



Published in final edited form as:

Science. 2014 August 1; 345(6196): 570–573. doi:10.1126/science.1256304.

Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection

Thor A. Wagner^{1,2,*}, Sherry McLaughlin^{1,2,*}, Kavita Garg³, Charles Y. K. Cheung³, Brendan B. Larsen², Sheila Styrchak¹, Hannah C. Huang¹, Paul T. Edlefsen^{2,3}, James I. Mullins^{2,*}, and Lisa M. Frenkel^{1,2,*†}

¹Seattle Children's Research Institute, 1900 9th Avenue, Seattle, WA 98101, USA

²University of Washington, Seattle, WA, USA

³Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Abstract

Antiretroviral treatment (ART) of HIV infection suppresses viral replication. Yet if ART is stopped, virus reemerges because of the persistence of infected cells. We evaluated the contribution of infected-cell proliferation and sites of proviral integration to HIV persistence. A total of 534 HIV integration sites (IS) and 63 adjacent HIV *env* sequences were derived from three study participants over 11.3 to 12.7 years of ART. Each participant had identical viral sequences integrated at the same position in multiple cells, demonstrating infected-cell proliferation. Integrations were overrepresented in genes associated with cancer and favored in 12 genes across multiple participants. Over time on ART, a greater proportion of persisting proviruses were in proliferating cells. HIV integration into specific genes may promote proliferation of HIV-infected cells, slowing viral decay during ART.

Despite suppression of viral replication during ART, HIV reservoirs, measured by the number of resting CD4⁺ T cells with infectious virus induced in cell culture, decline slowly (1). The mechanisms hypothesized to allow infectious proviruses to persist include long-lived latently infected cells (1); low-level HIV replication (2), potentially due to insufficient intracellular drug concentrations (3–5); and the proliferation of HIV-infected cells (2, 6–10). During ART, subpopulations of cells with identical HIV sequences comprise a progressively larger proportion of the persisting viral genomes (8), suggesting that proliferation of infected cells helps maintain the HIV reservoir. To further evaluate the contribution of infected-cell proliferation to HIV persistence, we developed a method [integration site loop amplification (ISLA)] to define sites of HIV integration in single cells and sequence up to 2.8 kb of the 3'

[†]Corresponding author. lfrenkel@uw.edu.

*These authors contributed equally to this work.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/345/6196/570/suppl/DC1

Materials and Methods

Figs. S1 to S3

Tables S1 to S3

References (46–59)

The authors report no conflicts of interest related to this research.

region of the viral genome adjacent to the integration site, allowing us to link specific viral variants to specific integration sites (fig. S1).

A total of 534 proviral integration sites were sequenced from three participants (B1, L1, and R1) at three time points each: after 1 to 2.3, 4.1 to 8.2, and 11.3 to 12.7 years of suppressive ART (Fig. 1, A to C). HIV integration at the same chromosomal site was found in multiple cells within each participant throughout follow-up, whereas no identical integration sites were shared by different participants, suggesting that HIV-infected cells proliferate, as reported (2, 6–8).

The hypothesis was further investigated by comparisons of the viral genomes in 63 integration sites. When HIV C2V5 *env* sequences shared a specific integration site, the *env* sequences [~625 base pairs (bp)] were identical ($n = 31$ C2V5 *env* sequences from 13 integration sites, except for one pair of sequences with a 1-bp difference). In contrast, among proviruses integrated at different positions in the human genome ($n = 45$ unique integration sites), C2V5 *env* sequences were distinct except for three groups from participant B1 (fig. S2) (13 out of 13 versus 3 out of 45; $P < 0.0001$). Sequences of the entire *C2env-nef-3'LTR* (long terminal repeat) region (~2.8 kb) from these three groups of pro-viruses had 2 to 39 nucleotide differences, indicating that many were distinct viruses (fig. S2). In comparison, eight pairs of *C2env-nef-3'LTR* sequences with identical integration sites differed by 0 to 2 bases (mean 0.9), an error rate consistent with misincorporations during the ISLA protocol. Phylogenetic analyses of 396 additional HIV *env* sequences (8), together with those from the ISLA method, revealed multiple additional identical viral sequences (8), strongly suggesting that multiple HIV-infected clonal cell populations persist during ART (Fig. 2 and figs. S2 and S3).

Three approaches were used to explore whether the distribution of HIV integration sites observed was random or shaped by selective forces. Specifically, we examined the distribution of HIV integrations in genes associated with cancer, regulation of cell proliferation, or cell survival.

First, proviral integration sites were examined for overrepresentation in cancer-associated genes from five combined sources ($n = 1332$ unique genes) (11–15). Across the three participants, 12.5% (36 out of 288) of the unique genes from all proviral integrations were annotated as associated with cancer (11–15), compared with only 5.19% (1332 out of 25,660) of the human genes in the human genome ($P < 0.0001$). In addition, unique integrations in proliferating HIV-infected cells (defined as identical integration sites derived from 2 separate cells) (table S1, A to C) were also increased in cancer-associated genes (6 out of 34, 17.65% versus 1332 out of 25,660, 5.2%; $P = 0.0076$), which suggests that HIV integration could disrupt the regulation of these genes, as is known to occur during tumor induction by nonacute oncogenic retroviruses (16).

Second, given that HIV integrates preferentially into actively transcribed genes, especially those activated upon HIV infection (17–19), we compared our participants' integration sites to the >44,000 integration sites mapped in acutely infected CD4⁺ T cells (Jurkat cells). The majority of our participants' proviruses were integrated within genes (425 out of 534,

79.6%) (Fig. 1, A to C), particularly within introns (337 out of 425, 79.1%), as previously observed (9, 17, 20). The frequency of unique proviral integration sites in cancer-associated genes (12.70%) was similar to those mapped in the acutely infected Jurkat cells (11.14%) (21). However, our participants' integration sites in proliferating cells were enriched for cancer-associated genes over samples of the same size from the Jurkat cell data set ($P = 0.0486$). In the Jurkat cell data set, 11.14% of integrations were found in cancer genes, compared with 15.97% in our participants ($P = 0.0828$). Importantly, the frequency of proliferating cells significantly increased through time on ART (Fig. 1, D to F), both when considering all integration sites ($P = 0.0013$) and when considering only unique integration sites ($P = 0.0048$) (Fig. 1, D to F). However, unique integrations into cancer genes did not increase over time. These findings suggest that HIV integrates preferentially into cancer-associated and other genes that promote cell proliferation in a manner distinct from homeostatic proliferation (7, 22). The increasing overrepresentation of proliferating cells containing proviruses suggests that HIV integration plays a mechanistic role in infected cell survival during ART.

Our third approach examined gene ontology in order to associate integration sites with biological processes (GO terms). The unique HIV integrations in our participants were enriched in genes controlling cell proliferation (table S2) and were overrepresented in hallmark cancer GO terms (23–25) when compared with integrations from the Jurkat cell data set (51.29% versus 41.96%; $P = 0.0001$). Integrations in proliferating cells were also highly enriched in the 10 hallmarks of cancer gene categories considered together compared with the Jurkat cell dataset ($P < 0.0001$). Among the 43 hallmark GO terms, significant enrichment was observed for three terms linked to the regulation of cell proliferation or survival (each $P < 0.0001$), including negative regulation of cell cycle (GO:0045786), cell adhesion (GO:0007162), and DNA repair (GO:0006281). This suggests that HIV integration may promote proliferation and persistence of infected cells.

Three genes have been favored for HIV integration in cell lines: *CREBBP*, *FBX11*, and *DNMT1* (18). We detected HIV integrated into the cancer gene *CREBBP* in participants B1 and R1 after 7 and 8.2 years of ART, respectively, indicating that some sites of preferential integration may affect cell survival. However, because we detected only a single integrant in each participant, our findings do not reveal whether integration into *CREBBP* potentiated cell proliferation. Interestingly, among the 288 unique virus integration sites detected within genes, 12 (*ANKRD13C*, *BACH2*, *C2CD3*, *CREBBP*, *HNRNPUL1*, *IKZF3*, *KCTD13*, *MAPK1*, *OXCT1*, *STAT5B*, *ST8SIA4*, and *TRAPPC10*) had proviruses in two of our three participants (table S1), an unusually high number ($P = 0.0264$), suggesting that these regions favor HIV integration and/or survival.

A comparison of our results to published studies of integration sites detected during ART (9, 10, 26, 27) revealed a striking persistence of cells with HIV integrated into *basic leucine zipper transcription factor 2* (*BACH2*), a transcription regulator with numerous biological functions—including tumor suppressor functions in B cells (28), regulating immune activation (29), maintaining naïve T cells, and regulating generation of effector-memory T cells (30)—and it is the target of *Menin* binding for regulation of CD4 T cell senescence and cytokine homeostasis (31). HIV integrations were mapped to seven positions in *BACH2* in

participant L1 and 2 positions in participant R1 (Fig. 3). Despite uneven sampling of HIV integration sites during ART, proviruses were detected in the *BACH2* gene of 7 of 28 individuals previously reported; at 0 out of 74 sites in 16 individuals (26, 33), 37 out of 442 sites in 2 of 3 individuals (9), 1 out of 63 sites in 1 of 8 individuals (27), and 4 out of 23 sites in peripheral blood mononuclear cells (PBMCs) from 1 person (10). All of these integrations occurred in intron 5, just upstream from the start codon of *BACH2*, and all in the same transcriptional orientation as the gene (Fig. 3). In contrast, among the >44,000 integration sites in the Jurkat cell data set (21), only one integration into intron 5 of *BACH2* was found (Fig. 3), and none among the integration sites evaluated in other studies of various cultured cells (17, 19, 34). The finding of disproportionate integrations into this intron of *BACH2* in proliferating and persisting cells strongly suggest that this event favors cell survival; HIV integration just upstream of the start of the coding sequence could lead to promoter insertion-mediated activation (16) and thus dysregulation of the *BACH2* gene. Indeed, another *BACH2* intron is a target for integration by murine leukemia viruses found in B cell lymphomas of mice (32).

The most extreme case of proliferation (31% of infected cells in participant L1) was found with provirus integrated into *MDC1* (*mediator of DNA damage checkpoint 1*), corresponding to cell expansion by a factor of $\sim 4 \times 10^8$ in vivo. However, our study design likely underestimates the occurrence of HIV integration into genes that favor cell proliferation. First, sampling of integration sites was limited, and examination of phylogenetic trees suggests additional proliferating cell populations for which integration sites were not derived (Fig. 2 and figs. S2 and S3). Second, ~20% of the integration sites were outside of gene coding regions, and these regions were not assessed for associations with cancer or cell survival, although many were noted within proliferating cells (Fig. 1). Third, cancer databases often omit genes associated infrequently with cancers, and given the rapid rate that new cancer genes are reported (12), some recently described cancer genes have not yet been added to these databases. For example, *BACH2*, despite being regarded as a tumor suppressor in B cells (28) and a cellular oncogene in mice (32), has not yet been classified as a cancer gene. Also, mutations in *IZKF3*, which also had HIV integrations detected in two of our three participants, was recently associated with a form of acute lymphoblastic leukemia (35). Fourth, as somatic mutations that “drive” cancers are estimated to only convey a 0.4% growth advantage (36), HIV integration into genes with subtle enhancement of cell proliferation may be difficult to detect as clonal due to our limited sampling.

HIV-infected cells that express viral proteins are likely to be eliminated by immune surveillance, or virus replication may lead to cell lysis. Whether the proliferating and persisting HIV-infected cells that we describe harbor replication-competent virus is critical to defining their role in perpetuating the infectious virus reservoir. Undoubtedly, some clonal populations persist due to defects in expression of the proviral genome (9, 10, 37). Although we did not evaluate viral sequences for replication competency, lethally hypermutated viral genomes were linked to three integration sites (Fig. 2). However, cells producing viremias with identical *env* sequences have been shown to harbor replication-competent virus (38). Also, approximately 12% of proviruses refractory to in vitro induction were found to have intact genomes and may be infectious (27). Although transcriptional

interference was not detected in the aforementioned noninduced viral transcripts (26, 27), others have observed that the site of integration may cause transcriptional interference (34, 39–43).

In conclusion, HIV integration into genes associated with cancer or cell cycle regulation appears to confer a survival advantage that allows these cells to persist during suppressive ART, with cell proliferation appearing to serve as an important mechanism of HIV persistence. To be defined are the mechanisms contributing to cell proliferation, the role of proliferating cells in perpetuating the infectious virus reservoir, and whether therapies that target HIV-infected proliferating cells, specific genes, or their products may contribute to a curative strategy for HIV infection.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We wish to express gratitude to the study participants who donated their time and provided specimens, and to N. H. Tobin, A. J. Melvin, and K. M. Mohan for their contributions to this study. Additionally, we thank Y. C. Ho, R. and J. Siliciano, H. Imamichi and C. Lane, and N. Malani and F. Bushman for providing their HIV integration site data. HIV *env* sequences generated from specific integration sites were deposited in GenBank under accession numbers (KM025062 to KM025124 and KM201287 to KM201295). Additional HIV *env* sequences were deposited in GenBank under accession numbers AY075701 to AY077450 (44), AY483287 to AY484389 (2), and KC314011 to KC315287 (8). Sequences of HIV-3'-LTR and adjacent chromosome integration sites are available in table S3. Other data are tabulated in the main paper and in the supplementary materials. Integration site sequences and HIV sequence alignments are available for download at <http://mullinslab.microbiol.washington.edu/publications/supplemental.html>. These studies were funded by R01AI091550 (L.M.F.), K23AI077357 (T.A.W.), R01 AI111806 (J.I.M.), the Molecular Profiling and Computational Biology Core of the University of Washington's Center for AIDS Research (P30 AI027757), and the Canadian Institutes of Health (HIV/AIDS Research Initiative award 201311CVI-322424-244686) (C.Y.K.C.). We wish to thank A. Embry and D. Lawrence at the National Institutes of Health for their support.

REFERENCES AND NOTES

1. Siliciano JD, et al. *Nat Med*. 2003; 9:727–728. [PubMed: 12754504]
2. Tobin NH, et al. *J Virol*. 2005; 79:9625–9634. [PubMed: 16014925]
3. Sigal A, et al. *Nature*. 2011; 477:95–98. [PubMed: 21849975]
4. Letendre S, et al. *Arch Neurol*. 2008; 65:65–70. [PubMed: 18195140]
5. Fletcher CV, et al. *Proc Natl Acad Sci USA*. 2014; 111:2307–2312. [PubMed: 24469825]
6. Bailey JR, et al. *J Virol*. 2006; 80:6441–6457. [PubMed: 16775332]
7. Chomont N, et al. *Nat Med*. 2009; 15:893–900. [PubMed: 19543283]
8. Wagner TA, et al. *J Virol*. 2013; 87:1770–1778. [PubMed: 23175380]
9. Ikeda T, Shibata J, Yoshimura K, Koito A, Matsushita S. *J Infect Dis*. 2007; 195:716–725. [PubMed: 17262715]
10. Imamichi H, et al. *AIDS*. 2014; 1
11. Forbes SA, et al. *Nucleic Acids Res*. 2011; 39(Database):D945–D950. [PubMed: 20952405]
12. Lawrence MS, et al. *Nature*. 2014; 505:495–501. [PubMed: 24390350]
13. Zhao M, Sun J, Zhao Z. *Nucleic Acids Res*. 2013; 41(D1):D970–D976. [PubMed: 23066107]
14. Kandoth C, et al. *Nature*. 2013; 502:333–339. [PubMed: 24132290]
15. Vogelstein B, et al. *Science*. 2013; 339:1546–1558. [PubMed: 23539594]
16. Fan H, Johnson C. *Viruses*. 2011; 3:398–422. [PubMed: 21994739]

17. Schröder AR, et al. *Cell*. 2002; 110:521–529. [PubMed: 12202041]
18. Bushman F, et al. *Nat Rev Microbiol*. 2005; 3:848–858. [PubMed: 16175173]
19. Mitchell RS, et al. *PLOS Biol*. 2004; 2:e234. [PubMed: 15314653]
20. Soto MJ, Peña A, Vallejo FG. *AIDS Res Hum Retroviruses*. 2011; 27:547–555. [PubMed: 20919923]
21. Wang GP, Ciuffi A, Leipzig J, Berry CC, Bushman FD. *Genome Res*. 2007; 17:1186–1194. [PubMed: 17545577]
22. Persaud D, et al. *J Virol*. 2004; 78:968–979. [PubMed: 14694128]
23. Hanahan D, Weinberg RA. *Cell*. 2000; 100:57–70. [PubMed: 10647931]
24. Hanahan D, Weinberg RA. *Cell*. 2011; 144:646–674. [PubMed: 21376230]
25. Plaisier CL, Pan M, Baliga NS. *Genome Res*. 2012; 22:2302–2314. [PubMed: 22745231]
26. Han Y, et al. *J Virol*. 2004; 78:6122–6133. [PubMed: 15163705]
27. Ho YC, et al. *Cell*. 2013; 155:540–551. [PubMed: 24243014]
28. Swaminathan S, et al. *Nat Med*. 2013; 19:1014–1022. [PubMed: 23852341]
29. Roychoudhuri R, et al. *Nature*. 2013; 498:506–510. [PubMed: 23728300]
30. Tsukumo S, et al. *Proc Natl Acad Sci USA*. 2013; 110:10735–10740. [PubMed: 23754397]
31. Kuwahara M, et al. *Nat Commun*. 2014; 5:3555. [PubMed: 24694524]
32. Liu J, et al. *BMC Mol Biol*. 2009; 10:2. [PubMed: 19159451]
33. Han Y, Wind-Rotolo M, Yang HC, Siliciano JD, Siliciano RF. *Nat Rev Microbiol*. 2007; 5:95–106. [PubMed: 17224919]
34. Lewinski MK, et al. *J Virol*. 2005; 79:6610–6619. [PubMed: 15890899]
35. Holmfeldt L, et al. *Nat Genet*. 2013; 45:242–252. [PubMed: 23334668]
36. Bozic I, et al. *Proc Natl Acad Sci USA*. 2010; 107:18545–18550. [PubMed: 20876136]
37. Josefsson L, et al. *PLOS Pathog*. 2013; 9:e1003432. [PubMed: 23818847]
38. Anderson JA, et al. *J Virol*. 2011; 85:5220–5223. [PubMed: 21367910]
39. Mok HP, Javed S, Lever A. *Gene Ther*. 2007; 14:741–751. [PubMed: 17330088]
40. Imai K, Togami H, Okamoto T. *J Biol Chem*. 2010; 285:16538–16545. [PubMed: 20335163]
41. Friedman J, et al. *J Virol*. 2011; 85:9078–9089. [PubMed: 21715480]
42. Eipers PG, Salazar-Gonzalez JF, Morrow CD. *Virology*. 2011; 410:151–160. [PubMed: 21115184]
43. Sherrill-Mix S, et al. *Retrovirology*. 2013; 10:90. [PubMed: 23953889]
44. Frenkel LM, et al. *J Virol*. 2003; 77:5721–5730. [PubMed: 12719565]
45. Armitage P. *Biometrics*. 1955; 11:375–386.

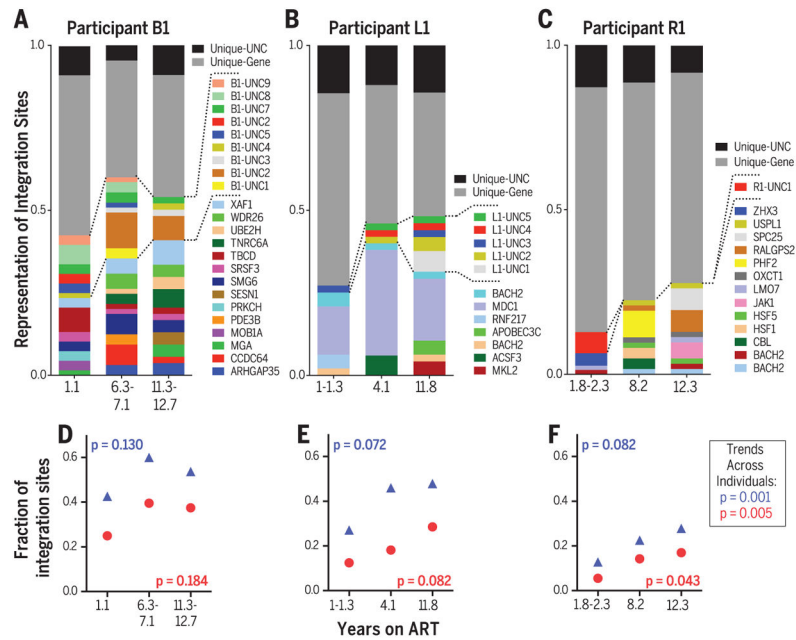


Fig. 1. Representation of HIV integration sites sampled through time

(A to C) show the scaled representation of each gene with integration sites mapped for the three participants at three intervals (times in years given along the x axis) after initiation of suppressive ART. Integration sites were detected in all chromosomes of all participants, except for chromosome 18 in participant B1, and chromosome 20 and the Y chromosome in L1, the only male studied. The gray and black regions at the top of each column indicate the proportion of uncharacterized noncoding regions (UNC, black) and genes (gray) that had only one integration site found at any time point. The other colors represent each of the genes (intragenic regions) and uncharacterized noncoding regions that were found to have virus integrated at the same position in multiple cells. (D to F) show the proportion of all HIV-infected cells found to be demonstrably proliferating (blue triangles), defined by ≥ 2 cells with identical integration sites (from any time point), and the proportion of unique integration sites with evidence of proliferation (red dots). Data are plotted by time (x axis) after initiation of suppressive ART. The P values shown report the significance of the upward trends of these values through time using nonparametric Cochran-Armitage tests (45).

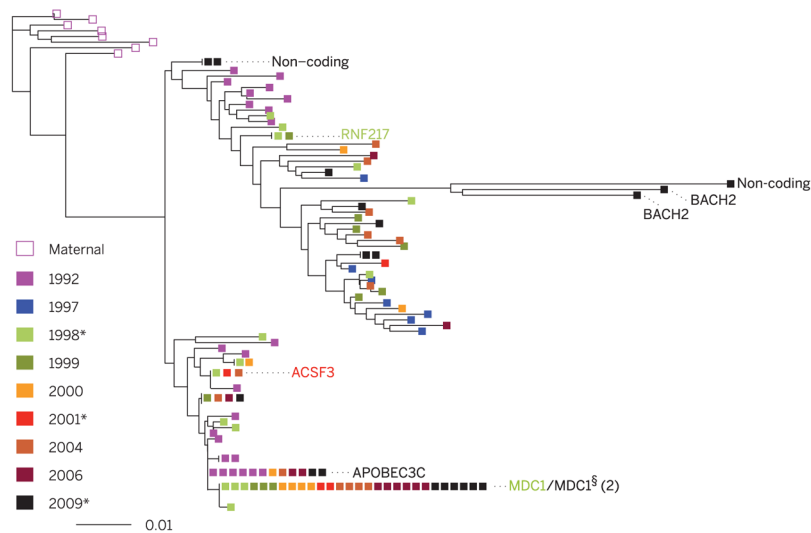


Fig. 2. Phylogenetic relationships between HIV-1 *env* (C2V5 region) genes sampled from participant L1 through time

A neighbor-joining tree was generated using viral gene sequences derived from PBMC DNA from participant L1 by single-genome sequencing, including from this (with integration sites determined) and a previous study (8). Identical sequences are collapsed into horizontal strings of symbols. The year the blood specimen was collected is shown by different colored symbols (see the color key). When an *env* gene was linked to an integration site, a dotted line was drawn from the symbols to the gene name corresponding to that integration site. The text color of the gene name indicates the year that the specimen was collected with a specific integration site, using the same color key. Participant L1 was perinatally infected with HIV-1, and viral sequences from his mother are shown with open boxes. The horizontal scale indicates a 1% difference between sequences. Asterisks after the year indicate cancer-associated genes time points that integration sites were determined, and § indicates genes with GO terms from the list of hallmarks of cancer (25). Three proviruses (indicated in the figure) were found to be defective as a result of hypermutation (assessed using the tool at www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html), resulting in the introduction of stop codons into the *env* genes of all three sequences.

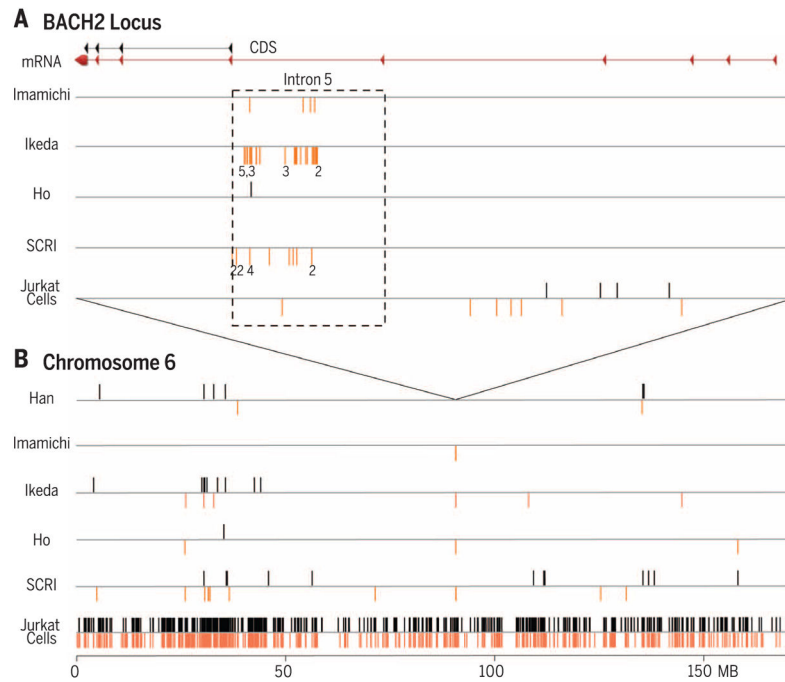


Fig. 3. HIV-1 integration sites in chromosome 6, including the BACH2 locus

(B) (bottom) shows the integration sites mapped in the human chromosome 6 in this (SCRI, Seattle Children's Research Institute) and other studies of individuals sampled while on suppressive ART, including from Han *et al.* (26), Imamichi *et al.* (10), Ikeda *et al.* (9), and Ho *et al.* (27). Data taken from the Wang *et al.* (21) study of virus integrations found 72 hours after infection of Jurkat cells are also shown. The vertical lines represent an integration site over the 172-MB chromosome. Vertical black lines above the horizontal line from each study represent virus integration in the same orientation (5' to 3') as the chromosome. Vertical orange lines below the horizontal indicate virus integrations in the opposite orientation. (A) shows an enlargement of the region of chromosome 6 encoding the BACH2 gene. The exons corresponding to a common mRNA transcript are shown on the red line with red arrowheads. The coding sequences (CDS) of the gene are shown with black arrowheads. The position of intron 5 is enclosed in the alignment below in a broken-line box.