BRIEF COMMUNICATIONS

A role for LEDGF/p75 in targeting HIV DNA integration

Angela Ciuffi[1], Manuel Llano[2], Eric Poeschla[2], Christian Hoffmann[1], Jeremy Leipzig[1], Paul Shinn[3], Joseph R Ecker[3] & Frederic Bushman[1]

HIV DNA integration is favored in active genes, but the underlying mechanism is unclear. Cellular lens epithelium-derived growth factor (LEDGF/p75) binds both chromosomal DNA and HIV integrase, and might therefore direct integration by a tethering interaction. We analyzed HIV integration in cells depleted for LEDGF/p75, and found that integration was (i) less frequent in transcription units, (ii) less frequent in genes regulated by LEDGF/p75 and (iii) more frequent in GC-rich DNA. LEDGF is thus the first example of a cellular protein controlling the location of HIV integration in human cells.

Targeting of retroviral DNA integration in human cells has been studied in detail for three retroviruses, and notably, they each show different integration site preferences. Sites of HIV integration are enriched in active transcription units[1–3]. Sites of murine leukemia virus (MLV) integration, in contrast, are enriched near transcription start sites and CpG islands[2]. Integration sites of avian sarcoma-leukosis virus are distributed in a near-random fashion, showing only a slight favoring of transcription units and no favoring of transcription start sites[3,4]. These differences among retroviruses support the idea that distinctive interactions between each viral integration complex and nuclear proteins enriched near favored sites may mediate target site selection[5,6]. LEDGF/p75 (also known as PSIP1) is a candidate for one such cellular protein, because LEDGF/p75 binds tightly to HIV integrase[7,8], and *in vivo*, coexpression of LEDGF/p75 with integrase causes integrase to accumulate on chromosomes[9–12].

To analyze the role of LEDGF/p75 in HIV integration site selection, we analyzed the distribution of 4,118 unique integration sites in three cell lines depleted for LEDGF/p75 and in matched controls (**Table 1**). In the first, we introduced two short hairpin (sh)RNAs into 293T cells, which yielded a strong knockdown of LEDGF/p75 termed siLL[10] (**Fig. 1a**). We compared this cell line to 293T cells containing an shRNA of scrambled sequence (siScram) and to unmodified 293T cells (293T-wt). In the second, we introduced an shRNA against LEDGF/p75 into Jurkat T cells, to make Jurkat-siJK2, which also showed a strong LEDGF/p75 knockdown[10]. We compared this cell line to Jurkat-siJK2 cells that were complemented with *PSIP1* (which encodes LEDGF/p75) insensitive to the *PSIP1* shRNA (siJK2BC–'back-complemented')[10] and unmodified Jurkat cells[13]. In

the third, we infected human osteosarcoma cells with a retroviral vector transducing an shRNA targeting a different part of *PSIP1* (ref. 14) to yield HOS-siL. We compared this cell line to cells transduced with the parent retroviral vector lacking the shRNA (HOS-U6). We compared the distributions of integration sites in these data sets to each other, and to previously published sets of HIV[1,3,13] and MLV[2] sites, and to a set of about 37,000 random sites generated *in silico* (**Supplementary Methods** online).

The frequency of integration in transcription units was reduced in all three cell lines containing the LEDGF/p75 knockdown compared to the paired controls (**Fig. 1b**). For siLL, 56% of integration sites were in transcription units compared to 67% in the control siScram ($P = 0.002$ by Chi-square test). For the Jurkat-siJK2 knockdown, 66% of sites were in transcription units compared to 72% in wild-type ($P = 0.0096$) and 71% for the back-complemented line siJK2BC ($P = 0.05$). For the HOS cells, 62% of integration events were in transcription units in the knockdown cells, compared to 67% in the control (as a result of the smaller number of sites sequenced, the difference in the HOS data set did not achieve statistical significance).

As a control for 'off-target' effects of the siScram shRNA, we determined 732 HIV integration site sequences generated by infection of 293T cells containing no shRNAs (293T-wt). This collection of integration sites was indistinguishable from the 293T-siScram, arguing against off-target siRNA effects. For the Jurkat cell experiment, the finding that the back-complemented line was indistinguishable from wild-type also argued against off-target effects.

The LEDGF/p75 knockdown cells showed only a partial reduction in integration in transcription units, leading us to search for additional

Table 1 Integration site data sets used in this study

Virus or vector	Cell type/integration target	Number of integration sites	Source
HIV vector	293T-siLL (LEDGF knockdown)	596	This work
HIV vector	293T-siScram (matched control)	453	This work
HIV vector	293T-wt	732	This work
HIV vector	293T-dKD (LEDGF and HRP2 knockdown)	292	This work
HIV vector	Jurkat-siJK2 (LEDGF knockdown)	602	This work
HIV vector	Jurkat-siJK2BC (LEDGF back-complemented)	476	This work
HIV-Tat vector	Jurkat	914	Ref. 13
HIV vector	HOS-siL (LEDGF knockdown)	248	This work
HIV vector	HOS-U6 (matched control)	105	This work
HIV vector	PBMC	542	Ref. 3
HIV vector	SupT1	542	Ref. 1
MLV vector	HeLa	917	Ref. 2

[1]University of Pennsylvania School of Medicine, Department of Microbiology, 3610 Hamilton Walk, Philadelphia, Pennsylvania 19104-6076, USA. [2]Molecular Medicine Program, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, Minnesota 55905, USA. [3]Genomic Analysis Laboratory, The Salk Institute, 10010 North Torrey Pines Road., La Jolla, California 92037, USA. Correspondence should be addressed to E.P. (poeschla.eric@mayo.edu) or F.B. (bushman@mail.med.upenn.edu).
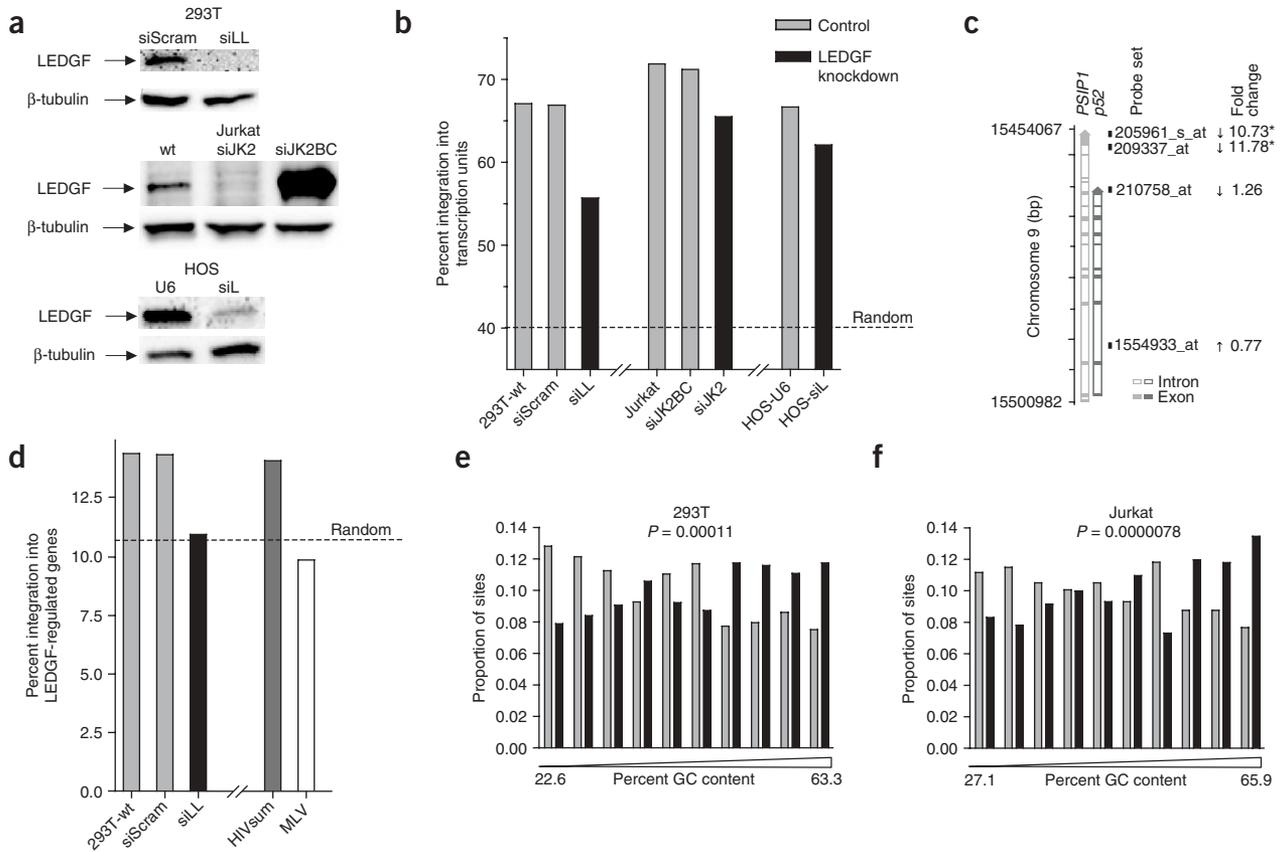
Figure 1 Depleting cells of LEDGF/p75 alters the distribution of HIV integration sites *in vivo*. (**a**) Western blot analysis documenting the LEDGF protein levels in knocked down and control cell lines. (**b**) Frequency of integration in transcription units. The frequency of integration in Unigenes in each data set is shown. Statistical significance for comparison between data sets was determined by the chi-squared test. 'Random' indicates a data set of ~37,000 matched random controls (matching procedure in ref. 3). (**c**) Transcription at the *PSIP1* locus assayed using microarray. Four independent RNA samples were prepared for each 293T-derived cell line and analyzed on eight Affymetrix HU133 Plus 2.0 arrays, which interrogate about 47,400 transcripts. Asterisks indicate probe sets querying *PSIP1* but not the p52 splice variant ($P < 0.01$, *t*-test). The shRNAs used in 293T-siLL targeted sequences present in *PSIP1* 3′ coding region that are not shared with the p52 splice variant, so only *PSIP1* was knocked down. The microarray data underestimate the extent of the knockdown because of the high noise level resulting from the use of mismatched control probes. (**d**) LEDGF/p75-regulated genes are favored targets for HIV integration. Genes hosting HIV integration events in the indicated data sets were divided into LEDGF/p75-regulated and LEDGF/p75-nonregulated sets based on the microarray data. The frequency of integration in LEDGF/p75-regulated genes in each data set is shown. *P* values were determined by chi-squared test. The numbers of sites in 293T-derived cells were too low to achieve significance compared to each other, though 293T-wt and siScram were each significantly different from random and siLL was not. (**e**) Effects of GC content on integration in 293T-derived cells. (**f**) Effects of GC content on integration in Jurkat-derived cells. In **e** and **f**, sites from LEDGF/p75 knockdown and control cells were pooled and divided into 10 equal sets based on GC content of the surrounding 5,000 bp, then the fraction from each cell type plotted. For the 293T-derived cells, the cut values of percent GC for the bins were (starting from the left): bin 1, 22.6–34.1; bin 2, 34.1–35.8; bin 3, 35.8–37.0; bin 4, 37.0–38.3; bin 5, 38.3–39.8; bin 6, 39.8–41.4; bin 7, 41.4–42.9; bin 8, 42.9–46.1; bin 9, 46.1–49.9; bin 10, 49.9–63.3 (those for Jurkat differed slightly).

possible tethering factors. One candidate is cellular hepatoma-derived growth factor–related protein 2 (HRP-2), which has an integrase binding domain similar to that of LEDGF/p75 (refs. 12,15). We analyzed HIV integration targeting in cells containing knockdowns of both LEDGF/p75 and HRP-2 (293T-dKD; mRNA encoding HRP-2 was diminished about 80%, data not shown). The frequency of integration in transcription units in 293T-dKD was indistinguishable from 293T-siLL ($P = 0.46$). Thus these data did not strengthen the idea that HRP-2 is involved in HIV integration and are consistent with the finding that HRP-2 does not tether integrase to chromatin[12].

To analyze the role of LEDGF/p75 using another approach, we reasoned that genes regulated by LEDGF/p75 should be enriched for bound LEDGF/p75 protein, and so LEDGF/p75-regulated genes could be preferential HIV integration targets. We carried out transcriptional profiling on RNAs from the 293T-siScram and 293T-siLL cells,

showing that 869 transcription units were upregulated and 1,139 were downregulated in 293T-siLL ($P < 0.01$, *t*-test). Two probe sets specific for *PSIP1* were significantly downregulated (more than ten times; **Fig. 1c**).

We tabulated the proportion of HIV integration sites in genes regulated by LEDGF/p75 for 293T-wt, siScram and siLL (**Fig. 1d**). Integration in LEDGF/p75-responsive genes was favored over random in the control data sets ($P = 0.005$ for 293T-wt and $P = 0.041$ for siScram), whereas in the 293T-siLL data set these genes were not favored ($P = 0.73$). We next extended the analysis to HIV data sets from other cell types, based on the assumption that many of the LEDGF/p75-regulated genes in 293T cells would also be regulated by LEDGF/p75 in other cell types as well. Analysis of a pool of HIV sites from SupT1, peripheral blood mononuclear cells and Jurkat cells (termed HIVsum) showed favored integration

in LEDGF/p75-responsive genes compared to random control sites ($P < 0.0001$). As another control, we analyzed MLV integration sites. MLV integrase does not bind LEDGF/p75, and accordingly, MLV did not show favored integration in LEDGF/p75-regulated genes ($P = 0.63$). We observed similar trends when we analyzed up-regulated or downregulated genes individually. These findings indicate that LEDGF/p75-regulated genes are favored targets for HIV integration.

Knockdown of LEDGF/p75 also resulted in increased HIV integration in chromosomal regions with higher GC content. **Figure 1e** shows data for the comparison in 293T cells (siLL versus siScram, $P = 1.1 \times 10^{-4}$) and **Figure 1f** shows the comparison for Jurkat cells (Jurkat-siJK2 knockdown versus unmodified Jurkat, $P = 7.8 \times 10^{-6}$). One potential explanation for this bias is that LEDGF/p75 is bound preferentially in AT-rich chromosomal regions through the AT hook motif identified in the LEDGF/p75 sequence, resulting in favored integration in AT-rich DNA. In support of this idea, the AT hook has been found to mediate tethering of LEDGF/p75 to chromatin (M.L. & E.P., unpublished data).

In summary, we present data indicating that cellular LEDGF/p75 affects the choice of target sites for HIV integration in cells. The 293T, Jurkat and HOS cells knocked down for LEDGF/p75 all showed reduced integration in transcription units compared to control cells. We found that LEDGF/p75-regulated genes identified by transcriptional profiling were preferential integration targets, consistent with the idea that genes with bound LEDGF/p75 are favored for HIV integration. The knockdown of LEDGF/p75 also resulted in increased HIV integration in target DNA with higher G-C content. Depleting LEDGF/p75 from cells, however, did not eliminate favored integration in transcription units, but rather resulted in a quantitative reduction, suggesting that additional factors may be involved in targeting HIV DNA integration. It has previously been reported that LEDGF/p75 binds to a variety of RNA polymerase subunits and transcription factors[16]. Our data are consistent with a model in which LEDGF/p75 becomes enriched on transcription units by binding to components of the transcriptional apparatus[16], thereby becoming positioned to target HIV integration to these sites.

Accession codes. For accession codes, see **Supplementary Table 1** online.

Note: Supplementary information is available on the Nature Medicine website.

Published online at http://www.nature.com/naturemedicine/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Schroder, A. *et al. Cell* **110**, 521–529 (2002).
2. Wu, X., Li, Y., Crise, B. & Burgess, S.M. *Science* **300**, 1749–1751 (2003).
3. Mitchell, R. *et al. PLoS Biol.* **2**, E234 (2004).
4. Narezkina, A. *et al. J. Virol.* **78**, 11656–11663 (2004).
5. Sandmeyer, S. *Proc. Natl. Acad. Sci. USA* **100**, 5586–5588 (2003).
6. Bushman, F. *et al. Nat. Rev. Micro.* advance online publication 19 September 2005 (doi:10.1038/nrmicro1263).
7. Cherepanov, P. *et al. J. Biol. Chem.* **278**, 372–381 (2003).
8. Turlure, F., Devroe, E., Silver, P.A. & Engelman, A. *Front. Biosci.* **9**, 3187–3208 (2004).
9. Maertens, G. *et al. J. Biol. Chem.* **278**, 33528–33539 (2003).
10. Llano, M. *et al. J. Virol.* **78**, 9524–9537 (2004).
11. Llano, M., Delgado, S., Vanegas, M. & Poeschla, E.M. *J. Biol. Chem.* **279**, 55570–55577 (2004).
12. Vanegas, M. *et al. J. Cell Sci.* **118**, 1733–1743 (2005).
13. Lewinski, M. *et al. J. Virol.* **79**, 6610–6619 (2005).
14. Devroe, E. & Silver, P.A. *BMC Biotechnol.* **2**, 15 (2002).
15. Cherepanov, P., Devroe, E., Silver, P.A. & Engelman, A. *J. Biol. Chem.* **279**, 48883–48892 (2004).
16. Ge, H., Si, Y. & Roeder, R.G. *EMBO J.* **17**, 6723–6729 (1998).