HIV Integration Site Analysis of Cellular Models of HIV Latency with a Probe-Enriched Next-Generation Sequencing Assay

Sara Sunshine,[a] Rory Kirchner,[b] Sami S. Amr,[c] Leandra Mansur,[c] Rimma Shakhbatyan,[c] Michelle Kim,[d,e] Alberto Bosque,[f] Robert F. Siliciano,[d,e] Vicente Planelles,[f] Oliver Hofmann,[b] Shannan Ho Sui,[b] Jonathan Z. Li[a]

Division of Infectious Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA[a]; Bioinformatics Core, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA[b]; Partners HealthCare Personalized Medicine, Cambridge, Massachusetts, USA[c]; Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA[d]; Howard Hughes Medical Institute, Chevy Chase, Maryland, USA[e]; Department of Pathology, University of Utah, Salt Lake City, Utah, USA[f]

ABSTRACT

Antiretroviral therapy (ART) is successful in the suppression of HIV but cannot target and eradicate the latent proviral reservoir. The location of retroviral integration into the human genome is thought to play a role in the clonal expansion of infected cells and HIV persistence. We developed a high-throughput targeted sequence capture assay that uses a pool of HIV-specific probes to enrich Illumina libraries prior to deep sequencing. Using an expanded clonal population of ACH-2 cells, we demonstrate that this sequence capture assay has an extremely low false-positive rate. This assay assessed four cellular models commonly used to study HIV latency and latency-reversing agents: ACH-2 cells, J-Lat cells, the Bcl-2-transduced primary CD4$^+$ model, and the cultured T$_{CM}$ (central memory) CD4$^+$ model. HIV integration site characteristics and genes were compared between these cellular models and to previously reported patient data sets. Across these cellular models, there were significant differences in integration site characteristics, including orientation relative to that of the host gene, the proportion of clonally expanded sites, and the proportion located within genic regions and exons. Despite a greater diversity of minority integration sites than expected in ACH-2 cells, their integration site characteristics consistently differed from those of the other models and from the patient samples. Gene ontology analysis of highly represented genes from the patient samples found little overlap with HIV-containing genes from the cell lines. These findings show that integration site differences exist among the commonly used cellular models of HIV latency and in comparison to integration sites found in patient samples.

IMPORTANCE

Despite the success of ART, currently there is no successful therapy to eradicate integrated proviruses. Cellular models of HIV latency are used to test the efficacy of latency-reversing agents, but it is unclear how well these models reflect HIV integration into the human genome *in vivo*. We have developed a novel probe-based sequence enrichment assay to sequence and analyze integrated HIV. We compared HIV integration site characteristics between four cellular models and to previously described patient data sets. Significant differences were detected in the distribution of HIV integration sites between cellular models of HIV latency and compared to data sets from patient samples. The results from this study have implications for how well these cellular models of HIV infection truly reflect HIV integration *in vivo* and their applicability in drug discovery for novel latency-reversing agents.

The main obstacle to eradicating human immunodeficiency virus (HIV) infection is the establishment of permanently integrated HIV genomes into the host cell chromosomes of long-lived immune cells. Importantly, HIV integration into the human genome is not a completely random process. HIV integration location is influenced by DNA structure and strongly favors transcriptionally active, genic regions of the human genome (1–3). Such targeting is guided in part by cellular factors such as the lens epithelium-derived growth factor (LEDGF)/p75, a chromatin-binding protein that tethers HIV integrase to transcriptionally active regions of the host chromosome (4–6). In addition, nuclear topology and the orientation of HIV integration relative to the host gene may impact transcriptional efficiency and latency (1, 7–9). The evaluation of HIV integration sites has taken on additional importance, given the recent evidence suggesting that specific integration sites associated with cell cycle control are associated with the clonal proliferation of HIV-infected cells, and may be a contributing factor to the persistence of the HIV reservoir despite sustained virologic suppression on antiretroviral therapy (ART)

(10, 11). However, there remains some controversy over the role that these clonally proliferating cells have in contributing to the replication-competent HIV reservoir and in viral rebound after treatment discontinuation (12–14).

Cellular models of HIV latency have become an integral com-

FIG 1 Probe-based sequence enrichment. (A) An overview of the probe-based sequence enrichment process for HIV integration site detection. (B) Compared to our probe library, >97% of the subtype B HIV sequences from the Los Alamos National Laboratory Sequence Database have 3 or fewer mismatches, with its closest matched sequence in the probe library.

ponent of drug discovery in the search for novel latency-reversing agents. Commonly used models include ACH-2 cells (15–17), J-Lat cells (18), the Bcl-2 transduced primary $CD4^+$ model (19–22), and the cultured T_{CM} (central memory) model (23–26). A previous analysis of integration sites among several cellular models showed significant variation in their integration site locations (27). However, it remains unclear if these models adequately reflect the diversity and location of HIV integration that is found within HIV-infected patients. This is reflected in the various and incomplete responses of *in vitro* cellular models to latency-reversing agents compared to responses from patient-derived cells (28).

The study of HIV integration sites historically has been performed with PCR-based methods, which employ a primer complementary to the HIV long terminal repeat (LTR) (10–12, 29). However, the HIV DNA reservoir is highly diverse, as it is derived from the successive archiving of circulating plasma viruses (30). In order to address the amount of viral diversity within the proviral reservoir, we have designed a novel targeted sequence capture-based method for HIV integration site sequencing that uses a pool of capture probes to account for HIV diversity (Fig. 1A). Using our probe-based sequence enrichment method, we have analyzed the degree of clonal expansion, the location, and the orientation of HIV integration sites in four commonly used HIV latency models and compared the findings to those of previously published analyses of HIV integration sites in virologically suppressed patients. The results from this study have implications for how well these cellular models of HIV infection truly reflect HIV integration *in vivo* and their applicability in drug discovery for novel latency-reversing agents.

MATERIALS AND METHODS

Cell lines. ACH-2 cells and J-Lat full-length cells (clone 9.2) were obtained from the NIH AIDS reagent program. ACH-2 clones were obtained by serially diluting cells to a concentration of 0.5 cells per well and were clonally expanded in the presence of the protease inhibitor darunavir and the fusion inhibitor enfuvirtide for 28 days. At the end of the cellular expansion process, approximately half of the wells contained viable cellular populations. DNA was extracted from ACH-2 cells (both bulk cultures grown in the absence of ART and clones as described above), J-Lat cells (31), Bcl-2 transduced primary $CD4^+$ HIV latency model cells (19, 20), and cells from a T_{CM} primary $CD4^+$ HIV latency model (23–25). DNA extractions were performed using the Gentra Puregene cell kit (Qiagen) and purified by isopropanol precipitation.

Probe library. All subtype B, long terminal repeat (LTR) sequences encompassing the U3 region ($n = 573$) were aligned using the Los Alamos National Laboratory Sequence (LANL) database (http://www.hiv.lanl.gov). A 72-bp fragment of the U3 region of the 5′ LTR was selected for probe design as a relatively conserved region. Sequences were clustered using DNACLUST, and 52 consensus probe sequences were identified to cover the diversity of LTR sequences found in LANL. Using this probe set, >97% of the subtype B HIV sequences within the LANL database have 3 or fewer mismatches with its closest matched sequence in the probe library (Fig. 1B; also see Table S1 in the supplemental material).

Sequence capture and deep sequencing. One microgram of DNA was sheared by the Covaris E220 sonicator to 400- to 600-bp fragments. The J-Lat cell line was assessed in duplicate. Illumina library construction was performed using the Beckman Coulter SPRI-TE nucleic acid extractor automated library preparation with Bioo Scientific NEXTflex adapters. Size selection was performed using Beckman-Coulter AMPure XP beads. The libraries then underwent PCR enrichment with Kapa HiFi DNA polymerase. Probe-based sequence enrichment was performed with our biotinylated probe library (xGEN lockdown probe; Integrative DNA Technologies). Probe hybridization was performed at 65°C for 18 h, and hybridized libraries were pulled down using Dynabeads MyOne Streptavidin T1 (Life Technologies). Captured DNA libraries were enriched by PCR enrichment with Kapa HiFi DNA polymerase. After QiaQuick PCR purification, samples were sequenced on the MiSeq platform (Illumina) to generate 300-bp paired-end reads.

Analysis of HIV integration site. Reads were aligned with bwa-mem (32) to hg19 with an artificial HXB2 chromosome, and duplicate alignments were marked using sambamba (https://github.com/lomereiter/sambamba). Unique, nonduplicate human-HXB2 chimeric reads with mapping quality greater than 20 were used to call integration sites at the human-HXB2 breakpoint. Chimeric sequences that mapped to multiple locations in the human genome were excluded from analysis. Source code to perform the integration site analysis is available at https://github.com/hbc/li_hiv. Integration sites within three base pairs were collapsed into the most common integration site. Sites were annotated with UCSC canonical gene annotations. Gene ontology analysis was performed using
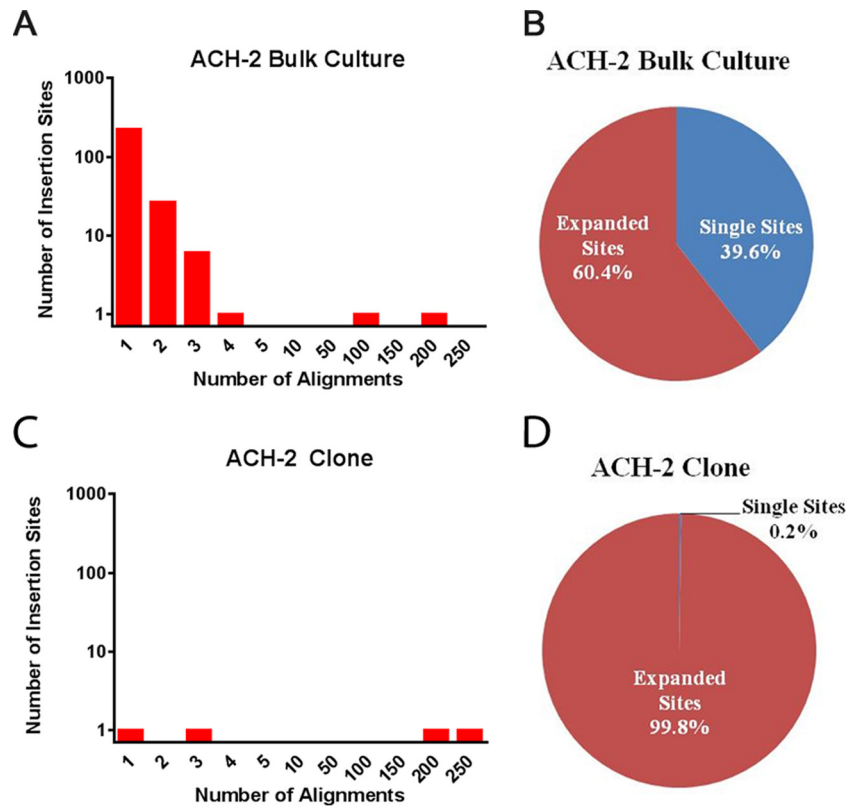
FIG 2 Integration site analysis of ACH-2 cells. (A) The ACH-2 bulk culture contained two major integration sites and over one hundred single integration sites. (B) A total of 39.6% of reads were at sites with a single integration site. Expanded sites represent integration sites with evidence of clonal expansion. (C) Only two minor integration sites were identified in an ACH-2 clone generated by limiting dilution. (D) The single integration site represents 0.2% of all reads from the ACH-2 clone sequencing.

the GEne SeT AnaLysis Toolkit (33). Confirmation of HIV integration sites was performed with PCR amplification using primers for both the HIV LTR and the chromosomal site of interest as well as Sanger population sequencing. Integration site analysis of participant data was obtained from two previously published data sets (10, 11). Median and range values of integration site characteristics (e.g., percent intron) were determined using all applicable reads for each participant. HIV integration site characteristics were compared between samples with chi-square or Fisher's exact test. All genes with viral integrants from patient samples were compared to genes from the cellular models. Genes identified from the cellular models also were compared to a smaller subset of highly represented genes from the two previously published data sets (10, 11). Highly represented genes from the Maldarelli et al. data set were defined by the authors as genes that have three or more different integration sites in the five patients studied or genes with highly expanded clones (10). Highly represented genes from the Wagner et al. study were defined by the authors as genes that had proviral integrants in two of three participants (11).

RESULTS

Integration analysis of ACH-2 and J-Lat cells. HIV integration site analysis was performed on a bulk culture of ACH-2 cells, a T cell line reported to have one integrated proviral copy (15, 16). Illumina sequencing resulted in 560 unique reads found to have a chimeric sequence containing both HIV and human sequence, representing an approximately 500-fold enrichment of HIV-containing reads. The most common HIV integration site in this bulk culture of ACH-2 cells occurred at position 33059398 in chromosome 7 (Chr7), with 200 clonal reads (Fig. 2A). This insertion site

has been described previously using Sanger sequencing (34). However, we found a second common integration at position 130873925 in Chr9 that was detected in 64 clonal reads. To confirm our deep sequencing findings, we successfully PCR amplified and Sanger population sequenced both of these major integration sites using primer pairs where one primer was specific to the human sequence and the other to HIV (Fig. 3A and B). While reported to be a clonal population, we found 247 unique integration sites within the ACH-2 bulk culture, 39.6% of which were single integration sites (Fig. 2A and B). To evaluate whether these minority integration sites represented PCR or sequencing error, ACH-2 cells were cloned to a concentration of 0.5 cells per well by serial dilution and reexpanded in the presence of ART to prevent new rounds of infection. Integration site analysis of the clonal ACH-2 population showed the near elimination of single integration sites (ACH-2 bulk culture versus ACH-2 clone, 39.6% versus 0.2%; $P < 0.01$ by Fisher's test) (Fig. 2), suggesting that this assay has an extremely low false-positive rate and that the minority integration sites detected in the bulk ACH-2 culture are unlikely to represent PCR or sequencing error. The same majority integration site in the bulk culture of ACH-2 cells, 33059398 in Chr7, was also a major integration site in the ACH-2 clone. This site was confirmed subsequently by PCR and Sanger population sequencing.

We also performed integration site analysis on J-Lat cells (clone 9.2). This is a clonal, Jurkat-based cell line with a construct that has a full-length HIV-1 genome deficient in functional *env*
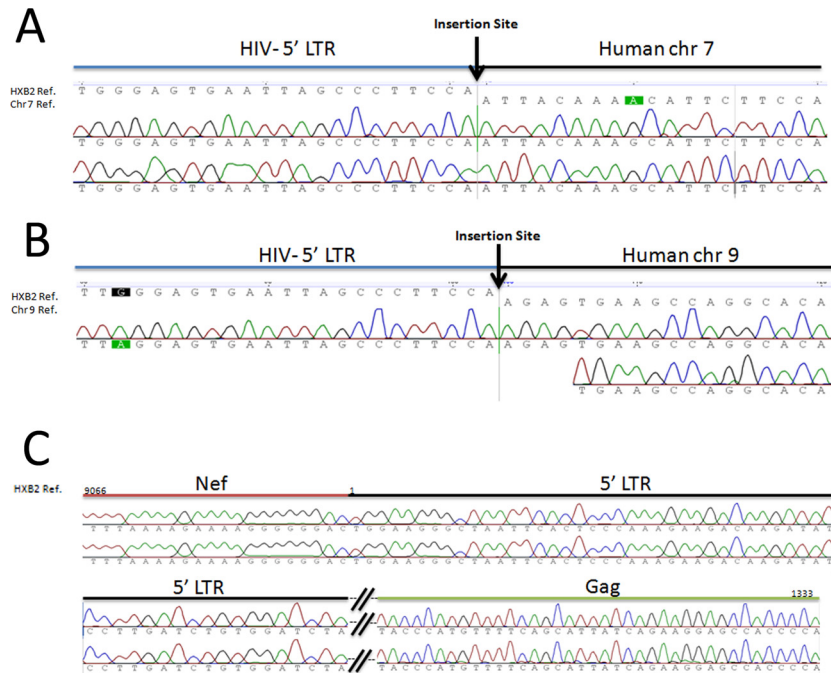
FIG 3 Confirmation of integration sites by PCR and Sanger population sequencing confirmation by PCR and Sanger population sequencing of the ACH-2 bulk culture of HIV integration sites into human chromosome 7 (Chr7; position 33059398) (A) and human chromosome 9 (Chr9; position 130873925) (B). (C) Confirmation of *nef* insertion site by amplification of a *nef*-5′ LTR-*gag* amplicon of the ACH-2 bulk culture. The double vertical line represents 5′ LTR sequences not shown for space reasons.

(31). Illumina sequencing resulted in 55 clonal reads with one insertion site at position 46884365 in Chr19. This integration site was previously reported in Lenasi et al. (35).

Evaluating cellular models of latency. In addition to ACH-2 and J-Lat cells, we also performed integration site sequencing for two other commonly used cell culture models of HIV latency, the Bcl-2 transduced primary CD4$^+$ T cell model (19, 20) and the cultured T$_{CM}$ (primary central memory T cell) model (Fig. 4; also see Table S2 in the supplemental material) (23). Unlike ACH-2 and J-Lat cells, no major HIV integration sites were detected in the Bcl-2 and T$_{CM}$ models (Fig. 4). We compared integration site features across the ACH-2 cells, the Bcl-2 transduced model, and the T$_{CM}$ model and in relation to integration obtained from 8 previously described patients (Table 1) (10, 11). The ACH-2 model was included in the analysis given our finding that these cells harbored a greater than expected diversity of integration sites. Across the three cellular models, there were significant differences in the proportion of clonally expanded integration sites identified within the human genome (ACH-2 versus Bcl-2 transduced model versus T$_{CM}$ model, 60% versus 26% versus 16%; $P < 0.01$ by chi-square test), with ACH-2 cells harboring a particularly high proportion of clonally expanded sites, most of which were found at two particular locations (Fig. 2A). The proportion of expanded sites in the ACH-2 cell line was greater than that of any of the 8 patient samples (60% versus a median of 30% [range, 6 to 47%]) (Table 1). The ACH-2 model also was found to have the highest percentage of integration sites located in exons and in the same orientation as the host gene ($P < 0.01$ by chi-square test for both comparisons) (Table 1). A higher percentage of ACH-2 reads was integrated in the same orientation as the host gene than for any of the patients samples (64% versus a median of 35% [range 23 to 56%]). ACH-2

and Bcl-2 models had a similar proportion of integration sites in genic regions, but a lower proportion was detected in the T$_{CM}$ model (ACH-2 versus Bcl-2 versus T$_{CM}$ model, 88% versus 82% versus 61%; $P < 0.01$ by chi-square test).

Gene overlap and ontology. We also performed a gene overlap and gene ontology analysis of integration sites found in the cellular models of HIV latency and genes with viral integrants identified from Maldarelli et al. and Wagner et al. (10, 11). Compared to all of the genes identified in the patient samples, approximately 20 to 30% of genes also were found in ACH-2 cells, the Bcl-2 transduced model, and the T$_{CM}$ model (Fig. 5A to C). However, there were relatively few genes that overlapped between viral integrants in these cellular models and highly represented genes from the human studies (e.g., BACH2 and MKL2) (Fig. 5D to F). Gene ontology analysis revealed that the ACH-2 cells and the Bcl-2 transduced model cells had viral integration in genes associated with cellular proliferation at percentages similar to those of the two previously published data sets, while the T$_{CM}$ model cells revealed no viral integrants in these genes (Fig. 6). We also compared the genes with viral integrants from the two published studies, and this revealed that 21% of the genes in Wagner et al. also were found in Maldarelli et al.

High efficiency of sequence capture of chimeric reads. Across all of the enriched and sequenced libraries, 94% of all chimeric reads aligned to the 5′ end of HIV (Fig. 7A), reflecting the enrichment of the 5′ LTR by our probe capture method. Analysis of all chimeric reads revealed that while the vast majority of HIV insertions occurred at or near the termini (base pair 1) of the HIV 5′ LTR, we identified a number of insertion sites involving truncated U3 regions of the LTR or within the distal portion of *nef* (Fig. 7B). Subsequent PCR and Sanger population sequencing of selected
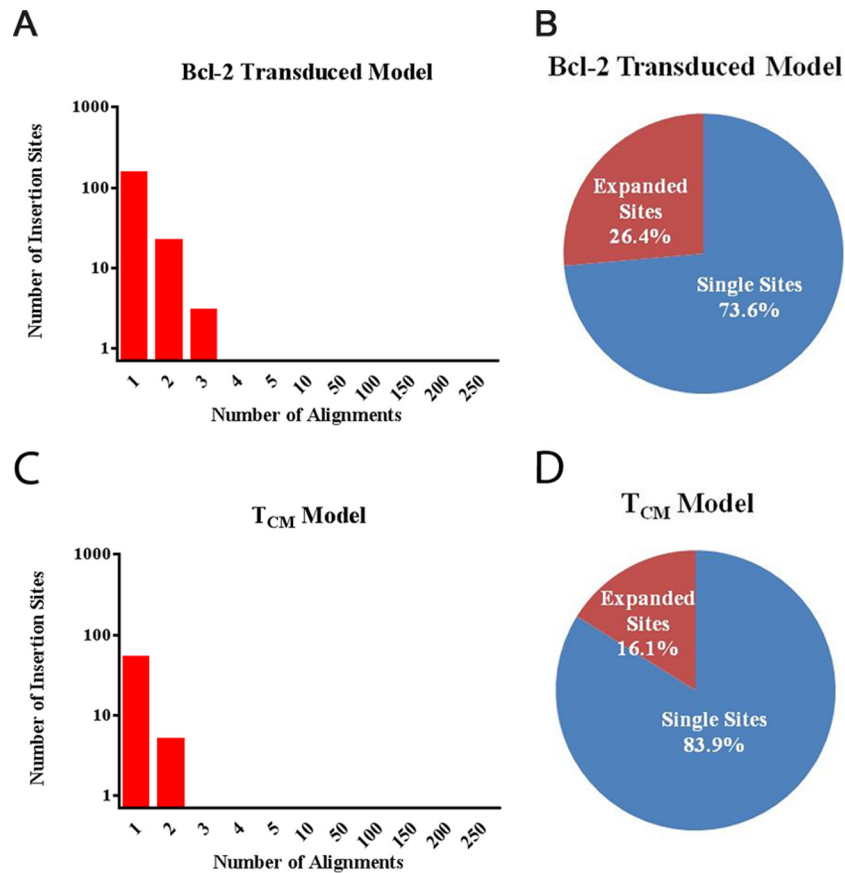
FIG 4 Integration site analysis of the Bcl-2 transduced CD4$^+$ model and the T$_{CM}$ model. (A and B) The Bcl-2 transduced model contained no dominant integration site, and only 26.4% of reads were at expanded sites. (C and D) The T$_{CM}$ model also contained only minor integration sites and 83.9% single integration sites.

sites confirmed the presence of truncated U3 and *nef* integration sites into the ACH-2 genome (Fig. 3C).

DISCUSSION

In this study, we have created a high-throughput assay that enriches Illumina deep sequencing libraries for the HIV 5′ LTR with a diverse biotinylated probe library. Using an expanded clonal population of ACH-2 cells, we demonstrate that this sequence capture assay has an extremely low false-positive rate. This assay

was used to assess four cellular models commonly used to study HIV latency and latency-reversing agents: ACH-2 cells, J-Lat cells, the Bcl-2 transduced primary CD4$^+$ model cells, and the cultured T$_{CM}$ model. There were significant differences in integration site characteristics, including the proportion of clonally expanded sites, proportion of sites located within genic regions and exons, and the orientation of the integrated HIV relative to the host gene among all of these cellular models. ACH-2 integration site characteristics consistently differed from those of the other models

TABLE 1 Integration site analysis

Sample	Unique integration sites[b] (%)		Genomic region (%)		Genomic annotation[c] (%)		HIV integration orientation relative to human genes (%)	
	Single site	Expanded site	Intergenic	Genic	Intron	Exon	Same	Opposite
ACH-2 cells	40	60	12	88	86	14	64	36
Bcl-2 transduced model	74	26	18	82	98	2	52	48
T$_{CM}$ model	84	16	39	61	97	3	38	62
Patients[a]	70 (53–94)	30 (6–47)	18 (13–56)	82 (44–87)	93 (58–98)	7 (2–42)	35 (23–56)	65 (44–77)

[a] Median values and ranges (in parentheses) for sample sets from Maldarelli et al. (10) and Wagner et al. (11).
[b] The percent of unique integration sites that were expanded clones represents integration sites with greater than one alignment, while single integration sites contained only one alignment.
[c] Genomic annotations only include alignments that were introns or exons, as determined by the UCSC canonical gene annotations; this analysis excludes integration sites in which there was an overlap between introns and exons in the human genome.
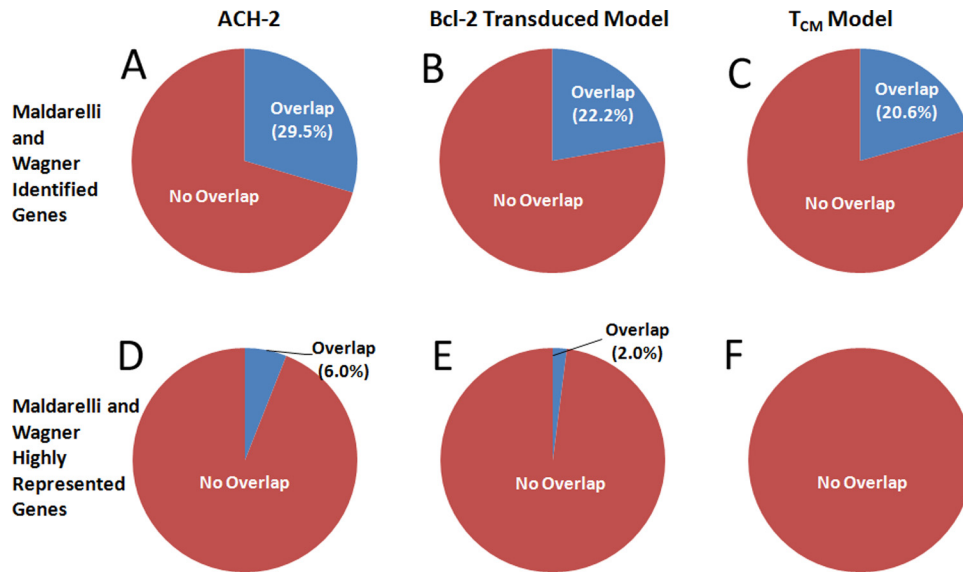
FIG 5 Gene overlap between previously described patient samples and cellular models. (A to C) Comparison between the ACH-2 model, the Bcl-2 CD4$^+$ transduced model, and the T$_{CM}$ CD4$^+$ model with all human genes containing HIV proviruses identified by Maldarelli et al. and Wagner et al. (D to F) Comparison between the ACH-2 model, the Bcl-2 CD4$^+$ transduced model, and the T$_{CM}$ CD4$^+$ model with only the highly represented genes as defined by Maldarelli et al. and Wagner et al.

and from the patient samples. Gene ontology analysis of highly represented genes from the patient samples found little overlap with HIV-containing genes from the cell lines. Finally, we found that the 5′ HIV integration junction does not always occur at the exact terminus of the HIV LTR and that both truncated U3 regions and *nef* additions appear to be present.

Targeted probe-based enrichment has been used for sequence enrichment of HIV-1-derived lentiviral vectors (36), other viruses, such as human herpesviruses (37), and Merkel cell polyomavirus (38), as well as the detection of human genetic mutations (39, 40). While Ustek et al. were able to enrich deep sequencing libraries for the targeted lentiviral vector, over 90% of fragments that were pulled down were not complementary to the target region (36). Rather than tiling across the target genome for sequence detection, as was done in these prior studies, we designed a library of diverse probes in the proximal HIV 5′ LTR. This allowed for sequence enrichment of the HIV-human junction sites for integration site analysis. Most previous studies of HIV integration sites used PCR-based methods that may bias integration site detection to those sequences that closely match the PCR primer sequence (10–12, 29). Our diverse probe library seeks to address this issue by avoiding the use of a single primer complementary to HIV. Using clonally derived ACH-2 cells, we show that the false-positive rate of this assay is extremely low. While this method does have an approximately 500-fold enrichment of HIV-specific sequences, further optimization is necessary for detection of rare HIV integration events, as seen in patient samples.

ACH-2 cells were produced by infection of A3.01 cells, serially diluted to 0.2 cells per well, and reexpanded (15, 16). Clouse et al. reported that most of the infected A3.01 cells that survived HIV infection are CD4 negative, and Southern blot hybridization revealed one integrated copy of HIV per cell (15). Using inverse PCR, it was found that the ACH-2 integration site was in chromosome 7p15 (34). While this also was determined to be a major integration site in the bulk culture of ACH-2 cells in our study, we

uncovered the previously undescribed diversity of integration sites within this cell line. Integration site analysis of the bulk ACH-2 cell line found that there was a second major integration site as well as a large number of minority integration sites. Using a clonally expanded population of ACH-2 cells, we show that these minority integration sites likely are not due to false positives and indicate that low-level viral replication is indeed occurring in ACH-2 cell cultures. The presence of two major integration sites within the clonal ACH-2 population likely represents two founder populations present in the limiting dilution culture, although we could not rule out the presence of a cellular population with two integration sites. J-Lat cells originally were constructed by infection of Jurkat cells with an HIV-R7/E-/green fluorescent protein (GFP) construct (41). The integration site of this clonal population (J-Lat clone 9.2) was reported to be in the PP5 gene of chromosome 19 (35). Using this novel method, we confirmed the integration site in this clonal population to be position 46884365 in Chr19.

Cellular models of latency are necessary for the screening and analysis of putative HIV latency-reversing agents (17, 21, 22, 26). We performed integration site analysis of four commonly used cellular models of HIV latency and compared how well they reflect HIV integration site characteristics in patient samples. The Bcl-2 transduced model was created by activation of primary human CD4$^+$ T cells and transduction with the EB-FLV vector containing the BCL-2 gene (19, 20). Transduced cells then were infected with reporter virus NL4-3-Δ6-drEGFP and sorted according to GFP expression to isolate latently infected cells. The T$_{CM}$ model was created by the isolation of naive CD4$^+$ T cells and induced to produce cultured T$_{CM}$ cells (23, 24). Cells subsequently were infected with replication-competent HIV-1 and treated with suppressive antiretroviral drugs, and latency was assessed by CD4 T cell sorting to isolate uninfected and latently infected cells (23). We found significant differences in integration site characteristics between the cellular models with ACH-2 cells. While ACH-2 cell
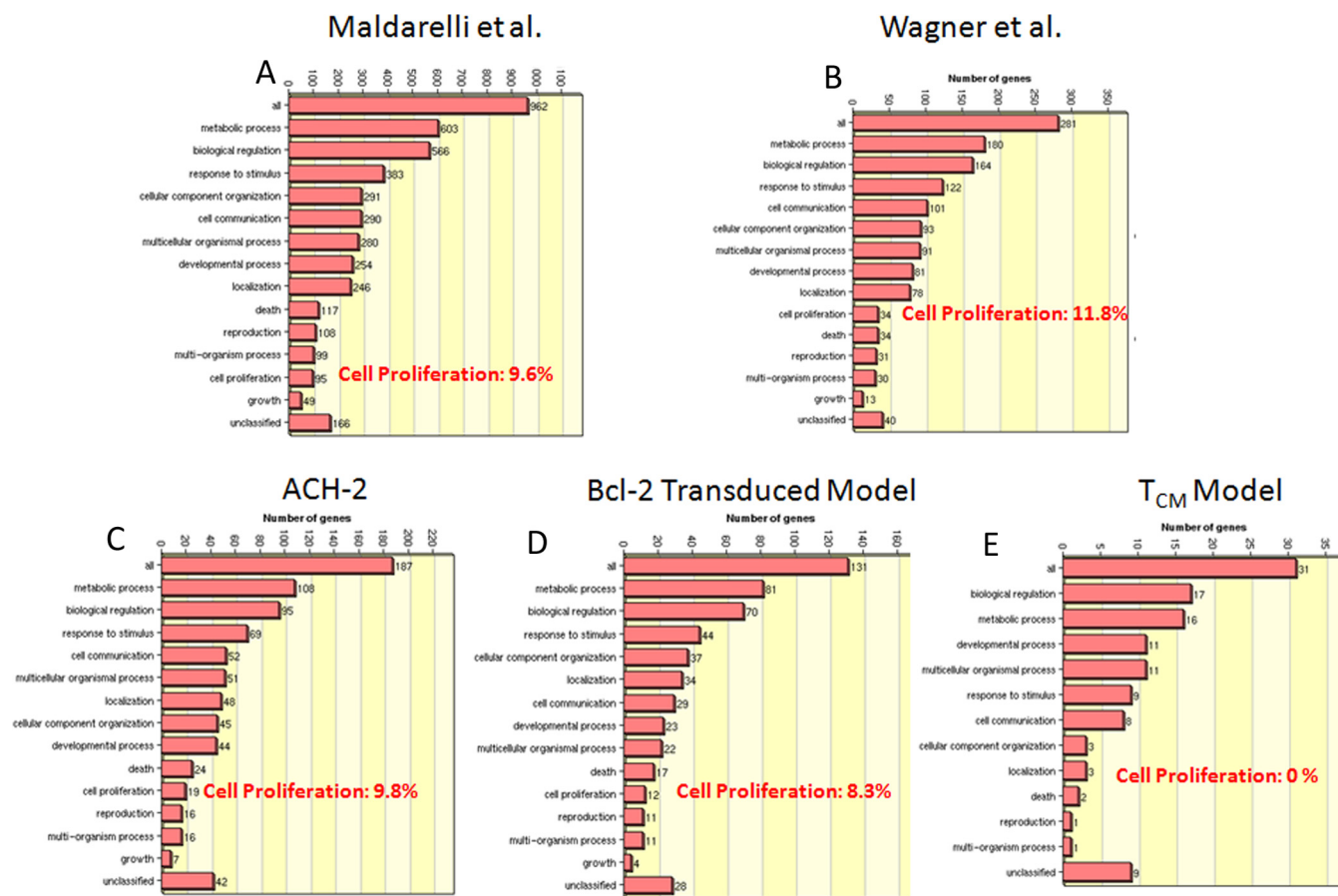
FIG 6 Gene ontology analysis using the GEne SeT AnaLysis Toolkit revealed the percentage of genes associated with cell proliferation for each library.

lines are readily available and easy to use, they are unlikely to accurately reflect the latently infected cellular environment. Despite a greater than expected diversity of minority integration sites detected in ACH-2 cells, their integration site characteristics consistently differed from those of the other models and from the patient samples. The Bcl-2 and T_{CM} cell models did not contain any major clonally expanded sites, and the integration site characteristics (e.g., percent integrated in genic regions, exons, and host cell orientation) were within the range of those found in the patient samples. However, gene ontology analysis of highly represented genes from the patient samples found relatively little overlap with HIV-containing genes from the cell lines. Overall, these results show that some cellular models of latency have insertion site characteristics similar to those found in participant samples, but that substantial differences may remain in the genes that contain proviruses. In the cellular models, no insertion sites were detected within BACH2 or MKL2 genes. One explanation is that the selective advantages of HIV integration into genes associated with clonal proliferation are blunted in cell lines that previously were immortalized and/or where sustained cellular proliferation is absent postinfection. This also may explain variations in the response of cellular models of HIV latency to stimulating agents and why response may be incomplete compared to that of patient-derived cells (28).

It is generally assumed that the human-HIV integration junction site is located exactly at the proviral terminus (base pair 1 of HIV viral DNA) (42). While the vast majority of the identified chimeric reads occurred at the terminus, we also found insertion sites where human sequence is attached to truncated U3 regions of the LTR or the distal portion of *nef*. These findings were confirmed by PCR and Sanger population sequencing, including the sequencing of a *nef-gag* amplicon (Fig. 3C). We postulate that a truncated U3 region is not lethal for viral replication, as RNA transcription occurs at the R region of the 5′ LTR and essential binding sites for transcriptional cofactors are not generally located at the proximal portion of the U3 region (i.e., NF-κB, SP1, and other cofactor binding sites are found primarily at or after HXB2 bp 350 in the U3 region). In addition, the U3 region of the proviral 5′ LTR is not included as part of the viral RNA progeny, as it is derived from the U3 region of the 3′ LTR at the time of reverse transcription. We hypothesize that the presence of a substantial number of *nef* sequences at the HIV integration junction site is a result of inadvertent partial *nef* duplication during the reverse transcriptase step that uses the U3 region of the 3′ LTR as the template for the proximal 5′ end (43). This finding only accounts for a relatively small proportion of integration events and still requires confirmation in patient samples, but it may have implications for primer/probe design and the accurate identification of HIV integration sites.

In this report, we demonstrate the efficiency of a targeted sequence capture-based next-generation sequencing method for HIV integration site analysis. Using this assay, we have evaluated
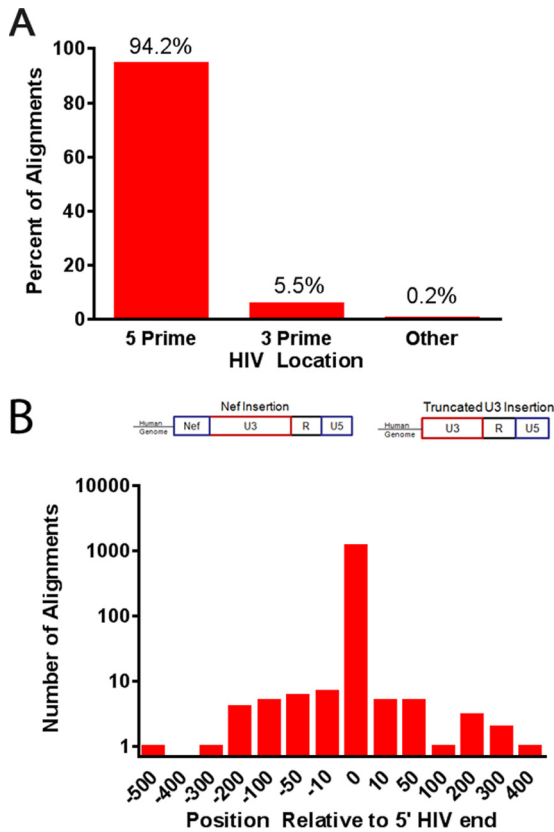
FIG 7 Efficiency of sequence enrichment and insertion site analysis. (A) Enrichment by sequence capture is highly efficient for the 5′ end of HIV. (B) While most chimeric alignments have an HIV insertion at the viral terminus (base pair 1), a subset of sequences showed truncated U3 or *nef* sequence at the human-HIV junction site. We postulate that the latter scenario is a result of inadvertent duplication of a fragment of *nef* during the U3 duplication process based off the *nef*-3′ LTR template.

how well four commonly used cellular models of HIV latency reflect HIV integration site characteristics found in patient samples. A deeper understanding of the HIV proviral environment within these cellular models will be crucial in evaluating how well they model HIV integration *in vivo* and the generalizability of their findings on potential drugs targeting the latent viral reservoir.

ACKNOWLEDGMENTS

FUNDING INFORMATION

REFERENCES

1. **Lewinski MK, Bisgrove D, Shinn P, Chen H, Hoffmann C, Hannenhalli S, Verdin E, Berry CC, Ecker JR, Bushman FD.** 2005. Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. J Virol **79:**6610–6619. http://dx.doi.org/10.1128/JVI.79.11.6610-6619.2005.
2. **Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F.** 2002. HIV-1 integration in the human genome favors active genes and local hotspots. Cell **110:**521–529. http://dx.doi.org/10.1016/S0092-8674(02)00864-4.
3. **Wang GP, Ciuffi A, Leipzig J, Berry CC, Bushman FD.** 2007. HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. Genome Res **17:**1186–1194. http://dx.doi.org/10.1101/gr.6286907.
4. **Marshall HM, Ronen K, Berry C, Llano M, Sutherland H, Saenz D, Bickmore W, Poeschla E, Bushman FD.** 2007. Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. PLoS One **2:**e1340. http://dx.doi.org/10.1371/journal.pone.0001340.
5. **Kvaratskhelia M, Sharma A, Larue RC, Serrao E, Engelman A.** 2014. Molecular mechanisms of retroviral integration site selection. Nucleic Acids Res **42:**10209–10225. http://dx.doi.org/10.1093/nar/gku769.
6. **Maertens G, Cherepanov P, Pluymers W, Busschots K, De Clercq E, Debyser Z, Engelborghs Y.** 2003. LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. J Biol Chem **278:**33528–33539. http://dx.doi.org/10.1074/jbc.M303594200.
7. **Shan L, Yang HC, Rabi SA, Bravo HC, Shroff NS, Irizarry RA, Zhang H, Margolick JB, Siliciano JD, Siliciano RF.** 2011. Influence of host gene transcription level and orientation on HIV-1 latency in a primary-cell model. J Virol **85:**5384–5393. http://dx.doi.org/10.1128/JVI.02536-10.
8. **Han Y, Lin YB, An W, Xu J, Yang HC, O'Connell K, Dordai D, Boeke JD, Siliciano JD, Siliciano RF.** 2008. Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough. Cell Host Microbe **4:**134–146. http://dx.doi.org/10.1016/j.chom.2008.06.008.
9. **Marini B, Kertesz-Farkas A, Ali H, Lucic B, Lisek K, Manganaro L, Pongor S, Luzzati R, Recchia A, Mavilio F, Giacca M, Lusic M.** 2015. Nuclear architecture dictates HIV-1 integration site selection. Nature **521:**227–231. http://dx.doi.org/10.1038/nature14226.
10. **Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, Spindler J, Ferris AL, Mellors JW, Kearney MF, Coffin JM, Hughes SH.** 2014. HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. Science **345:**179–183.
11. **Wagner TA, McLaughlin S, Garg K, Cheung CY, Larsen BB, Styrchak S, Huang HC, Edlefsen PT, Mullins JI, Frenkel LM.** 2014. HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. Science **345:**570–573.
12. **Cohn LB, Silva IT, Oliveira TY, Rosales RA, Parrish EH, Learn GH, Hahn BH, Czartoski JL, McElrath MJ, Lehmann C, Klein F, Caskey M, Walker BD, Siliciano JD, Siliciano RF, Jankovic M, Nussenzweig MC.** 2015. HIV-1 integration landscape during latent and active infection. Cell **160:**420–432. http://dx.doi.org/10.1016/j.cell.2015.01.020.
13. **Simonetti FR, Sobolowski MD, Hill S, Shao W, Fyne E, Wu X, Coffin J, Hughes SH, Mellors J, Maldarelli F.** 2015. Residual viremia caused by clonally expanded tumor-infiltrating CD4$^+$ cells. Abstr Conf Retrovir Opportunistic Infect, abstr 105.
14. **Kearney MF, Shao W, Gandhi RT, Keele BF, Li JZ.** 2015. Identifying HIV variants that rebound after treatment interruption. Abstr Conf Retrovir Opportunistic Infect, abstr 378.
15. **Clouse KA, Powell D, Washington I, Poli G, Strebel K, Farrar W, Barstad P, Kovacs J, Fauci AS, Folks TM.** 1989. Monokine regulation of human immunodeficiency virus-1 expression in a chronically infected human T cell clone. J Immunol **142:**431–438.
16. **Folks TM, Clouse KA, Justement J, Rabson A, Duh E, Kehrl JH, Fauci AS.** 1989. Tumor necrosis factor alpha induces expression of human immunodeficiency virus in a chronically infected T-cell clone. Proc Natl Acad Sci U S A **86:**2365–2368. http://dx.doi.org/10.1073/pnas.86.7.2365.
17. **Choi BS, Lee HS, Oh YT, Hyun YL, Ro S, Kim SS, Hong KJ.** 2010. Novel histone deacetylase inhibitors CG05 and CG06 effectively reactivate la-

tently infected HIV-1. AIDS **24**:609–611. http://dx.doi.org/10.1097/QAD
.0b013e328333bfa1.

18. **Venkatachari NJ, Zerbato JM, Jain S, Mancini AE, Chattopadhyay A,
Sluis-Cremer N, Bar-Joseph Z, Ayyavoo V.** 2015. Temporal transcriptional response to latency reversing agents identifies specific factors regulating HIV-1 viral transcriptional switch. Retrovirology **12**:85. http://dx
.doi.org/10.1186/s12977-015-0211-3.

19. **Kim M, Hosmane NN, Bullen CK, Capoferri A, Yang HC, Siliciano JD,
Siliciano RF.** 2014. A primary CD4(+) T cell model of HIV-1 latency
established after activation through the T cell receptor and subsequent
return to quiescence. Nat Protoc **9**:2755–2770. http://dx.doi.org/10.1038
/nprot.2014.188.

20. **Yang HC, Xing S, Shan L, O'Connell K, Dinoso J, Shen A, Zhou Y,
Shrum CK, Han Y, Liu JO, Zhang H, Margolick JB, Siliciano RF.** 2009.
Small-molecule screening using a human primary cell model of HIV latency identifies compounds that reverse latency without cellular activation. J Clin Investig **119**:3473–3486.

21. **Shan L, Xing S, Yang HC, Zhang H, Margolick JB, Siliciano RF.** 2014.
Unique characteristics of histone deacetylase inhibitors in reactivation of
latent HIV-1 in Bcl-2-transduced primary resting CD4+ T cells. J Antimicrob Chemother **69**:28–33. http://dx.doi.org/10.1093/jac/dkt338.

22. **Xing S, Bhat S, Shroff NS, Zhang H, Lopez JA, Margolick JB, Liu JO,
Siliciano RF.** 2012. Novel structurally related compounds reactivate latent
HIV-1 in a bcl-2-transduced primary CD4+ T cell model without inducing global T cell activation. J Antimicrob Chemother **67**:398–403. http:
//dx.doi.org/10.1093/jac/dkr496.

23. **Martins LJ, Bonczkowski P, Spivak AM, De Spiegelaere W, Novis CL,
DePaula-Silva AB, Malatinkova E, Typsteen W, Bosque A, Vanderkerckhove L, Planelles V.** 2015. Modeling HIV-1 latency in primary T cells
using a replication-competent virus. AIDS Res Hum Retrovir **32**:187–193.

24. **Bosque A, Planelles V.** 2011. Studies of HIV-1 latency in an ex vivo model
that uses primary central memory T cells. Methods **53**:54–61. http://dx
.doi.org/10.1016/j.ymeth.2010.10.002.

25. **Bosque A, Planelles V.** 2009. Induction of HIV-1 latency and reactivation
in primary memory CD4+ T cells. Blood **113**:58–65. http://dx.doi.org/10
.1182/blood-2008-07-168393.

26. **Wei DG, Chiang V, Fyne E, Balakrishnan M, Barnes T, Graupe M,
Hesselgesser J, Irrinki A, Murry JP, Stepan G, Stray KM, Tsai A, Yu H,
Spindler J, Kearney M, Spina CA, McMahon D, Lalezari J, Sloan D,
Mellors J, Geleziunas R, Cihlar T.** 2014. Histone deacetylase inhibitor
romidepsin induces HIV expression in CD4 T cells from patients on suppressive antiretroviral therapy at concentrations achieved by clinical dosing. PLoS Pathog **10**:e1004071. http://dx.doi.org/10.1371/journal.ppat
.1004071.

27. **Sherrill-Mix S, Lewinski MK, Famiglietti M, Bosque A, Malani N,
Ocwieja KE, Berry CC, Looney D, Shan L, Agosto LM, Pace MJ,
Siliciano RF, O'Doherty U, Guatelli J, Planelles V, Bushman FD.** 2013.
HIV latency and integration site placement in five cell-based models. Retrovirology **10**:90. http://dx.doi.org/10.1186/1742-4690-10-90.

28. **Spina CA, Anderson J, Archin NM, Bosque A, Chan J, Famiglietti M,
Greene WC, Kashuba A, Lewin SR, Margolis DM, Mau M, Ruelas D,
Saleh S, Shirakawa K, Siliciano RF, Singhania A, Soto PC, Terry VH,
Verdin E, Woelk C, Wooden S, Xing S, Planelles V.** 2013. An in-depth
comparison of latent HIV-1 reactivation in multiple cell model systems
and resting CD4+ T cells from aviremic patients. PLoS Pathog **9**:e1003834.
http://dx.doi.org/10.1371/journal.ppat.1003834.

29. **Han Y, Lassen K, Monie D, Sedaghat AR, Shimoji S, Liu X, Pierson TC,
Margolick JB, Siliciano RF, Siliciano JD.** 2004. Resting CD4+ T cells
from human immunodeficiency virus type 1 (HIV-1)-infected individuals
carry integrated HIV-1 genomes within actively transcribed host genes. J
Virol **78**:6122–6133. http://dx.doi.org/10.1128/JVI.78.12.6122-6133
.2004.

30. **Lambotte O, Chaix ML, Gubler B, Nasreddine N, Wallon C, Goujard C,
Rouzioux C, Taoufik Y, Delfraissy JF.** 2004. The lymphocyte HIV reservoir in patients on long-term HAART is a memory of virus evolution.
AIDS **18**:1147–1158. http://dx.doi.org/10.1097/00002030-200405210
-00008.

31. **Jordan A, Bisgrove D, Verdin E.** 2003. HIV reproducibly establishes a
latent infection after acute infection of T cells in vitro. EMBO J **22**:1868–
1877. http://dx.doi.org/10.1093/emboj/cdg188.

32. **Li H.** 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio.GN].

33. **Zhang B, Kirov S, Snoddy J.** 2005. WebGestalt: an integrated system for
exploring gene sets in various biological contexts. Nucleic Acids Res **33**:
W741–W748. http://dx.doi.org/10.1093/nar/gki475.

34. **Ishida T, Hamano A, Koiwa T, Watanabe T.** 2006. 5′ Long terminal
repeat (LTR)-selective methylation of latently infected HIV-1 provirus
that is demethylated by reactivation signals. Retrovirology **3**:69. http://dx
.doi.org/10.1186/1742-4690-3-69.

35. **Lenasi T, Contreras X, Peterlin BM.** 2008. Transcriptional interference
antagonizes proviral gene expression to promote HIV latency. Cell Host
Microbe **4**:123–133. http://dx.doi.org/10.1016/j.chom.2008.05.016.

36. **Ustek D, Sirma S, Gumus E, Arikan M, Cakiris A, Abaci N, Mathew J,
Emrence Z, Azakli H, Cosan F, Cakar A, Parlak M, Kursun O.** 2012. A
genome-wide analysis of lentivector integration sites using targeted sequence capture and next generation sequencing technology. Infect Genet
Evol **12**:1349–1354. http://dx.doi.org/10.1016/j.meegid.2012.05.001.

37. **Depledge DP, Palser AL, Watson SJ, Lai IY, Gray ER, Grant P, Kanda
RK, Leproust E, Kellam P, Breuer J.** 2011. Specific capture and whole-genome sequencing of viruses from clinical samples. PLoS One **6**:e27805.
http://dx.doi.org/10.1371/journal.pone.0027805.

38. **Duncavage EJ, Magrini V, Becker N, Armstrong JR, Demeter RT, Wylie
T, Abel HJ, Pfeifer JD.** 2011. Hybrid capture and next-generation sequencing identify viral integration sites from formalin-fixed, paraffin-embedded tissue. J Mol Diagn **13**:325–333. http://dx.doi.org/10.1016/j
.jmoldx.2011.01.006.

39. **Gao R, Liu Y, Gjesing AP, Hollensted M, Wan X, He S, Pedersen O, Yi
X, Wang J, Hansen T.** 2014. Evaluation of a target region capture sequencing platform using monogenic diabetes as a study-model. BMC
Genet **15**:13. http://dx.doi.org/10.1186/1471-2156-15-13.

40. **Stoddard JL, Niemela JE, Fleisher TA, Rosenzweig SD.** 2014. Targeted
NGS: a cost-effective approach to molecular diagnosis of PIDs. Front Immunol **5**:531.

41. **Jordan A, Defechereux P, Verdin E.** 2001. The site of HIV-1 integration
in the human genome determines basal transcriptional activity and response to Tat transactivation. EMBO J **20**:1726–1738. http://dx.doi.org
/10.1093/emboj/20.7.1726.

42. **Craigie R, Bushman FD.** 2012. HIV DNA integration. Cold Spring Harb
Perspect Med **2**:a006890.

43. **Coffin JM, Hughes SH, Varmus HE (ed).** 1997. Reverse transcription of
the viral genome in vivo. Cold Spring Harbor Laboratory Press, Cold
Spring Harbor, NY.