size
515 rather than frequency. PLoS Comput Biol 6.
516 Spano, J.P., Costagliola, D., Katlama, C., Mounier, N., Oksenhendler, E.,
517 Khayat, D., 2008. AIDS-related malignancies: state of the art and
518 therapeutic challenges. J. Clin. Oncol. 26, 4834-4842.
519 Summa, V., Petrocchi, A., Bonelli, F., Crescenzi, B., Donghi, M., Ferrara, M.,
520 Fiore, F., Gardelli, C., Gonzalez Paz, O., Hazuda, D.J., Jones, P., Kinzel, O.,
521 Laufer, R., Monteagudo, E., Muraglia, E., Nizi, E., Orvieto, F., Pace, P.,
522 Pescatore, G., Scarpelli, R., Stillmock, K., Witmer, M.V., Rowley, M., 2008.
523 Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase
524 inhibitor for the treatment of HIV-AIDS infection. J. Med. Chem. 51,
525 5843-5855.
526 Suzuki, K., Ohbayashi, F., Nikaido, I., Okuda, A., Takaki, H., Okazaki, Y.,
527 Mitani, K., 2010. Integration of exogenous DNA into mouse embryonic stem
528 cell chromosomes shows preference into genes and frequent modification at
529 junctions. Chromosome Res. 18, 191-201.
530 Suzuki, Y., Misawa, N., Sato, C., Ebina, H., Masuda, T., Yamamoto, N.,
531 Koyanagi, Y., 2003. Quantitative analysis of human immunodeficiency virus
532 type 1 DNA dynamics by real-time PCR: integration efficiency in stimulated
533 and unstimulated peripheral blood mononuclear cells. Virus Genes 27,
534 177-188.
535 Wang, G.P., Ciuffi, A., Leipzig, J., Berry, C.C., Bushman, F.D., 2007. HIV
536 integration site selection: analysis by massively parallel pyrosequencing
537 reveals association with epigenetic modifications. Genome Res. 17,
538 1186-1194.
539

540

Figure legend

Fig.1. The transduction and integration efficiency of HIV under IN-deficient conditions. Jurkat cells were infected with a MOI of 1-10 with WT EGFP-expressing HIV vector in the absence or presence of 100 nM Elv or 1 μM Ral and indicated as WT, WT+Elv and WT+Ral, respectively. Alternatively, the IN-deficient mutant virus (D64V) was also used. (A) Two weeks after infection, the copy number of integrated, 2LTR and unintegrated forms of viral DNA were analyzed by qPCR. The cell number was determined quantity by qPCR detecting ß-actin. (B) The efficiency of integration under IN-deficient conditions is shown. The values were calculated by dividing the integrated copy number of provirus derived from IN-deficient conditions by that of WT. (C) GFP positive cells were detected by flow cytometry 2 weeks after infection. Pseudo plots of the raw data analyzed by flow cytometry are shown. The numbers indicated on the plot show the percentage of GFP positive cells. These results are summarized in the bar graphs depicted below.

557 (D) Coefficient of correlation. The correlation of GFP positive cells (%) and
558 HIV DNA (copies/1000 cells) are shown. Gray squares indicate WT+Elv (n=9)
559 and triangles indicate D64V (n=9). Pearson's product-moment correlation
560 coefficient was calculated from populations of WT+Elv and D64V (n=18).
561 (E-H) DNA damage enhanced IN-independent integration. Jurkat cells were
562 exposed to various doses of gamma radiation, IR (E and G) or hydrogen
563 peroxide, H_2O_2 (F and H). The cells induced with DNA breaks were infected
564 with GFP expressing HIV vector. (E and F) GFP positive cells were analyzed
565 as mentioned in Fig.1C. Pseudo plots and the summarized graph are shown.
566 (G and H) The efficiency of integration under IN-deficient conditions was
567 calculated as in 1B. The amount of integrated DNA detected in DNA
568 damage-induced cultures was divided by that obtained in the culture
569 infected with WT virus without inhibitor and under non-damaged conditions.
570 All experiments were performed in triplicate (n=3) in A, B, C, E, F and H.
571 The results of flow cytometry in C, E, and F are data from one experiment,
572 which is representative of independent experiment. The error bars in A, B, C,

573 E, F, G and H show standard deviations.

574

575 **Fig. 2.** HIV expression from the LTR promoter of IN-independently
576 generated provirus. Pre-irradiated Jurkat cells were infected with VSV-G
577 pseudotyped LTIG vector. (A-C) Transduction efficiency of LTIG vector under
578 IN-deficient conditions with either WT virus in the presence of Elv (WT+Elv)
579 or D64V mutant virus (D64V). The percentage of GFP positive cells was
580 analyzed by flow cytometry at 2 weeks after infection. (A) Pseudo plots of the
581 raw data analyzed by flow cytometry are shown. The bottom numbers
582 indicated on the plot show the percentage of GFP positive cells, while
583 underlined numbers indicate the mean fluorescence intensity. The results
584 are data from one experiment, which is representative of three independent
585 experiments. (B) The percentage of GFP positive cells is summarized as the
586 graph. (C) Magnitude of virus expression from LTR promoter. The MFI of the
587 GFP expressing cells generated after non-irradiated (0 Gy) or irradiated (5
588 Gy or 10 Gy) stress is shown. All experiments were performed in triplicate

589 (n=3) in B and C. The error bars in A and B show standard deviations. (D)
590 Clonal cell analysis of viral expression. The GFP positive cells were isolated
591 from cells shown C and generated clonal cell lines. The MFI of GFP in each
592 cell clones were shown. The average value of population was shown by
593 horizontal line. (E) Frequency distribution table. IN-dependent (n=17) and
594 IN-independent (n=30) were distributed by MFI.
595
596 **Fig. 3.** The involvement of IN-independent integration in replication
597 competent HIV-1 replication.
598 (A) The experimental procedure is shown. Irradiated (IR(+)) or untreated
599 (IR(-)) Jurkat cells were infected with replication competent HIV-1 NL4-3 in
600 the presence of an IN inhibitor. At 3 dpi, the IN inhibitor and free viral
601 particles were removed with extensive wash. To expand viral replication,
602 fresh Jurkat cells were added and the p24CA viral antigen in culture
603 medium was monitored over the course of 15 days. (B) The p24CA viral
604 antigen in culture medium. (C) The copy number of integrated and 2LTR

605 form of HIV DNA at day 0. All experiments were performed in triplicate
606 (n=3) in B and C. The error bars in B and C show standard deviations. Heat
607 inactivated virus treated at 60 ˚C for 2 hr was used as the negative control.
608 IN inhibitors, Ral or Elv, is indicated in top (C and D). The control
609 experiment was performed in the absence of IN inhibitor and indicated as
610 WT without inhibitor.

Fig. 1

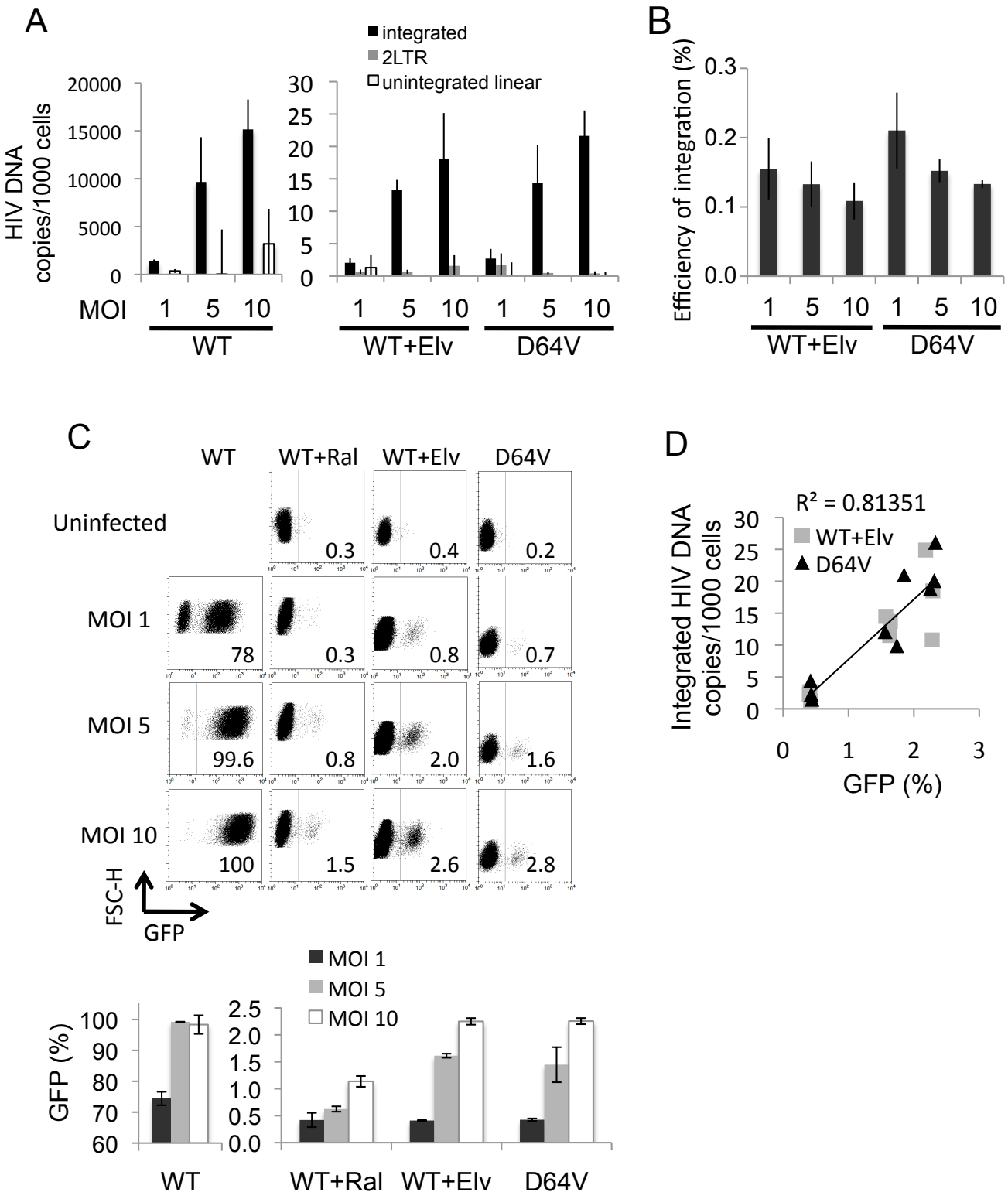
A
integrated
2LTR
unintegrated linear
HIV DNA copies/1000 cells
20000
15000
10000
5000
0
30
25
20
15
10
5
0
MOI 1 5 10 1 5 10 1 5 10
WT
WT+Elv
D64V
B
Efficiency of integration (%)
0.3
0.2
0.1
0.0
1 5 10 1 5 10
WT+Elv
D64V
C
WT
WT+Ral
WT+Elv
D64V
Uninfected
0.3
0.4
0.2
MOI 1
78
0.3
0.8
0.7
MOI 5
99.6
0.8
2.0
1.6
MOI 10
100
1.5
2.6
2.8
FSC-H
GFP
MOI 1
MOI 5
MOI 10
GFP (%)
100
90
80
70
60
2.5
2.0
1.5
1.0
0.5
0.0
WT
WT+Ral
WT+Elv
D64V
D
R² = 0.81351
WT+Elv
D64V
Integrated HIV DNA copies/1000 cells
30
25
20
15
10
5
0
0
1
2
3
GFP (%)

Fig. 1

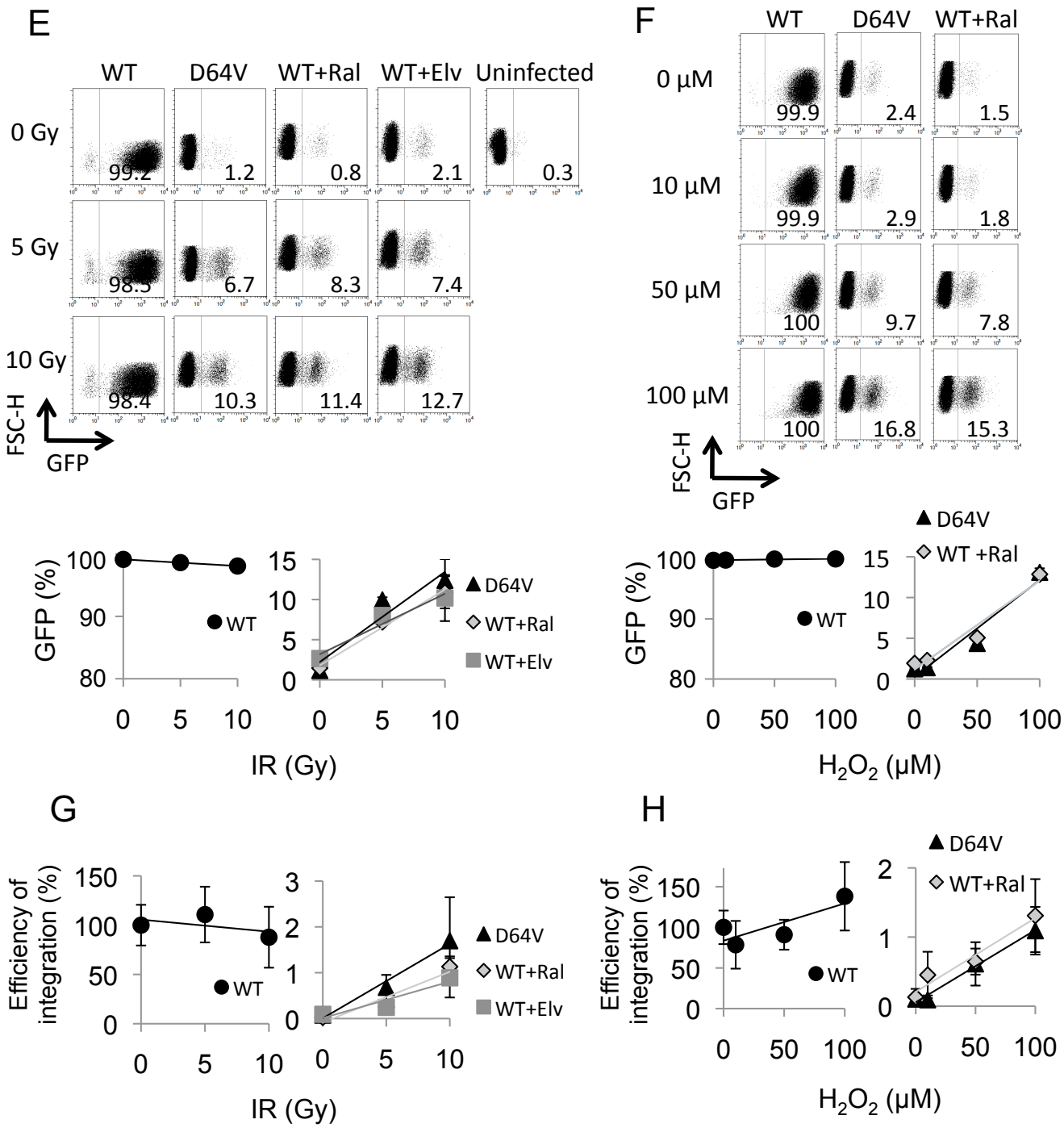
E
WT
D64V
WT+Ral
WT+Elv
Uninfected
0 Gy
5 Gy
10 Gy
99.2
1.2
0.8
2.1
0.3
98.5
6.7
8.3
7.4
98.4
10.3
11.4
12.7
FSC-H
GFP
F
WT
D64V
WT+Ral
0 µM
10 µM
50 µM
100 µM
99.9
2.4
1.5
99.9
2.9
1.8
100
9.7
7.8
100
16.8
15.3
FSC-H
GFP
GFP (%)
WT
D64V
WT+Ral
WT+Elv
IR (Gy)
GFP (%)
WT
D64V
WT +Ral
H2O2 (µM)
G
H
Efficiency of integration (%)
WT
D64V
WT+Ral
WT+Elv
IR (Gy)
Efficiency of integration (%)
WT
D64V
WT+Ral
H2O2 (µM)

Fig. 2

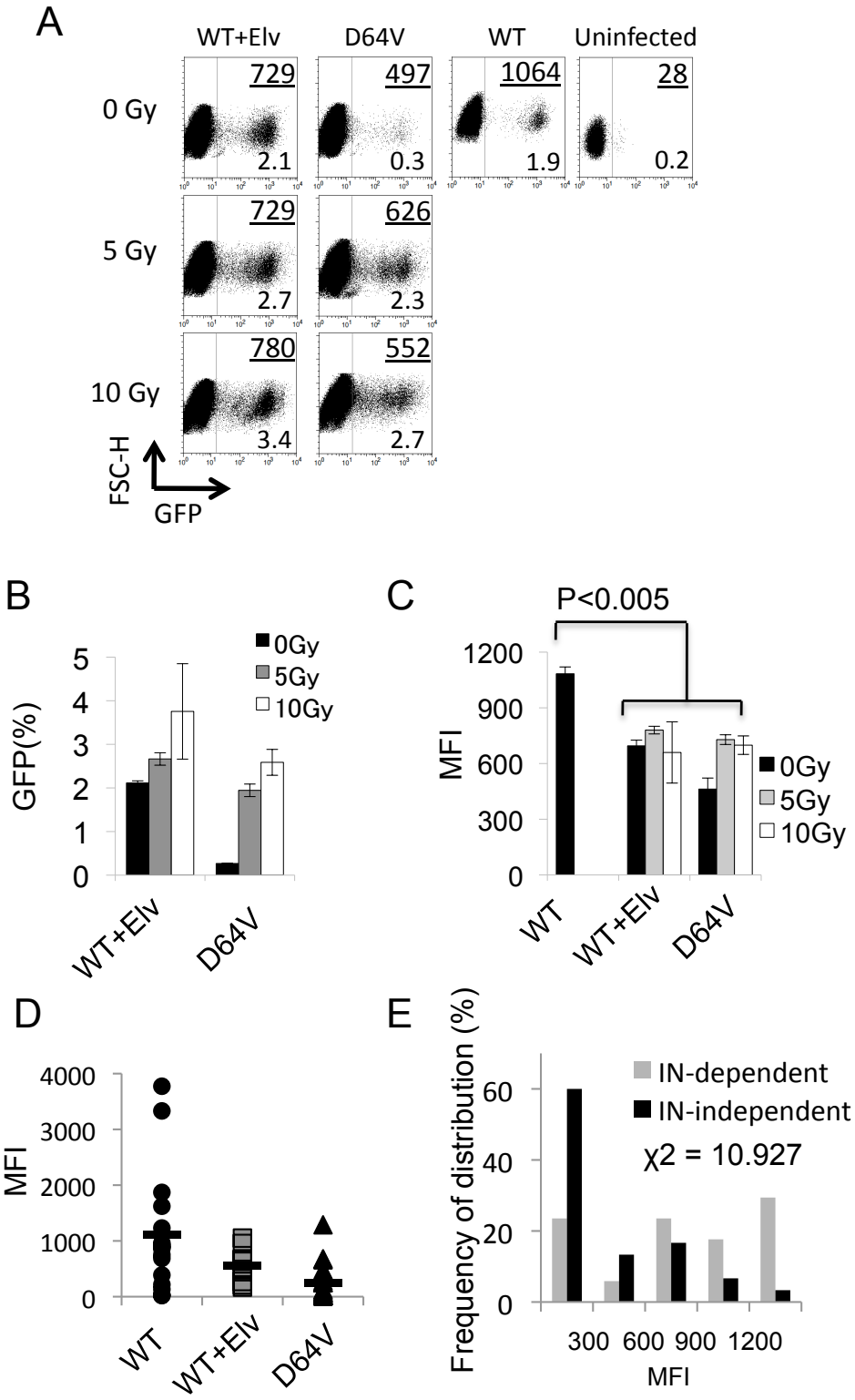
A
WT+Elv
D64V
WT
Uninfected
0 Gy
5 Gy
10 Gy
729
2.1
497
0.3
1064
1.9
28
0.2
729
2.7
626
2.3
780
3.4
552
2.7
FSC-H
GFP
B
GFP(%)
0Gy
5Gy
10Gy
WT+Elv
D64V
C
P<0.005
MFI
0Gy
5Gy
10Gy
WT
WT+Elv
D64V
D
MFI
WT
WT+Elv
D64V
E
Frequency of distribution (%)
IN-dependent
IN-independent
χ2 = 10.927
MFI

Fig. 3

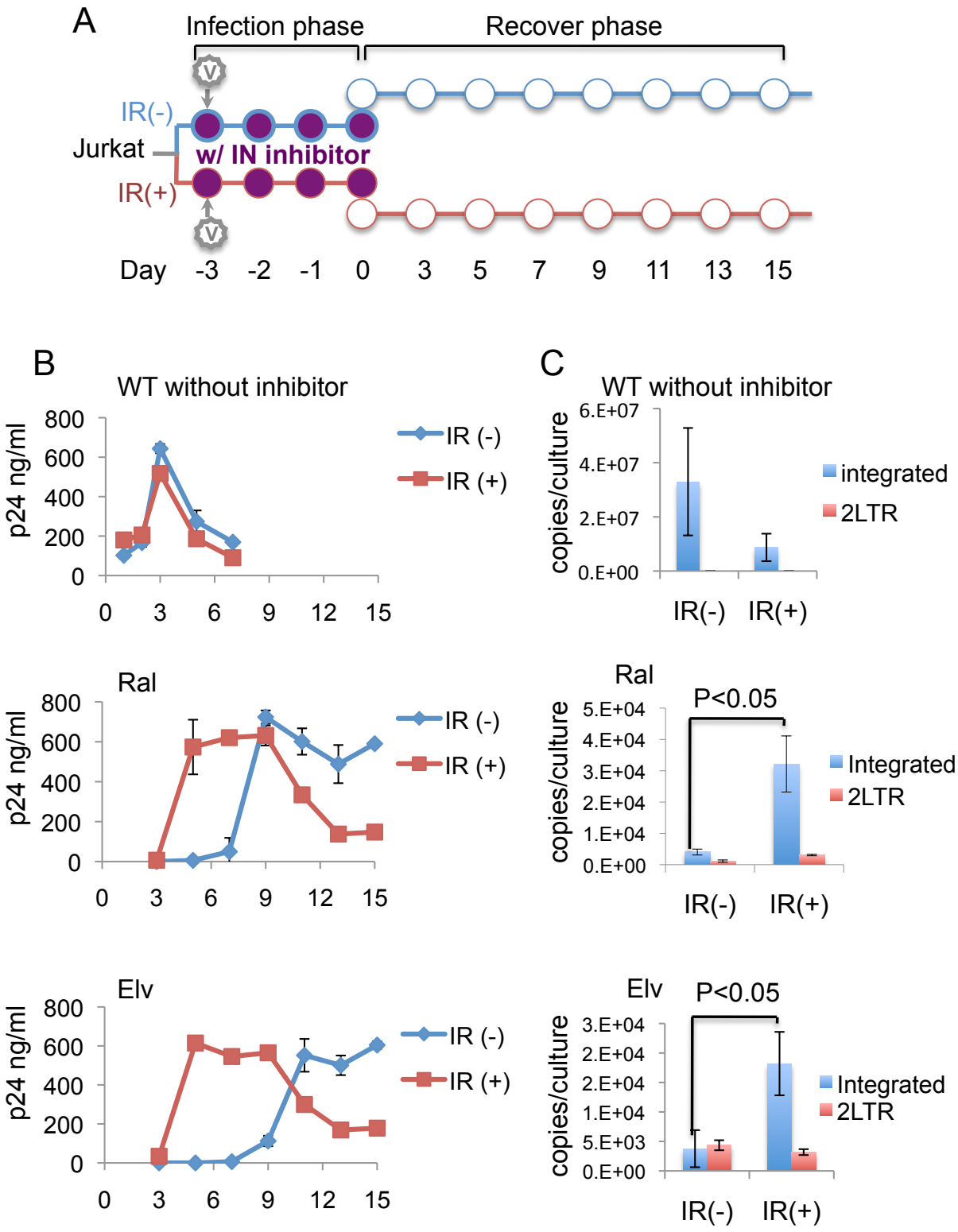
A
Infection phase
Recover phase
V
IR(-)
Jurkat
w/ IN inhibitor
IR(+)
V
Day -3 -2 -1 0 3 5 7 9 11 13 15
B
WT without inhibitor
p24 ng/ml
800
600
400
200
0
0 3 6 9 12 15
IR (-)
IR (+)
Ral
p24 ng/ml
800
600
400
200
0
0 3 6 9 12 15
IR (-)
IR (+)
Elv
p24 ng/ml
800
600
400
200
0
0 3 6 9 12 15
IR (-)
IR (+)
C
WT without inhibitor
copies/culture
6.E+07
4.E+07
2.E+07
0.E+00
IR(-) IR(+)
integrated
2LTR
Ral
P<0.05
copies/culture
5.E+04
4.E+04
3.E+04
2.E+04
1.E+04
0.E+00
IR(-) IR(+)
Integrated
2LTR
Elv
P<0.05
copies/culture
3.E+04
2.E+04
2.E+04
1.E+04
5.E+03
0.E+00
IR(-) IR(+)
Integrated
2LTR

Table 1. Integration sites analysis

Pathway	Total Events	In RefSeq[*3]	(%)[*4]	In repeat Seq	(%)[*5]	Deletion or Insertion	(%)[*6]
IN-dependent[*1]	49	40	84.4	1	2.2	3	6.1
IN-independent[*2]	79	54	68.4	15	19.0	21	26.6
			P=0.098		***P*=0.0048**		***P*=0.0039**

*1: This result summarizes 40 sites derived from LTIG vector infected and 9 sites derived from CS-CDF-EG-Pre infected cells.

*2: This result summarizes 9 sites derived from LTIG vector infected and 75 sites derived from CS-CDF-EG-Pre infected cells. Of these sites, 6 are the results from WT+Elv and 76 six are from D64V. Then, 13 out of 79 results were derived from pre-irradiated culture.

*3: Integration events counted as insertion into reference sequence (=gene coding resion).

*4: A frequency of integration into RefSeq. Random integration into RefSeq is expected to be 33% of total integration sites.

*5: A frequency of integration into repeat sequence.

*6: A frequency of integration with deletion in LTR sequence or with up to 50 bp insertion in LTR-host genome junction.