Journal of Virus Eradication 8 (2022) 100306

Contents lists available at ScienceDirect

Journal of Virus Eradication

journal homepage: www.sciencedirect.com/journal/journal-of-virus-eradication

Original research

Intact provirus and integration sites analysis in acute HIV-1 infection and changes after one year of early antiviral therapy

Gabriella Rozera [a], Giuseppe Sberna [a], Giulia Berno [a], Cesare Ernesto Maria Gruber [a], Emanuela Giombini [a], Pietro Giorgio Spezia [b], Nicoletta Orchi [c], Vincenzo Puro [c], Annalisa Mondi [d], Enrico Girardi [e], Francesco Vaia [f], Andrea Antinori [d], Fabrizio Maggi [a], Isabella Abbate [a,*]

[a] Laboratory of Virology, National Institute for Infectious Diseases Lazzaro Spallanzani IRCCS, Rome, Italy
[b] Department of Translational Research, Retrovirus Center, University of Pisa, Pisa, Italy
[c] AIDS Referral Center, National Institute for Infectious Diseases Lazzaro Spallanzani IRCCS, Rome, Italy
[d] Clinical and Research Infectious Department, National Institute for Infectious Diseases Lazzaro Spallanzani IRCCS, Rome, Italy
[e] Scientific Direction, National Institute for Infectious Diseases Lazzaro Spallanzani IRCCS, Rome, Italy
[f] General Direction, National Institute for Infectious Diseases Lazzaro Spallanzani IRCCS, Rome, Italy

ARTICLE INFO

Keywords:
HIV
Acute infection
Provirus
Integration site

ABSTRACT

Background and objectives: HIV-1 provirus integration in host genomes provides a lifelong reservoir of virally infected cells. Although not able to generate viral progeny, the expression of defective proviruses has been associated with activation. Provirus integration may influence host gene transcription and shifts may occur during disease progression or antiretroviral therapy (ART). The study aimed to analyze intact/defective provirus and sites of provirus integration in acute infections: changes after 48 weeks of early therapy were also evaluated.
Methods: DNA from peripheral blood lymphomonocytes of 8 acute HIV-1 infections at serodiagnosis (T0) and after 48 weeks of therapy (T1) was used to quantify intact and defective provirus by digital-droplet PCR and to analyze provirus integration sites, by next-generation sequencing of libraries derived from ligation-mediated PCR.
Results: A high variability in the amount of intact proviral DNA was observed at both T0 and T1, in the different subjects. Although the ratio of intact/total proviral HIV-1 DNA did not dramatically change between T0 (8.05%) and T1 (9.34%), after early therapy both intact and total HIV-1 DNA declined significantly, $p = 0.047$ and $p = 0.008$, respectively. The median number of different (IQR) integration sites in human chromosomes/subject was 5 (2.25-13.00) at T0 and 4 (3.00-6.75) at T1. Of all the integration sites observed at T1, 64% were already present at T0. Provirus integration was observed in introns of transcriptionally active genes. Some sites of integration, among which the most represented was in the neuregulin 2 gene, were shared by different patients, together with the orientation of the insertion. Provirus integration was also observed in intergenic regions, with median (IQR) % of 15.13 (6.81-21.40) at T0 and 18.46 (8.98-22.18) at T1 of all read matches.
Conclusions: In acute HIV-1 infection, the amount of intact proviral DNA in peripheral lymphomonocytes did not exceed 10% of total HIV-1 DNA, a percentage that was not substantially changed by early administrated ART. Provirus displayed a relatively small number of recurrent integration sites in introns of transcriptionally active genes, mainly related to cell-cycle control. Consideration should be given to therapeutic strategies able to target the cells harboring defective proviruses, that are not reached by conventional antiviral drugs, these potentially also impacting on replicative competent integrated provirus.

1. Introduction

The integration of the retrotranscribed genome (i.e. the proviral DNA) into the host cell chromosomes is an essential step in the viral replication cycle of all retroviruses, including HIV-1. Once integrated, the proviral DNA is replicated with cellular DNA, establishing a lifelong

* Corresponding author.
E-mail address: isabella.abbate@inmi.it (I. Abbate).

https://doi.org/10.1016/j.jve.2022.100306
Received 22 November 2022; Accepted 7 December 2022
Available online 10 December 2022

reservoir of virally infected cells.[1,2] The integration process is mediated by the interaction between viral and host factors, and it is almost random, also strongly favored in active transcription units, and able to promote efficient viral gene expression. Integration is mediated by the virus-encoded Integrase (IN) protein, which is introduced into cells during infection, along with reverse transcriptase, the viral RNA, and other proteins as a part of the viral core. The HIV-1 integration machinery interacts with many host factors including nuclear trafficking and pore proteins during nuclear entry, histones for initial target capture, and DNA repair proteins in the completion of the DNA joining steps.[3,4] During the retrotranscription process, deletions often occur during minus-strand synthesis, before the second strand transfer event, leading to a consistent amount of defective proviral DNA.[5] Only the intact integrated proviral DNA can generate viral progeny, however, it has been reported that defective proviruses may be translationally competent, producing HIV-1 *gag* and *nef* proteins.[6-8] The HIV-1 RNA transcripts expressed from these defective proviruses may trigger innate immunity and help in explaining persistent immune activation, also in effectively antiretroviral (ART)-treated patients.[9] Intact/defective proviruses are under distinct selective pressure.[10,11] The sites of provirus integration may influence HIV gene transcription of both intact and defective provirus and may change during disease progression and antiviral therapy.[12,13] The present study aimed to characterize proviral DNA in peripheral blood lymphomonocytes of acute HIV-1 infections, in terms of both ratios of intact/defective proviral genomes and site of integration in human chromosomes. Changes 48 weeks after early therapy were also evaluated.

2. Materials and methods

2.1. Patients and virological characterization

Eight acute HIV-infected patients were participants from the SIREA cohort (SIndrome REtrovirale Acuta), an observational longitudinal cohort of primary HIV infection established at the National Institute for Infectious Diseases Spallanzani (INMI) in Rome, composed of all adult individuals with a primary HIV-1 infection (PHI) diagnosis performed at the INMI Virology Laboratory. From this cohort, we have identified 8 individuals who had a positive HIV Ab/Ag (Architect, Abbott) and negative ImmunoBlot assay (Geenius, Bio-Rad) with stored blood peripheral mononuclear cell (PBMC) samples at serodiagnosis (T0) and at 1 year of ART (T1) administrated after a median (IQR) of 2.5 (1.25–5.75) days after serodiagnosis. Emtricitabine (FTC)/tenofovir (TDF) associated with either darunavir (DRV)/ritonavir (RTV) and raltegravir (RGV) or dolutegravir (DLG) regimens were administrated depending on the availability of the drugs. Plasma HIV-1 RNA was measured with a commercial assay (Abbott Real-Time HIV-1 assay, Abbott Molecular, Inc., Des Plaines, IL, USA) at T0 and week 2, 4, 8, 12, 24, 36, and 48 and thereafter every six months. HIV- 1 subtype was determined by molecular phylogeny. To improve the accuracy of the analysis for recombinant forms, RDP3 software was used (http://web.cbio.uct.ac.za/~darren/rdp.html): moreover, subtype classification was confirmed by the HIV Resistance database (https://hivdb.stanford.edu/), and by the COMET subtyping tool (https://comet.lih.lu/), as previously described.[14]

2.2. Intact and defective provirus quantification by digital-droplet PCR

Cryopreserved PBMCs from each participant were thawed, and genomic DNA was isolated using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). For each sample, DNA concentration was determined by fluorometry (Qubit dsDNA BR Assay Kit, Thermo Fisher Scientific, Waltham, Massachusetts, USA) and underwent digital droplet PCR (ddPCR) for intact provirus (IPDA, intact provirus DNA assay) as previously described.[15] Briefly, to ensure consistent quantification, DNA extracted from each clinical sample and time point was run in triplicate wells, which then were merged during the analysis: a total amount of extracted nucleic acid containing a median (IQR) 138,209 (110, 448–173,880) number of PBMC were analyzed. DNA samples underwent ddPCR on the Bio-Rad QX200 AutoDG Digital Droplet PCR system, using the appropriate manufacturer-supplied consumables and the ddPCR Supermix for probes (no dUTPs) (Bio-Rad Laboratories, S.r.l., Milan, Italy). The primer/probe mix consists of oligonucleotides for two independent hydrolysis probe reactions that interrogate conserved regions of the HIV-1 genome to discriminate intact from defective proviruses i.e. the packaging signal (Ψ) with the amplicon positioned at HXB2 coordinates 692–797 recognized by a 5′ 6-FAM-labeled hydrolysis probe, and the HIV-1 *rev* response element (RRE) of the proviral genome, with the amplicon positioned at HXB2 coordinates 7736–7851. The second amplicon may be recognized by two hydrolysis probes: a 5′ VIC-labeled probe specific for wild-type proviral sequences, and a 5′ unlabelled probe specific for APOBEC-3G hypermutated proviral sequences. After the PCR reaction, the droplets were subsequently read by a QX100 droplet reader, and data were analyzed using QuantaSoft software (Bio-Rad); ddPCR droplets containing intact proviruses exhibit probe signal from both discriminatory amplicons, while droplets containing defective proviruses exhibit only 1 probe signal from a single discriminatory amplicon. In addition, since assay results may be affected by DNA shearing between IPDA amplicons, which artificially reduces intact provirus counts while increasing defective provirus counts as described in Ref. 15, the amount of intra-amplicon shearing which occurred during DNA isolation and assay setup was directly measured for each sample by duplex ddPCR reaction with primers and probes for human albumin present in the genome at a distance of 5110 nucleotides, a similar distance between the packaging HIV-1 sequence and env in the IPDA assay. To this aim, a specific Bio-Rad albumin assay was created, putting together the ddPCR Copy Number Variation assay (dHsaCNS864404398), in silico validated assay labeled with HEX fluorophore, with the dHsaCNS344298713 custom assay, specifically designed considering 5110 nucleotides from the above-mentioned assay, labeled with FAM fluorophores, using the Biorad Web Tool for ddPCR custom assay (Bio-Rad Droplet Digital PCR Assays, Copy Number Determination). For this evaluation, one-third of the total amount of extracted nucleic acid used for IPDA ddPCR for each sample was used. For each sample, the intra-amplicon shearing was calculated using the ratio of dual fluorescent to single fluorescent droplets to determine the frequency of the U1 cell line[16] and pNL4- 3 plasmid, both obtained from the NIH AIDS reagent program, were used as controls. The number of intact and defective provirus was referred to as one million PBMC using the DNA concentration obtained by Qubit fluorimetric reads.

2.3. Provirus integration site analysis

The analysis of provirus integration in host genomes was carried out as previously described,[17] and next-generation sequencing (NGS) and genomic DNA mapping of retroviral integration sites were performed as in Ref. 18. Briefly, an amount of 1 μg of PBMC DNA underwent fragmentation by restriction enzymes, ligation to double-stranded DNA of a linker, and semi-nested PCR using primers complementary to both the linker DNA and the long terminal repeat (LTR) end of the HIV provirus. NGS of the obtained amplicons was performed with the shotgun approach on the Ion Torrent S5 platform (Termofsher Scientific, Waltham, MA, USA), following the manufacturer's protocols. High-quality reads were filtered using an in-house pipeline (raw reads with mean Phread quality score <20 were discarded) and mapped on HIV-1 reference sequence using BWA v.0.7.12 [https://doi.org/10.1093/bioinformatics/btp324]; reads containing LTR sequence were aligned on an HIV-1 reference genome [NCBI Accession Number KO3455.1] and the Human Reference Genome [GRCh38], discarding all reads that mapped on multiple sites with SAMTOOLS software v.1.3.1 [https://doi.org/10.1093/bioinformatics/btp352]. Reads mapping over each site position within a range of 100 nucleotides were collected using

homemade software.

2.4. Statistical evaluation

Wilcoxon matched-pairs signed rank test and Mann Whitney unpaired *t*-test were used as appropriate.

3. Results

3.1. Quantification of intact and defective HIV-1 proviral DNA

Of the 8 patients included in the study, 3 were infected with HIV-1 subtype B, 3 with subtype C, 1 with the recombinant form CRF02_AG n = 1, and the last one with the other recombinant form CRF63_02A1.

At T0, median (IQR) log HIV-1 RNA copies/ml was 7.00 (6.14-7.06), CD4 T cell counts/ml were 685 (539–883), and the CD4/CD8 ratio 1.061 (0.773-1.606). At T1, all patients were virologically HIV-1 suppressed (HIV- 1 RNA not detectable in plasma). The results of quantification of intact and total HIV-1 proviral DNA in lymphomonocytes at T0 and after 48 weeks of early administrated ART (T1) were reported for each patient in Table 1, together with the ratio between intact and total provirus. At both time points, high variability of the amount of intact proviral DNA was observed. At T0, intact provirus was not detected, i.e., was under the limit of detection for the number of cells analyzed, in 2 patients (Pt 4 and Pt 8); at T1, the number of patients in which intact provirus was not detected increased from 2 to 4 (in addition to Pt4 and Pt8, IPDA was not detectable in Pt5 and Pt 7). For quantitative evaluation, considering an amount of one unit below the limit of detection when provirus was not detected, both intact and total HIV-1 DNA declined significantly between T0 and T1, p = 0.047 and p = 0.008, in Wilcoxon matched-pairs signed rank test, respectively. The median ratio of intact/total proviral HIV-1 DNA did not dramatically change between T0 and T1, remaining almost 8–9% of total HIV-1 DNA. Regarding the nature of defective provirus, deletions in the 5′ region were less frequent than deletions and/or hypermutations in 3’ of the provirus (% median (IQR) 48.50 (20.50-81.00) vs 74.50 (58.75-89.00), p = 0.012 in unpaired *t*-test), with no major differences between the two observation times. The median (IQR) intra-amplicon DNA shearing assessed by duplex ddPCR albumin assay, considering all samples and time-points was 0.75 (0.61-0.76). This index was comparable to that obtained by analyzing DNA extracted from U1-infected control cells, whereas in pNL4 samples this index was almost 1. From these data, it can be inferred that IPDA values were underestimated by technical conditions at about 25%, therefore reaching about 10–11% of total proviral DNA in our patient population.

Table 1
Intact proviral DNA associated with PBMC in acutely infected patients at diagnosis (T0) and after one year of effective ART (T1).

Patient and time point	Intact proviral HIV-1 DNA (copies/million PBMC)	Total HIV-1 DNA (copies/million PBMC)	% intact/total proviral HIV-1 DNA
Pt1 T0	522	3919	13.32
Pt2 T0	618	6054	10.22
Pt3 T0	869	8511	10.21
Pt4 T0	4[a]	8184	0.05
Pt5 T0	11	1557	0.74
Pt6 T0	1829	18608	9.83
Pt7 T0	172	2741	6.28
Pt8 T0	8[a]	3857	0.21
Median (IQR)	347 (9–806)	4987 (3020–8429)	8.05 (0.34–10.22)
Pt1 T1	35	470	7.41
Pt2 T1	13	144	8.72
Pt3 T1	28	288	9.57
Pt4 T1	14[a]	134	10.45
Pt5 T1	4[a]	117	3.41
Pt6 T1	72	791	9.11
Pt7 T1	11[a]	90	12.22
Pt8 T1	8[a]	50	16.00
Median (IQR)	32 (17–63)	139 (97–425)	9.34 (7.74–11.78)

[a] Below the limit of detection.

3.2. Provirus integration site analysis

In Fig. 1, Circos plots are shown for each infected patient, indicating all the proviral integration sites in the human genome observed at T0 and T1, along with the percentages of LTR-containing reads that matched for a given chromosome position at each time-point. For each patient and time-point, a median (IQR) of 617 (481–935) LTR-containing reads was obtained. For the evaluation, we considered only provirus integration sites highlighted by at least 8 matches, i.e., the sites of the human genome recognized by the alignment of at least 8 LTR-containing reads. The median (IQR) of different proviral integration sites in human chromosomes/subject was 5 (2.25-13.00) and 4 (3.00-6.75) at T0 and T1, respectively. Of all the different integration sites observed at T1, 64% were already present at T0. Some provirus integration sites were shared by different patients, being also particularly represented. Among these, provirus integration in NRG2 (neuregulin 2 gene), was observed in all patients at T0 median % (IQR) 56.94 (32.01-86.36) and at T1 in all except one patient, with a median % (IQR) 72.65 (62.27-83.56) of all the LTR read matches. Integration sites observed in more than 2 patients were in TXNDC16 (thioredoxin domain containing 16) observed in 5 patients; in 3 patients at T0 median % (IQR) 0.97 (0.57–1.75) and at in 3 patients at T1 1.12 (0.90-2.21), with only one patient with this provirus insertion observed at both T0 and T1 and in STUM (mechanosensory transduction mediator homolog gene) observed in 2 patients at T0 median % (IQR) 19.97 (18.00-21.93) and in other 2 different patients at T1 median % (IQR) 4.37 (2.81-5.93). Provirus integration observed at least in two patients were in B4GALT5 (beta-1,4-galactosyl transferase 5), CNGB1 (cyclic nucleotide-gated channel subunit beta1), FLT1 (fms-related receptor tyrosine kinase) and GATAD2A (zinc finger domain containing 2A) all of them not reaching more than 2% of the of integration sites for single time point, except for GATAD2A present in the T1 of 2 patients at 51.48 and 3.93%, respectively. All the provirus integration sites in human chromosomes at both T0 and T1 observed in genic regions were in introns of transcriptionally active genes, when considering the most characterized splice variants, expressed in white blood cells, lymph nodes, and stem cells, as reported in relevant databases (ensambl.org and genecards.org). In particular, insertion in NRG2 was observed within the large intron 1–2 region of the most characterized splice variants https://jul2022.archive.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=ENSG00000158458;r=5:139846779-140043299;t=ENST00000361474). About the transcriptional orientation of the provirus in relation with the transcription of the gene in which it is inserted, we found that this was always the same for each insertion event, within and among different subjects. Insertion of the provirus in the NRG2 gene was always in the same orientation with the transcription of the human gene, as well as the insertions in B4GALT5, CNGB1, and FLT1 genes, whereas insertions in TXNDC16, STUM, and GATAD2A genes were always convergent. Overall, no major differences in the proportion of the orientation of the provirus insertions with respect to human gene transcription were observed between T0 and T1 (54 and 50% in the same orientation with host transcription at T0 and T1). In addition, in all patients, proviral integration was observed also in inter-genic regions (IGR), representing a median (IQR) percentage of 15.13 (6.81-21.40) at T0 and of 18.46 (8.98-22.18) at T1 of all LTR read matches. In all patients and time-points, 10 different provirus insertions in IGR were observed and located, as for provirus insertion in genes, at the same chromosome position within and among patients (Table 2). The insertion in the IGR of chromosome 2 at nucleotide 16.092.405 (named IGR1 in the Figure) was the most prevalent, being observed in all patients at a median (IQR) percentage of 6.15 (4.63-16.38) at T0 and of 14.13 (8.98-18.34) at T1.

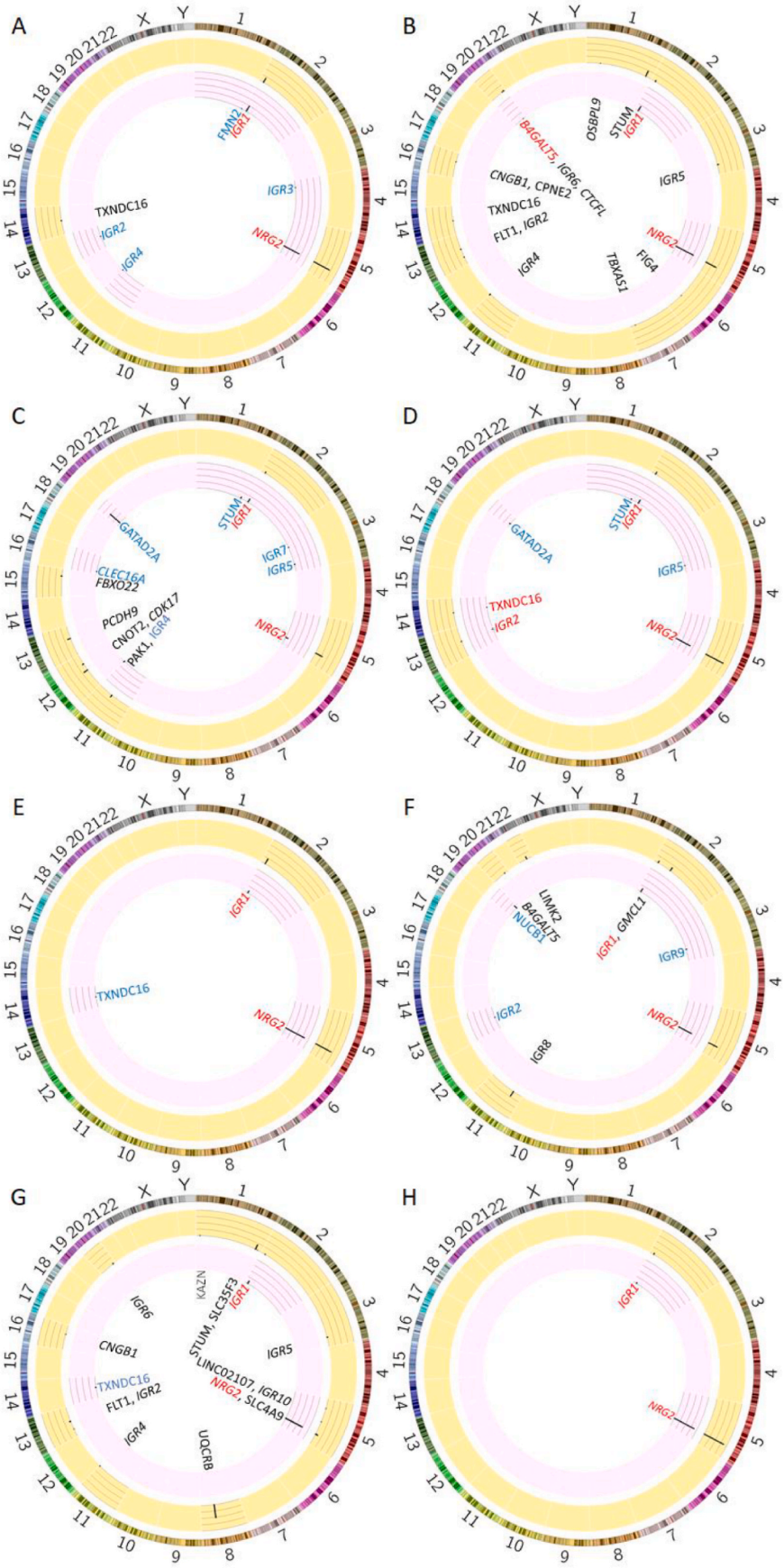

Fig. 1. Circos plots of HIV-1 proviral integration sites in human chromosomes observed in the analyzed patients from Pt1 (A) to Pt8 (H). The positions of the provirus insertions found at T0 in the outer ring (yellow) and at T1 in the inner ring (pink) are marked in the chromosomes slice with the lengths of the tick marks proportional to the frequencies of the total LTR reads matching that particular chromosome position. Each ring is divided into internal rings indicating intervals of 25% of the frequencies. The insertions are shown with the name of the gene or with the intergenic region number (IGR) in which the provirus was found. IGR were numbered from 1 to 12 as consecutively reported. The names of the genes and the IGR were colored in black if provirus insertion was observed only at T0, in blue if found only at T1, and in red if observed at both T0 and T1. The names of the genes were in Italics if the orientation of the provirus insertion with respect to the host gene transcription was convergent. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Chromosome positions for provirus insertions in genes (A) and IGR (B).

A

Chromosome (NCBI Reference Sequence)	Nucleotide position	Gene
NC_000005.10	139967566	NRG2, neuregulin 2
NC_000014.9	52551000	TXNDC16, thioredoxin domain containing 16
NC_000001.11	226567704	STUM, mechanosensory transduction mediator homolog
NC_000020.11	49,687,838	B4GALT5, beta-1,4-galactosyltransferase 5
NC_000016.10	57923016	CNGB1, cyclic nucleotide gated channel subunit beta 1
NC_000013.11	28414033	FLT1, fms related receptor tyrosine kinase
NC_000019.10	19409139	GATAD2A, GATA zinc finger domain containing 2A
NC_000008.11	96229643	UQCRB, ubiquinol-cytochrome c reductase binding protein
NC_000005.10	38066142	LINC02107, long intergenic non-protein coding RNA 2107
NC_000001.11	14548767	KAZN, kazrin, periplakin interacting protein
NC_000001.11	234143179	SLC35F3, solute carrier family 35 member F3
NC_000016.10	57923016	CNGB1, cyclic nucleotide gated channel subunit beta 1
NC_000005.10	140366572	SLC4A9, solute carrier family 4 member 9
NC_000002.12	69838525	GMCL1, germ cell-less 1, spermatogenesis associated
NC_000022.11	31220676	LIMK2, LIM domain kinase 2
NC_000001.11	51,687,823	OSBPL9, oxysterol binding protein like 9
NC_000006.12	109759890	FIG4, 4 phosphoinositide 5-phosphatase
NC_000016.10	57096107	CPNE, copine 1
NC_000007.14	139808617	TBXAS1, thromboxane A synthase 1
NC_000020.11	57517885	CTCFL, CCCTC-binding factor like
NC_000012.12	96358019	CDK17, cyclin dependent kinase 17
NC_000015.10	75922334	FBXO22, F-box protein 22
NC_000011.10	77403760	PAK1, p21 (RAC1) activated kinase 1
NC_000013.11	67087153	PCDH9, protocadherin 9
NC_000019.10	48918858	NUCB1, nucleobindin 1
NC_000016.10	11077559	CLEC16A, C-type lectin domain containing 16A
NC_000001.11	240464202	FMN2, formin 2

B

Chromosome (NCBI Reference Sequence)	Nucleotide position	IGR
NC_000002.12	16097405	1
NC_000013.11	59598069	2
NC_000004.12	35007318	3
NC_000011.10	133588816	4
NC_000003.12	163832269	5
NC_000020.11	56101666	6
NC_000003.12	75154636	7
NC_000011.10	35075651	8
NC_000003.12	140322623	9
NC_000005.10	89862808	10

All other sites of proviral integration in IGR found in more than 1 patient at T0 or T1, or at both time-points, were in chromosomes 13, 11, 3, and 20 (named IGR2, IGR4, IGR5, and IGR6, respectively), all of them not exceeding the 5.5% of all total matches for a patient and time-point. It has to be underlined that for both the provirus insertions in human chromosomes in genic and in IGR observed, the nucleotide position of the insertions was the same, considering the technical interval confidence of 100 nucleotides (Table 2).

4. Discussion

The present data highlighta that, as early as about three weeks from the acquisition of HIV-1 infection, the amount of intact proviral DNA may be only about 10% of the total HIV-1 DNA associated with peripheral lymphomonocytes. Previous reports regarding the intact provirus in these first phases of the infection were focused on the intact fraction after a period of ART initiated during primary infection.[19] A large amount of defective provirus, also in this early phase of infection, may be explained by the inefficient work of the reverse transcriptase that during minus-strand DNA synthesis frequently jumps, generating large internal deletions in proviral DNA.[5] In our study, early administrated ART induced a significant reduction of both intact and defective provirus in PBMCs after 48 weeks of effective therapy, with substantially no changes in the percentages of intact provirus of total HIV1-DNA, this value remaining almost stable at around 10%. In the defective provirus, at both baseline and after early therapy, deletions and/or hypermutation of the 3' end were more frequently observed. In the patients with acute HIV-1 infection, provirus displayed a relatively small number of recurrent integration sites in human chromosomes, some of them shared by different patients in the same genetic position and orientation to the direction of host cell transcription. A very recurrent site of provirus integration in our patient population was in the NRG2 gene, which encodes an epidermal growth factor-like protein, that is a direct ligand for ERBB3 and ERBB4 tyrosine kinase receptors.[20] Others were in TXNDC16, STUM, B4GALT5, CNGB1, FLT1 and GATAD2A genes. All proviral integrations were within introns of transcriptionally active genes and involved genes that encode growth factors, signal transduction proteins, and metabolism-related genes.[12,21-23] Most of the insertion sites observed at T1 were already found at T0 and some of them increased between T0 and T1. An intriguing aspect of the present work is the finding that the same location and orientation of integration of the HIV-1 provirus in human chromosomes may be observed within a subject and more importantly between different patients. This finding, when analyzed within a subject, reinforces the clonal nature of the insertion and possible expansion of cells harboring provirus during the natural course of the infection, as observed for NRG2 insertions. It is in line with previous studies demonstrating that clones of infected cells may arise early in HIV-1 infected individuals.[24,25] If we consider that different subjects may harbor HIV-1 provirus in the same chromosomal location and orientation to host gene transcription, this could be related to the specific nature of the integration process and the putative presence of hot spots for provirus integration. Overall, besides this, provirus

integration in genes involved in cell replication cycles may be more frequently observed if the integration process has an impact on the human gene expression and leads to a selective cell expansion which harbors the provirus in these sites, especially when infected cells may not be eliminated by the immune response. A limitation of our study is that defective provirus has been evaluated in bulk and each provirus insertion in human chromosomes maybe not be directly correlated with the nature of the inserted provirus, i.e., defective, or not. This was because we chose a scalable method to easily quantify defective provirus as in Ref. 26. However, our data also indicated that at the least for the most represented provirus insertion, that is in the NRG2 gene, this was found as the most represented site of insertion in human genes, both in patients displaying the highest and the lowest percentage of the intact provirus. In conclusion, it can be summarized that, from very early in HIV infection, most of the viral reservoir is represented by defective provirus. Since defective provirus may be able to produce HIV-1 proteins[7,8] and these may be related to the persistent stat of immune activation in patients,[9,27] consideration should be given to therapeutic strategies able to target cells harboring defective provirus that may not be reached by the antiviral mechanism of conventional drugs. In this regard, therapeutic approaches able to enhance the CD8 T cell killing of cells harboring defective provirus may be useful in supporting standard antiviral therapies and also impact on replicative competent integrated provirus.

Funding

This work has been funded by the Italian Ministry of Health (Ricerca Corrente to INMI L.Spallanzani, IRCCS). The funders had no role in the study design, data collection, analysis, or preparation of the manuscript.

Institutional review board statement

Human participants in this study were enrolled into the locally established INMI observational cohort SIREA (SIndrome REtrovirale Acuta). Protocols were approved by the INMI ethical committee on 18 February 2014 and patients were requested to sign an informed consent.

Author contributions

I. Abbate and G. Rozera conceived the project; N. Orchi, V.Puro, A. Mondi, A. Antinori provided clinical diagnosis and care for patients; G. Rozera, G. Sberna, G. Berno, and I. Abbate performed the experiments; C. Gruber, E. Giombini analyzed the data; I. Abbate and G. Rozera wrote the manuscript and prepared the table; Pietro Giorgio Spezia prepared the figure; F. Maggi, E. Girardi, F. Vaia, and I. Abbate revised the manuscript. All authors read and approved the final version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

1 Symons J, Cameron PU, Lewin SR. HIV integration sites and implications for maintenance of the reservoir. *Curr Opin HIV AIDS*. 2018;13:152–159. https://doi.org/10.1097/COH.0000000000000438.
2 Anderson EM, Maldarelli F. The role of integration and clonal expansion in HIV infection: live long and prosper. *Retrovirology*. 2018;15:71. https://doi.org/10.1186/s12977-018-0448-8.
3 Craigie R, Bushman FD. HIV DNA integration. *Cold Spring Harb Perspect Med*. 2012;2, a006890.
4 Wong RW, Mamede JI, Hopec TJ. Impact of nucleoporin-mediated chromatin localization and nuclear architecture on HIV integration site selection. *J Virol*. 2015; 89:9702–9705.
5 Hu WS, Hughes SH. HIV-1 reverse transcription. *Cold Spring Harb Perspect Med*. 2012; 2, a006882.
6 Imamichi H, Dewar RL, Adelsberger JW, et al. Defective HIV-1 proviruses produce novel protein- coding RNA species in HIV-infected patients on combination antiretroviral therapy. *Proc Natl Acad Sci U S A*. 2016;113:8783–8788.
7 Pollack RA, Jones RB, Pertea M, et al. Defective HIV-1 proviruses are expressed and can be recognized by cytotoxic T lymphocytes which shapes the proviral landscape. *Cell Host Microbe*. 2017;21:494–506. https://doi.org/10.1016/j.chom.2017.03.008. e4.
8 Imamichi H, Smith M, Adelsberger JW, et al. Defective HIV-1 proviruses produce viral proteins. *Proc Natl Acad Sci U S A*. 2020;117:3704–3710. www.pnas.org/cgi/doi/10.1073/pnas.1917876117.
9 Ishizaka A, Sato H, Nakamura H, et al. Short intracellular HIV-1 transcripts as biomarkers of residual immune activation in patients on antiretroviral therapy. *J Virol*. 2016;90:5665–5676.
10 Gandhi RT, Cyktor JC, Bosch RJ, et al. Decay of intact HIV-1 proviral DNA on antiretroviral therapy. *J Infect Dis*. 2021;223:225–233.
11 Peluso MJ, Bacchetti P, Ritter KD, et al. Differential decay of intact and defective proviral DNA in HIV-1–infected individuals on suppressive antiretroviral therapy. *JCI Insight*. 2020;5, e132997. https://doi.org/10.1172/jci.insight.132997.
12 Cohn L, Silva IT, Oliveira TY, et al. HIV-1 integration landscape during latent and active infection. *Cell*. 2015;160:420–432.
13 Han Y, Lin YB, An W, et al. Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough. *Cell Host Microbe*. 2008;4: 134–146. https://doi.org/10.1016/j.chom.2008.06.008.
14 Fabeni L, Berno G, Fokam J, et al. Comparative evaluation of subtyping tools for surveillance of newly emerging HIV-1 strains. *J Clin Microbiol*. 2017;55:2827–2837.
15 Bruner KM, Wang Z, Simonetti FR, et al. A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. *Nature*. 2019;566:120.
16 Symons J, Chopra A, Malatinkova E, et al. HIV integration sites in latently infected cell lines: evidence of ongoing replication. *Retrovirology*. 2017;14:23. https://doi.org/10.1186/s12977-017-0340-y.
17 Serrao E, Cherepanov P, Engelman AN. Amplification, next-generation sequencing, and genomic DNA mapping of retroviral integration sites. *JoVE*. 2016;109, 53840. https://doi.org/10.3791/53840.
18 Rozera G, Visco-Comandini U, Giombini E, et al. Analysis of HIV quasispecies and virological outcome of an HIV D+/R+ kidney-liver transplantation. *Virol J*. 2022;19: 4. https://doi.org/10.1186/s12985-021-01730-w.
19 Bruner KM, Murray AJ, Pollack RA, et al. Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat Med*. 2016;22:1043–1049. https://doi.org/10.1038/nm.4156.
20 Kermit L, Carraway I, Webert JL, et al. Neuregulin-2, a new ligand of ErbB3/ErbB4-receptor tyrosine kinases. *Nature*. 1997;387:512–516.
21 Schroder ARW. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell*. 2002;110:521–529.
22 Maldarelli F, Wu X, Su L, et al. HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science*. 2014;345:179–183. https://doi.org/10.1126/science.1254194.
23 Wagner TA, McLaughlin S, Garg K, et al. HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science*. 2014;345: 570–573.
24 Kok YL, Vongrad V, Chaudron SE, et al. HIV-1 integration sites in CD4+ T cells during primary, chronic, and late presentation of HIV-1 infection. *JCI Insight*. 2021;6, e143940. https://doi.org/10.1172/jci.insight.143940.
25 Coffin JM, Wells DW, Zerbato JM, et al. Clones of infected cells arise early in HIV-1 infected individuals] and may suggest that clonally expanded cells can arise already during early phases of the infection. *JCI Insight*. 2019;4(12), 128432.
26 Simonetti FR, White JA, Tumiotto C, et al. Intact proviral DNA assay analysis of large cohorts of people with HIV provides a benchmark for the frequency and composition of persistent proviral DNA. *Proc Natl Acad Sci U S A*. 2020;117:18692–18700. https://doi.org/10.1073/pnas.2006816117.
27 Miedema F, Hazenberg MD, Tesselaar K, van Baarle D, de Boer RJ, Borghans JAM. Immune activation and collateral damage in AIDS pathogenesis. *Front Immunol*. 2013;4:298.