IDENTICAL HIV-1 PROVIRUSES ORIGINATE FROM CELL PROLIFERATION OR INFECTION WITH A COMMON VIRAL ANCESTOR

by

Aurelie Niyongabo

B.S. Biochemistry (State University of New York at Binghamton) 2017

THESIS

Submitted in partial satisfaction of the requirements

for the degree of

MASTER OF SCIENCE

in

BIOMEDICAL SCIENCE

in the

GRADUATE SCHOOL

of

HOOD COLLEGE

May 2020

Accepted:

_____	_____
Dr. Ann Boyd, Ph.D. Committee Member	Dr. Ann Boyd, Ph.D. Director, Biomedical Science Program
_____	
Dr. John Coffin, Ph.D. Committee Member	
_____	_____
Dr. Mary Kearney, Ph.D. Thesis Adviser	Dr. April M. Boulton, Ph.D. Dean of the Graduate School

STATEMENT OF USE AND COPYRIGHT WAIVER

I do authorize Hood College to lend this thesis, or reproductions of it, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

`

DEDICATION

I dedicate my Master's Thesis work to my entire family. To my father and mother, Célestin Niyongabo and Séraphine Manirambona, I am completely grateful for all your constant love, support, and encouragement. Thank you both for inspiring me to further my education in the field that I am passionate about. You taught me that "akagumye bagumako", meaning that success comes from hard work. This master's degree and thesis is dedicated most to you both. To my sisters, Deborah, Muriel, and Martine, I am incredibly grateful for all your constant support, prayers, and love throughout my life and especially during this process. I love and appreciate you all and wish you greatness in all that you are pursuing. I also dedicate this thesis to my church family, thank you for the constant support and prayers.

Thank you all again for the inspiration, guidance, love, and support that have encouraged me to pursue and complete this thesis in the field that I am passionate about!

ACKNOWLEDGEMENTS

I wish to thank my committee members for their time and support. I would like to thank to Dr. Ann Boyd for all the help and support. I learned so much in all the classes you taught. Your passion for teaching has inspired me to pursue a Ph.D. and go into academia. I would like to thank Dr. John Coffin for all the fantastic edits, insights, and discussions. You helped me improve my presentation skills, scientific thinking, and problem solving and always raised important points that improved the project. I strongly appreciate all the time and help you dedicated to this work. I wish to thank my mentors, Dr. Mary Kearney and Dr. Sean Patro, for the discussions, support, and help. Thank you, Mary, for all you have done to increase my knowledge of the HIV field. From attending international meetings to presenting my own research at lab meetings, I learned a great deal. Thank you, Sean, for insightful and helpful discussions regarding this project. You taught me a great deal and I gained invaluable laboratory skills that made me a better scientist. I wish to also thank everyone in the Kearney lab for all the helpful discussions regarding this project. Finally, I wish to thank Dr. Frank Maldarelli for the donor samples and helpful discussions that helped this work.

Thank you, Dr. Ann Boyd, Dr. John Coffin, and Dr. Mary Kearney for agreeing to be on my thesis committee and supporting this project. I am incredibly grateful.

TABLE OF CONTENTS

ABSTRACT

Understanding the mechanisms for HIV-1 persistence during antiretroviral therapy (ART) is crucial for developing curative strategies. Due to the high intra-patient genetic diversity of HIV-1, I hypothesized that integrated HIV-1 DNA (proviruses) with identical sub-genomic sequences are sustained during ART through cellular proliferation. To test this hypothesis, a method called "Multiple-Displacement Amplification Single-Genome Sequencing" (MDA-SGS) wherein the site of proviral integration in the host genome and the full-length HIV-1 sequence can be determined, was applied to 34 sets of proviruses with identical P6, protease, reverse transcriptase (P6-PR-RT) sequences in a single donor with viremia suppressed on ART reported by Musick, et al[(1)]. The MDA-SGS workflow includes the isothermal amplification of DNA from cells containing single proviruses of interest within their sites of host integration followed by integration site analysis on the products to determine if proviruses identical in P6-PR-RT also have identical integration sites and, therefore, result from proliferation of a single infected cell. If different sites of integration are observed, then I conclude that the identical proviruses result from infection of two or more different cells by a closely related viral ancestor. Of the 20 populations of proviruses with identical P6-PR-RT sequences successfully assayed by MDA-SGS, I found 9 to contain only identical integration sites (cell clones), 6 to have only unique integration sites (infection with a common ancestor), and 5 to contain a combination. The finding that proviruses identical in P6-PR-RT often have different sites of integration suggests infection of multiple cells prior to ART or during an ART interruption with a common viral ancestor, each establishing a latent infection allowing them to survive and, likely, divide. Targeting such long-lived, infected, proliferating cells is necessary to achieve HIV-1 remission without ART.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ART	Antiretroviral Therapy
CTM	Central Transitional Memory
EM	Effector Memory
HIV-1	Human Immunodeficiency Virus Type 1
SGS	Single Genome Sequencing
MDA-SGS	Multiple Displacement Amplification Single-Genome Sequencing
PBMC	Peripheral Blood Mononuclear Cells
PCR	Polymerase Chain Reaction
VOA	Viral Outgrowth Assay
QVOA	Quantitative Viral Outgrowth Assay

INTRODUCTION

Human Immunodeficiency Virus Type 1 (HIV-1) is a primate lentivirus that infects immune cells, specifically CD4+ T cells and macrophages. There are currently 37.9 million people living with HIV worldwide [(2)] with the majority of infections occurring in sub-Saharan Africa, making it a major public health issue. Of the 37.9 million people living with HIV, less than 60% have access to Antiretroviral Therapy (ART)[(2)]. Without treatment, HIV infection depletes CD4+ T cells, and opportunistic infections lead to eventual death in most individuals.

The HIV replication cycle includes 7 basic steps: attachment, entry, reverse transcription, integration, assembly, release, and maturation. During attachment, the HIV envelope protein binds to the CD4 receptor and the CCR5 or CXCR4 coreceptor on the surface of T cells then fuses the virion envelope` with the cell for entry. After entry, the virus core is transported to the nucleus while its RNA genome is reverse transcribed into double-stranded DNA (**Figure 1**)[(3)]. The double-stranded DNA genome is integrated into the host genomic DNA. The integrated viral genome is referred to as a "provirus" and can be transcribed and translated as any host gene. When expressed, new viral proteins assemble into particles that are released, which can result in cell death. However, if the provirus is not expressed, the cell can persist and divide, passing the provirus on to its cellular progeny[(4, 5)].

ART targets multiple steps in the HIV replication cycle including entry, reverse transcription, integration, and maturation, suppressing the plasma viral load to below detection by commercial assays[(6)]. However, highly sensitive assays reveal low-level, persistent virus production during ART, and viremia rebounds to pre-therapy levels if ART is interrupted[(7, 8)]. If ART is not

adhered to, HIV can replicate and mutate in the presence of the drugs and lead to HIV drug resistance and possible transmission of drug resistant strains[9].

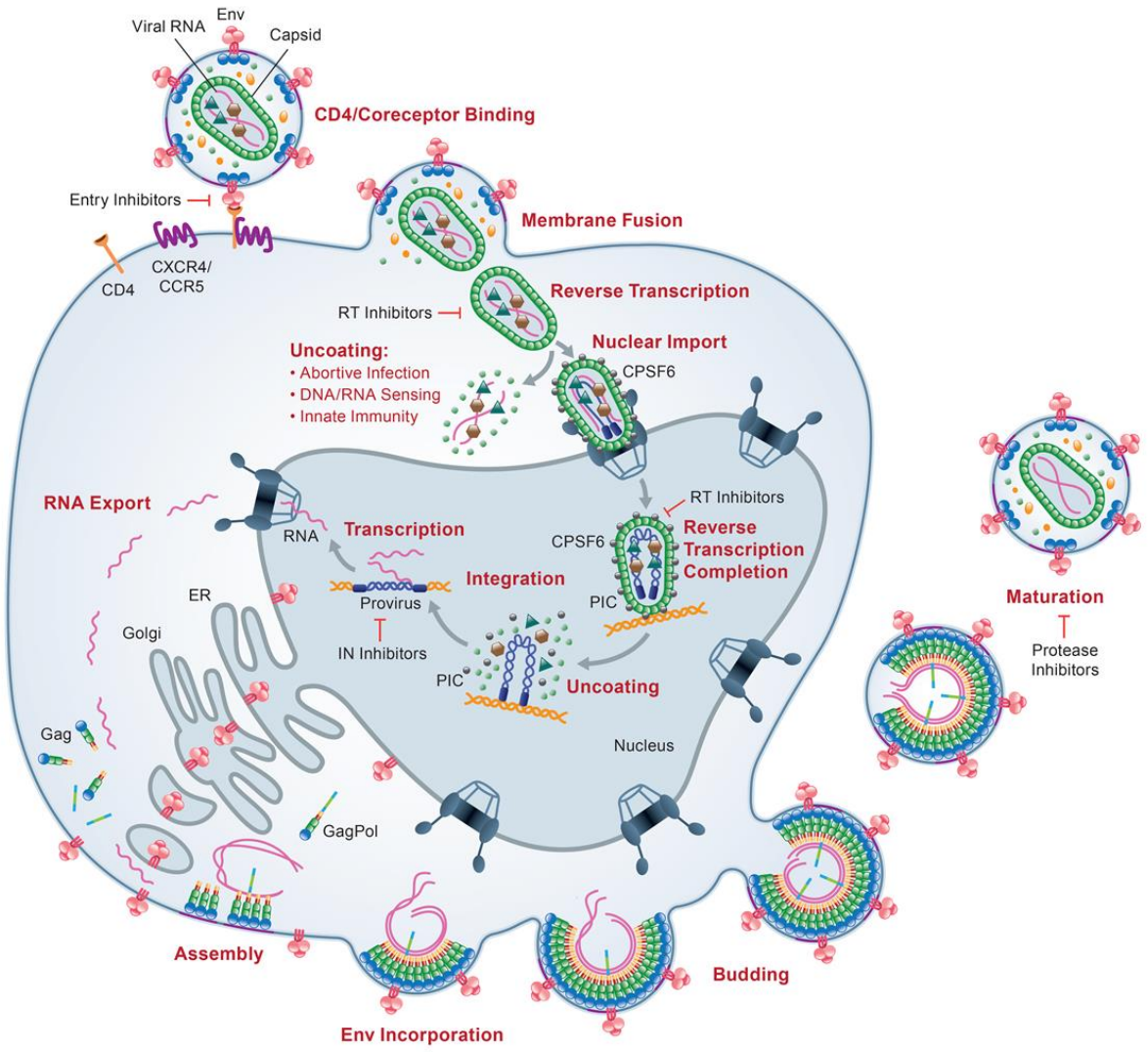

Figure 1. HIV Replication Cycle. The HIV-1 life cycle includes virion attachment, entry, reverse transcription, uncoating, integration, assembly, budding, and maturation. Steps in the lifecycle that are inhibited by ART are shown. Figure by Burdick, et al.[10]

The HIV provirus is ~9.8kb in length with both ends being flanked by long terminal repeats (LTRs)[11] **(Figure 2).** The major proviral genes are *gag*, *pol*, and *env*. *Gag* encodes the structural proteins: matrix (MA), capsid (CA), nucleocapsid (NC), and P6. *Pol* encodes the viral enzymes: protease (PR), reverse transcriptase (RT), and integrase (IN). *Env* encodes the envelope glycoprotein gp160 (cleaved into gp120, gp41)[12]. The genes *tat* and *rev* encode regulatory proteins and *vpu*, *vpr*, *vif*, and *nef* encode accessory proteins[12]

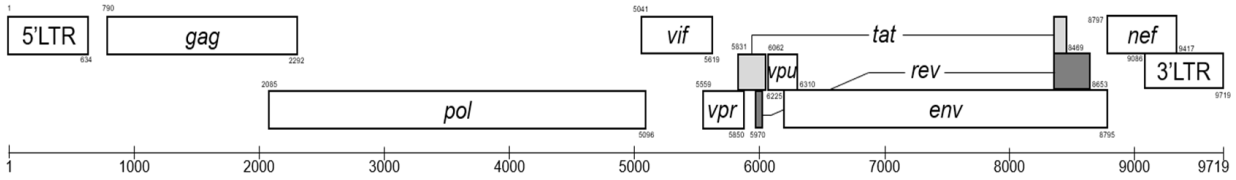

Figure 2. HIV Genome Map. HIV genes *gag, pol, env, tat, rev, vpu, vpr, vif,* and *nef* and locations are shown here. Figure from the HIV sequence database: https://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html

Previous studies have shown that HIV replication is effectively halted when individuals adhere to combination ART and that persistent, low-level viremia results from the expression of proviruses in cells that were infected prior to ART initiation[13-16]. Understanding the mechanisms that sustain the HIV reservoir during ART (the cells that harbor replication-competent proviruses) is crucial to the development of potential curative strategies. One mechanism for HIV persistence during ART is the clonal expansion of cells that were infected prior to treatment initiation[4, 15, 17-22]. Most clones of infected cells carry defective proviruses due to errors made during reverse transcription[23]; however, some cell clones harbor replication-competent proviruses that can be a source of persistent[24] and rebound[15, 24] viremia if ART is interrupted. It has been shown that clonal expansion, not viral replication, maintains HIV during ART[4, 16]. Characterizing the HIV reservoir

includes profiling the genetics of HIV proviruses that persist in T cell clones during ART and determining their sites of integration in the host genome.

Simonetti et al.[24] were the first to show that an infected, expanded T cell clone carrying a replication-competent provirus could persist during ART and be the source of persistent, low-level viremia in a donor denoted "Patient 1." Patient 1 was a 58 year-old African American man diagnosed with advanced HIV infection (16 CD4+ T cells/µl) in May of 2000[24]. He was on ART for 13 years and treatment interruptions occurred unexpectedly throughout (**Figure 3A**). Initially on ART, his HIV plasma RNA levels declined to below the limit of detection and his CD4+ T cell count partially recovered[24]. Single-genome sequencing (SGS) on P6-PR-RT in plasma virus had previously revealed that the virus present at the time of diagnosis was genetically diverse[13]. When experiencing ART failure at a late timepoint (starting around year 12), a population of diverse drug-resistant variants and a population of identical wildtype variants dominated the plasma virus [4, 24]. The ART regimen was adjusted to suppress the rebounding viremia. Although the viral variants that were resistant to the previous regimen declined to below the limit of detection on the new drug regimen, the identical wildtype (i.e., drug sensitive) virus persisted, indicating that the wildtype variants resulted from proviral expression of a large cell clone and not from full cycles of viral replication (**Figure 3B**). Instances of identical plasma virus during ART had been previously reported[4, 13, 19, 24]. Simonetti et al. demonstrated that the wildtype virions that comprised the persistent low-level viremia in patient 1 were produced from a large cell clone that he called "ABMI-1".

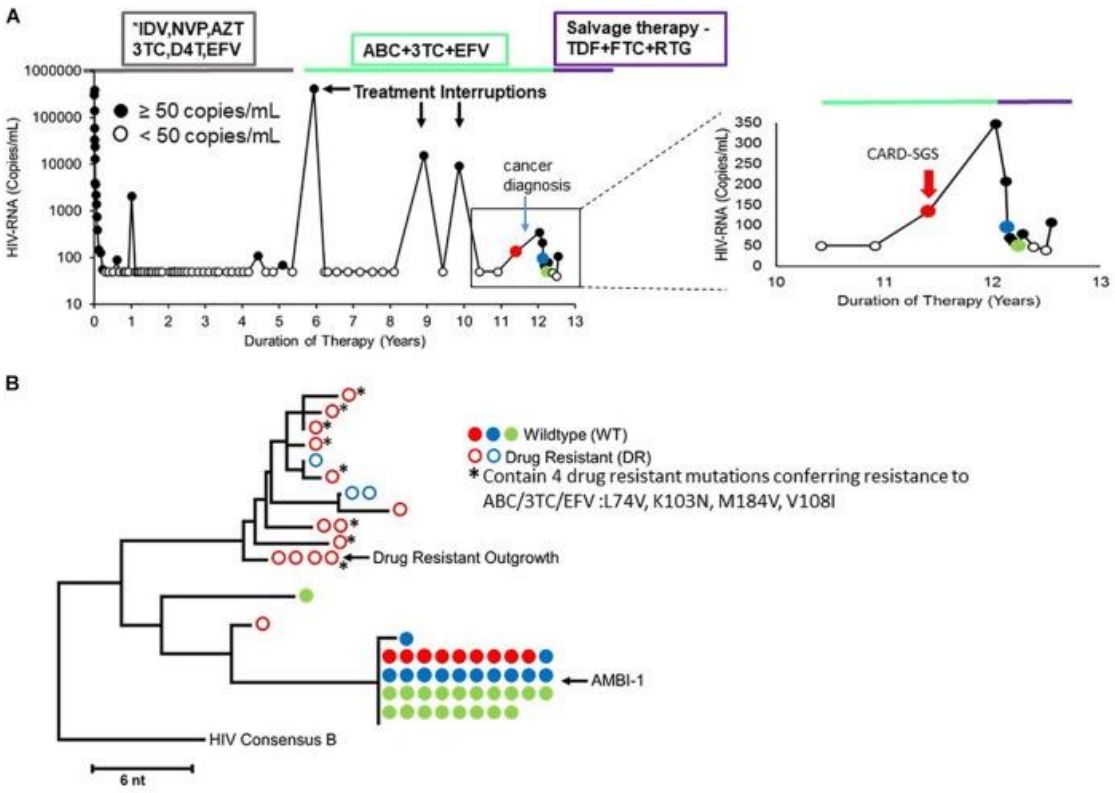

Figure 3. Patient 1 ART Regimen, Levels of Viremia, and Single-Genome Sequences from Plasma Viremia. A) ART regimen and HIV viremia time course and viral load (open circles are below the limit of detection). B) Neighbor joining trees of plasma RNA sequences from timepoints in A. Wildtype sequences are represented by closed red, blue, and green circles. Drug resistant sequences are represented by open red and blue circles. Figure by Musick et al[(1)].

Subsequent to Simonetti, et al., Musick, et al. performed SGS on the proviral population in naïve, central and transitional memory, and effector memory T cells collected during virologic failure in the same individual. He identified 34 populations of identical, wild-type P6-PR-RT sequences (also called phylogenetic "rakes") within and across the T cell subsets (**Figure 4**). Due to their sequence identity, Musick, et al. called these populations (or rakes) "possible cell clones", to express the possibility that, like AMBI-1, they derived from the infection of a single cell that had undergone clonal expansion. The 34 possible cell clones included 3 with replication-competent proviruses determined by outgrowth in co-culture experiments[(24)], called AMBI-1, OG-1, OG-2.

Five rakes contained sequences with obvious genetic defects rendering them defective. Twenty-six had proviruses without obvious defects in P6-PR-RT but did not replicate in culture [25, 26].

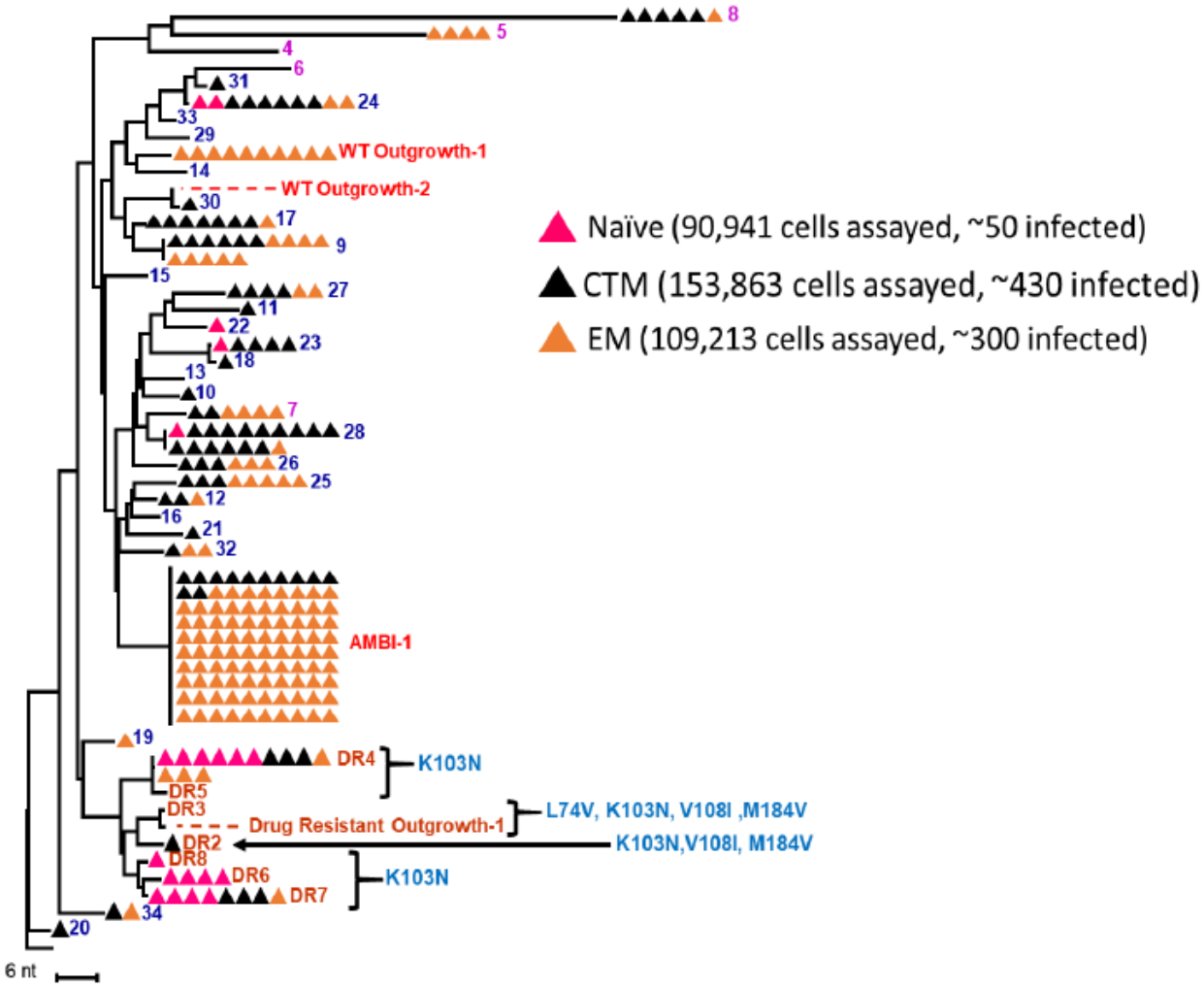

Figure 4. Patient 1 Neighbor Joining Phylogenetic Tree of P6-PR-RT SGS. PBMC were sorted into naïve (pink triangles), central and transitional memory (CTM) (black triangles), and effector memory (EM) (orange triangles) CD4+ T cell subsets. The 34 populations of identical P6-PR-RT sequences described in Musick et al.[1] are labeled. Those rakes labeled in red show P6-PR-RT sequences that match replication-competent proviruses (AMBI-1, OG-1, and OG-2), those in purple have genetic defects in P6-PR-RT, and those in blue have no apparent defects in P6-PR-RT but are not known to be replication-competent. Identical sequences are shown with no horizontal distances on the phylogenetic tree. Figure from Musick et al.[1]

Musick, et al. estimated the fraction of cells within the 34 possible clonally expanded cell populations that contained unspliced HIV RNA and found that it ranged from ~2% to 65% (**Figure**

5). His findings demonstrated that most cells in clones persisting on ART contained proviruses that are transcriptionally silent or latent. Their latent infection likely allowed these infected cells to escape immune detection and proliferate despite ART. Furthermore, Musick, et al. found no difference in the fraction of infected cells with HIV RNA in populations carrying replication-competent proviruses vs. those that had defective proviruses.

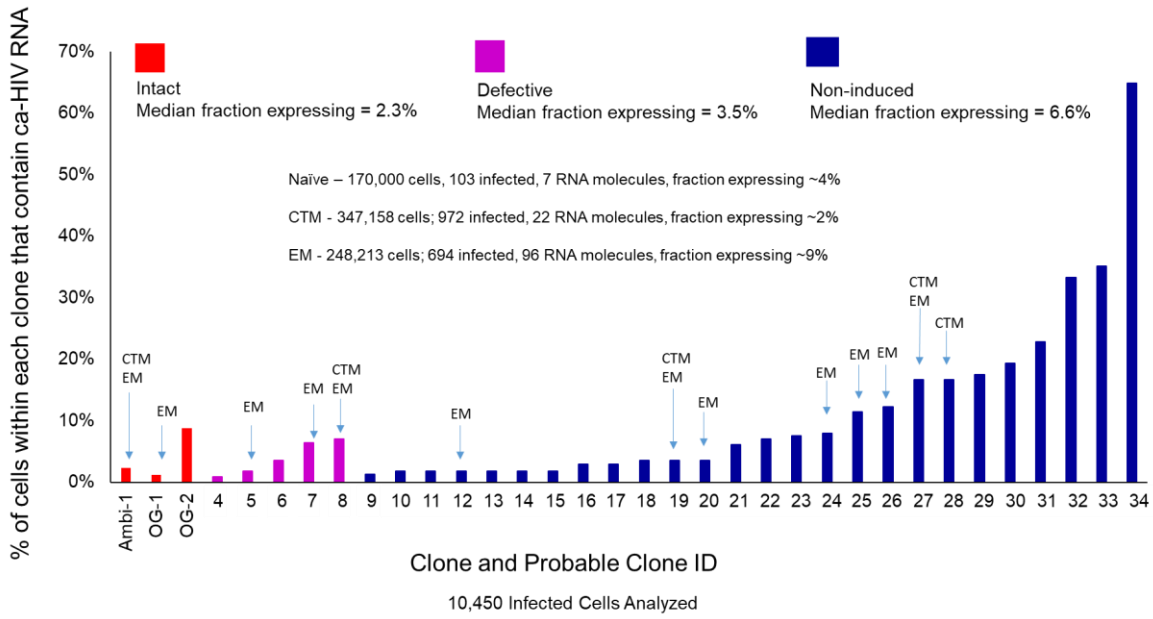

Figure 5. The Fraction of Cells Within Possible Cell Clones that Have HIV RNA. Possible cell clones with replication-competent proviruses are shown in red. Possible cell clones with defective proviruses are shown in purple. The remainder of the possible cell clones are in blue. The cell subsets in which HIV RNA was detected is indicated. Figure from Musick et al.[(1)]

Because assays that recovered both proviral sequences and their site of integration in the host genome did not exist, evidence for HIV-infected cell clones often relied on SGS data alone[(13, 24, 27)]. Integration site analysis provided direct evidence of clonal expansion of infected cells but did not yield information regarding proviral sequences[(4, 16, 28, 29)]. Therefore, we and others developed the multiple displacement amplification single-genome sequencing (MDA-SGS) method wherein

both the proviral integration site and the full-length HIV-1 proviral sequence can be determined[28, 29]. For my thesis research, I applied the MDA-SGS assay to DNA extracted from the same T cell subsets of Patient 1 described in Musick, et al. to determine if the 34 "possible cell clones" resulted from the proliferation of a single infected cell (identical P6-PR-RT sequences have the same integration site into the host DNA) or if they resulted from the infection of two or more different cells by a related viral ancestor (identical P6-PR-RT sequences have different integration sites). Given the extensive clinical history, integration site characterization, and plasma/proviral sequence information for Patient 1, my thesis research also aimed to further investigate the genetics of the full-length proviruses in the 34 possible cell clones. Of the rakes that were confirmed to result from clonal expansion of a single cell, I also analyzed the results to determine if their proviral structures or integration sites could be associated with their RNA expression reported by Musick, et al.[1]. Identifying and reporting the sources of identical sub-genomic sequences is imperative for understanding the establishment, maintenance, and composition of the viral reservoir and the source of rebound viremia when ART in interrupted.

MATERIALS AND METHODS

Donor and Samples in Study

Peripheral blood mononuclear cells (PBMC) were obtained from "Patient 1" reported in previous studies[1, 4, 24, 29] (**Table 1**). The donor was enrolled in NIH protocol 00-I-0110 conducted at the NIH Clinical Center in Bethesda, MD and was approved by the NIH internal review board. Informed, written, consent was obtained for the collection of blood samples.

Table 1: **Patient 1 Samples in Study Information.** Sample dates match the timepoints in Figure 3a

Years on ART	Cell Type	# Cells Recovered
11	Effector Memory (EM)	400,000
11	Central/Transitional Memory (CTM)	800,000
11.5	Effector Memory (EM)	495,425

Fluorescence-Activated Cell Sorting (FACS), Gating, T-Cell Subset Definitions

The FACS and gating methods are described from Music et. al[1]. In brief, PBMC from Patient 1 were obtained by leukapheresis. Cell sorting for single CD4+ T cells was done with the following antibodies: CD3-APC-H7 (BD Biosciences, 641406), CD4-BV785 (BioLegend, 317442), CD8-QDot655 (Invitrogen, Q10055), CD11c-PE (BD Biosciences, 347637), CD14-PE (BD Biosciences, 555398), CD27-PE/Cy5 (Beckman Coulter, 6607107), CD45RO-ECD (Beckman Coulter, IM2712U), CD56-APC (BioLegend, 304610), CD57-BV421 (NIAID Vaccine Research Center (VRC) Ab), CCR7-Ax700 (VRC Ab), and TCR γδ (BD Biosciences, 555718) for 15 minutes at 22°C. Cells were washed, kept on ice, and sorted on a BD FACSAria into CD4+ memory subsets. The Naïve markers include CD3+Aqua-CD8-CD4hiCD56-TCRγδ-CD14-CD11c-

CD27+CD45RO-CCR7+CD57-. the Central and Transitional Memory markers include CD3+Aqua-CD8-CD4hiCD56-TCRγδ-CD14-CD11c-CD27+CD45RO+. The Effector Memory markers include CD3+Aqua-CD8-CD4hiCD56-TCRγδ-CD14-CD11c-CD27-.

Genomic DNA Extraction

100µl of guanidinium hydrochloride (8M, 100 mL, SIGMA #G9284) + proteinase K (20mg/mL, APPLIED BIOSYSTEMS #AM2548) were added to the cell pellets, vortexed immediately for 10 seconds, and incubated at 42°C for 1 hour. 400µl of guanidinium isothiocyanate (6M, 50mL, SIGMA #50983) + glycogen (20mg/mL, ROCHE #10901393001) were added, vortexed to mix, and incubated at 42°C for another 10 minutes. 500µl of 100% isopropanol (SIGMA) at room temperature was added, vortexed for 10 seconds at high intensity, and centrifuged at 21,000g at room temperature for 10 minutes. The supernatant was removed without disturbing the pellet and the samples were centrifuged at 21,000g for 15 seconds. 750µl of 70% ethanol (SIGMA) was added to wash the nucleic acid pellet, vortexed, and spun at 21,000g at room temperature for 10 minutes. The residual ethanol was removed using 10µl pipette tip. The pellets were air dried for 5-10 minutes. The extracted nucleic acid was resuspended in 100µl of 5mM Tris-HCl (pH 8.0) and incubated at 42°C for up to 2 hours. The nucleic acid was stored at -80°C.

Multiple Displacement Amplification Single-Genome Sequencing (MDA-SGS) Workflow

The MDA-SGS workflow (**Figure 6**) included first identifying proviral sequences of interest by standard SGS. MDA was performed as described below on single proviruses and the host genomic DNA in which they were integrated. Integration site analysis (ISA) (described below) and near full-length (NFL) proviral PCR amplification and sequencing (described below) was performed on the MDA products [(29)].

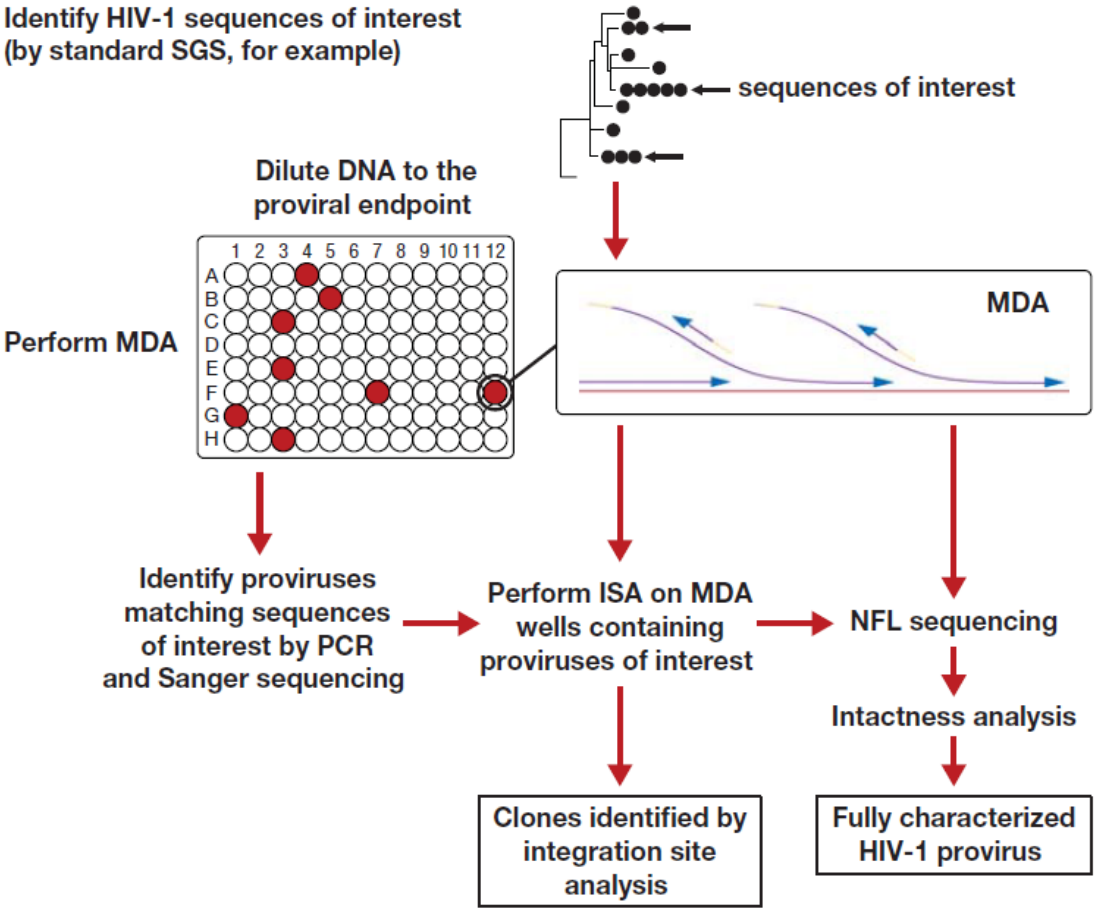

Figure 6. Multiple Displacement Amplification Single Genome Sequencing (MDA-SGS) Workflow. Identical subgenomic sequences of interest are identified by standard SGS. MDA is performed at a proviral endpoint to identify proviruses matching standard SGS sequences. Integration site assay (ISA) and near-full length assay (NFL) are performed on MDA wells of interest for clonal and intactness analysis. Figure from Patro et al. [29].

Multiple Displacement Amplification (MDA)

Extracted genomic DNA was diluted in 5ml Tris-HCl pH 8.0 to a P6-PR-RT endpoint (i.e. 0-1 P6-PR-RT per well)[27]. DNA in each sample was chemically denatured by the addition of 2µl 1x D-solution (1x D-solution is a dilution of 8x D-solution containing 3.2M KOH, 80mM EDTA)

followed by incubation at room temperature for 3-6 minutes. Reactions were neutralized by the addition of 4µl 1x N solution (0.4M Trizma-HCl) and immediately kept at 4°C to minimize any renaturation of genomic DNA. 2µl of the random hexamer primer (p6N, Integrated DNA Technologies: /5phos/NNNN*N*N) was added which contains phosphorothioate linkages at the ultimate and penultimate 3'-terminal positions and is 5' phosphorylated. To each well, the following were added: 4µl 10x phi29 DNA polymerase reaction buffer (New England BioLabs), 4µl 0.5M KCl, 4µl 10mM dNTPs, 10µl 0.8 trehalose (Sigma; T5251-10G), 7µl water and 1µl phi29 DNA polymerase (New England BioLabs). The enzyme mixtures were assembled and dispensed to nucleic acid mixtures on ice and MDA reactions were incubated at 30°C for 18hr in a thermal cycler utilizing a heated lid to minimize condensation. Completed reactions were incubated at 65°C for 10min to denature the polymerase and stored at -20°C.

Integration Sites Assay (ISA)

MDA products were purified using Zymo10 DNA clean and concentration kit (Zymo; D4011). The integration site assay[30] was used to identify HIV integration sites based on linker-mediated amplification of sheared MDA samples (**Figure 7**). ISA included DNA fragmentation, end repair, and linker ligation using the NEB Next Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs, Cat No. E7805S). Patient virus-specific primers in the 5' and 3' LTRs (listed below) and primers in the linkers were used for PCR amplification. Linker details were previously published[30]. Following PCR amplification, DNA sequencing was performed using the Illumina MiSeq platform system. All integration sites were reported using the 3' LTR mapped to hg19[4, 30]. The patient virus-specific primer sequences are as follows:

5'LTR: TCAGGGAAGTAGCCTTGTGTGTGGT

3'LTR: TGTGACTCTGGTAACTAGAGATCCCTC

Nested PCR Primers:

5’LTR:
AATGATACGGCGACCACCGAGATCTACACTATAGCCTACACTCTTTCCCTACACGAC
GCTCTTCCGATCTNNNNNNACGTCTCGTCCTGTCTTTTCTGGGAGTGAACTA

3’LTR:
AATGATACGGCGACCACCGAGATCTACACATAGAGGCACACTCTTTCCCTACACGAC
GCTCTTCCGATCTNNNNNNATAGAGGCCCCTTTTAGTCAGTGTGGAAAATC

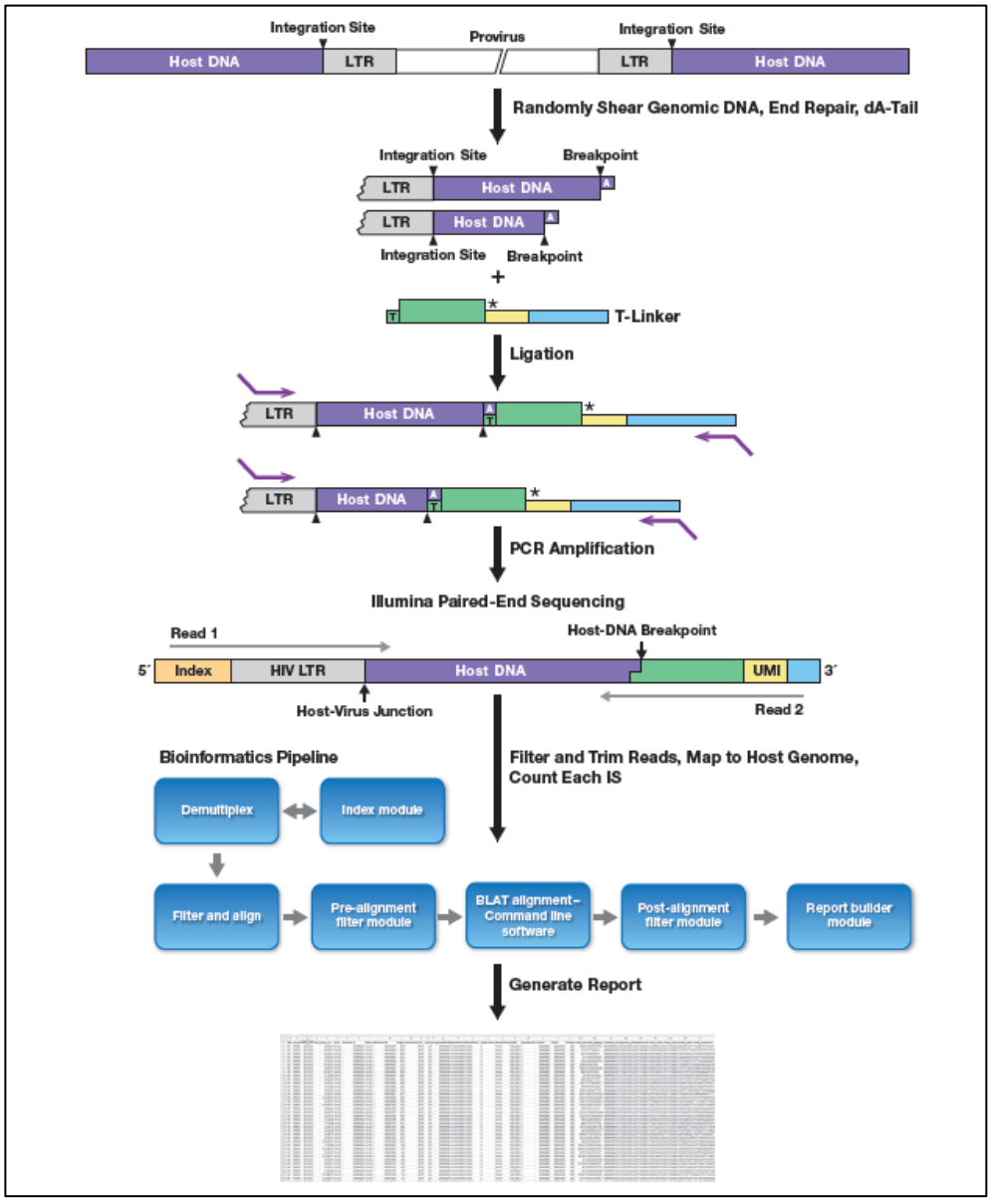

Figure 7. Workflow of Integration Site Analysis. MDA samples were randomly sheared followed by a ligation of a T-linker (containing the unique modifier identifier (UMI)). The samples were PCR amplified and sequenced by MiSeq. A bioinformatics pipeline was used to filter, trim reads, map to host genome, count each integration site, and generate a report[30].

Near-Full-Length (NFL) Amplification and Primers for Sanger Sequencing

Near-full-length (NFL) polymerase chain reaction (PCR) amplification was performed using the Ranger enzyme (BIOLONE-BIO-25052). The first round of PCR, denoted NFL1, was used to generate a 9kb amplicon from U5-5'LTR to U5-3'LTR. The NFL1 PCR product was diluted 1:5 and used as template for nested PCR to generate four 2.9-3.2kb genome-spanning amplicons, denoted F1, F2, F3, and F4 (**Figure 8**). Alternatively, a 9kb NFL2 amplicon was generated. Amplicon sizes were confirmed by ethidium bromide gel visualization and sequenced by Sanger. The resulting data was used to generate the sequence of the NFL amplicon. The primer sequences used for PCR and sequencing are shown in Table 2.

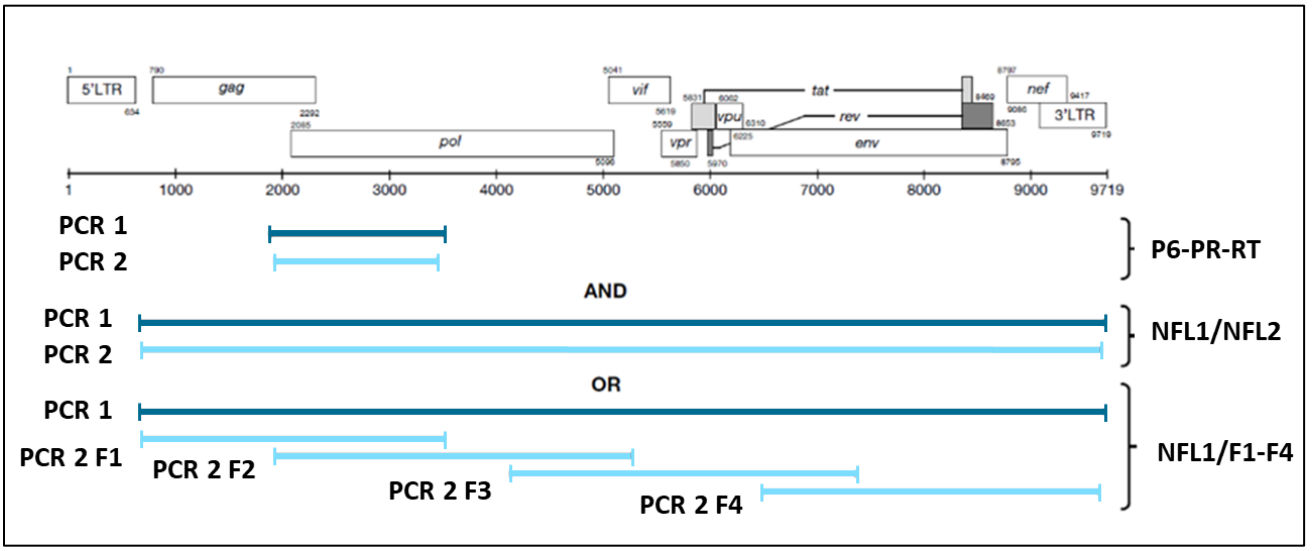

Figure 8. **Method for Near-Full-Length (NFL) Amplification.** MDA wells of interest screened in the P6-PR-RT region are NFL PCR amplified. PCR1 amplifies NFL1. That NFL1 is the template for nested NFL2 amplification or 4 overlapping nesting PCR. The HIV genome map is shown at the top.

Table 2. List of All Primers Used for PCR NFL Amplification and Sequencing

Gene	Primer Location (HXB2 reference)	Sequence
LTR	611(+)	AGTCAGTGTGGAAAATCTCT*A*G
LTR	618(+)	GTGGAAAATCTCTAGCAGT*G*G
LTR	U3-3(-)	ATATCTTGTCTTTTTTGGGAGTAAATTAGCCCTTC
LTR	U3-74(-)	GGTGTGTAGTTCTGCCAATCAGGG
LTR	U3-124(-)	TACTAGCTTGAAGCACCATCCAAAGG
LTR	R-519(-)	GCACTCAAGGCAAGCTTTATTGAGGCTTA
LTR	9662(-)	TTACCAGAGTCACACAACAG*A*C
LTR	U5-577(-)	TGAGGGATCTCTAGTTACCAGAGTC
LTR	9675(-)	GAGGGATCTCTAGTTACCAG*A*G
PBS	PBS-653(+)	AGTGGCGCCCGAACAGGGAC
gag	1466(-)	CCTTGGTTCTCTCATCTGGC
gag	1501(-)	TGAAGGGTACTAGTAGTTCC
gag	1849(+)	GATGACAGCATGTCAGGGAG
gag	1870(+)	GAGTTTTGGCTGAGGCAATGAG
gag	1061(+)	GGATAGAGGTAAAAGACACCAA
gag	1488(+)	AAGTGACATAGCAGGAACTACTAG
gag	2012(+)	CTAGGAAAAAGGGCTGTTGGAAATG
gag	895(-)	AATTTTTCCAGCTCCCTGCTTGCCCA
gag	1294(-)	CTGATAATGCTGAAAACATGGGTAT
gag	1606(-)	GACAGGGCTATACATTCTTACTAT
pol	3410(-)	CAGTTAGTGGTATTACTTCTGTTAGTGCTT

pol	3500(-)	CTATTAAGTATTTTGATGGGTCATAA
pol	3410(+)	AAGCACTAACAGAAGTAATACCACTAACTG
pol	3500(+)	TTATGACCCATCAAAATACTTAATAG
pol	3996(+)	CATCTAGCTTTGCAGGATTCG
pol	4133(+)	GGAAAAGGTCTACCTGGCATG
pol	5270(-)	CTGACCCAAATGCCAGTCTC
pol	5248(-)	TCTCCTGTATGCAGACCCCA
pol	2385(+)	AAAATGATAGGGGGAATTGGAGGTTT
pol	2869(+)	CAGTACTAGATGTGGGGGATGCATA
pol	3298(+)	ACAGCTGGACTGTCAATGATATACA
pol	3676(+)	CCACAGAAAGCATAGTAATATGGGG
pol	4162(+)	CACACAAAGGGATTGGAGGAAATGA
pol	2557(-)	TTACTGGTACAGTTTCAATAGGAC
env	6445(-)	CTTGTGGGCTGGGGTCTGTGGGTACAC
env	6426(-)	CTGTGGGTACACAGGCATGTGTGGCCC
env	E20(+)	GGGCCACACATGCCTGTGTACCCACAG
env	E30(+)	GTGTACCCACAGACCCCAGCCCACAAG
env	E125(-)	CAATTTCTGGGTCCCCTCCTGAGG
env	E115(-)	AGAAAAATTCCCCTCCACAATTAA
env	6497(-)	ACCATGTTATTTTTCCACATGTTAAA
env	7001(-)	CTGCCATTTAACAGCAGTTGAGTTGA
env	7531(-)	ATGGGAGGGGCATACATTGCT
env	7943(-)	CCTGGAGCTGTTTAATGCCCCAGAC

env	8352(-)	GGTGAGTATCCCTGCCTAACTCTAT
env	6552(+)	TATGGGACCAAAGCCTAAAGCCATGTG
env	6961(+)	CAGCACAGTACAATGTACACATGGAA
env	7807(+)	AGCAGCAGGAAGCACTATGGGCGC
env	8257(+)	CATATCAAATTGGCTGTGGTATAT
env	8517(+)	GGAACCTGTGCCTCTTCAGCTACC
nef	9021(+)	CCACAGGTACCTTTAAGACCAATGAC
vif	5451(-)	AGAGATCCTACCTTGTTATGTCCT
tat	5955(-)	CTTCCTGCCATAGGAGATGCCTAAG

Illumina Sequencing

The NFL amplicons generated from MDA samples were also sequenced using the MiSeq Illumina sequencing platform[31] (**Figure 9**). In order to increase throughput, sequencing of the NFL amplicons was done in a multiplexed library using the Nextera XT kit (Illumina-20018705). The Nextera XT Illumina DNA Library prep workflow is described in detail in the protocol included with the kit. In brief, first the sample DNA is fragmented and tagged with adapter sequences using Nextera transposomes. The libraries are then amplified using a limited-cycle PCR program (parameters in the kit protocol) and purified with single-sided bead purification. Purification is followed by a quality control check with Agilent Technology 2100 Bioanalyzer and a High Sensitivity DNA kit (Agilent-5067-4626). Typical broad size distribution of libraries is ~250-1000bp. Normalization of libraries is done with Nextera XT Index v2 or Nextera XT Index Kits (Illumina-catalog # FC-131-1096) to ensure equal representation in the pooled library for sequencing.

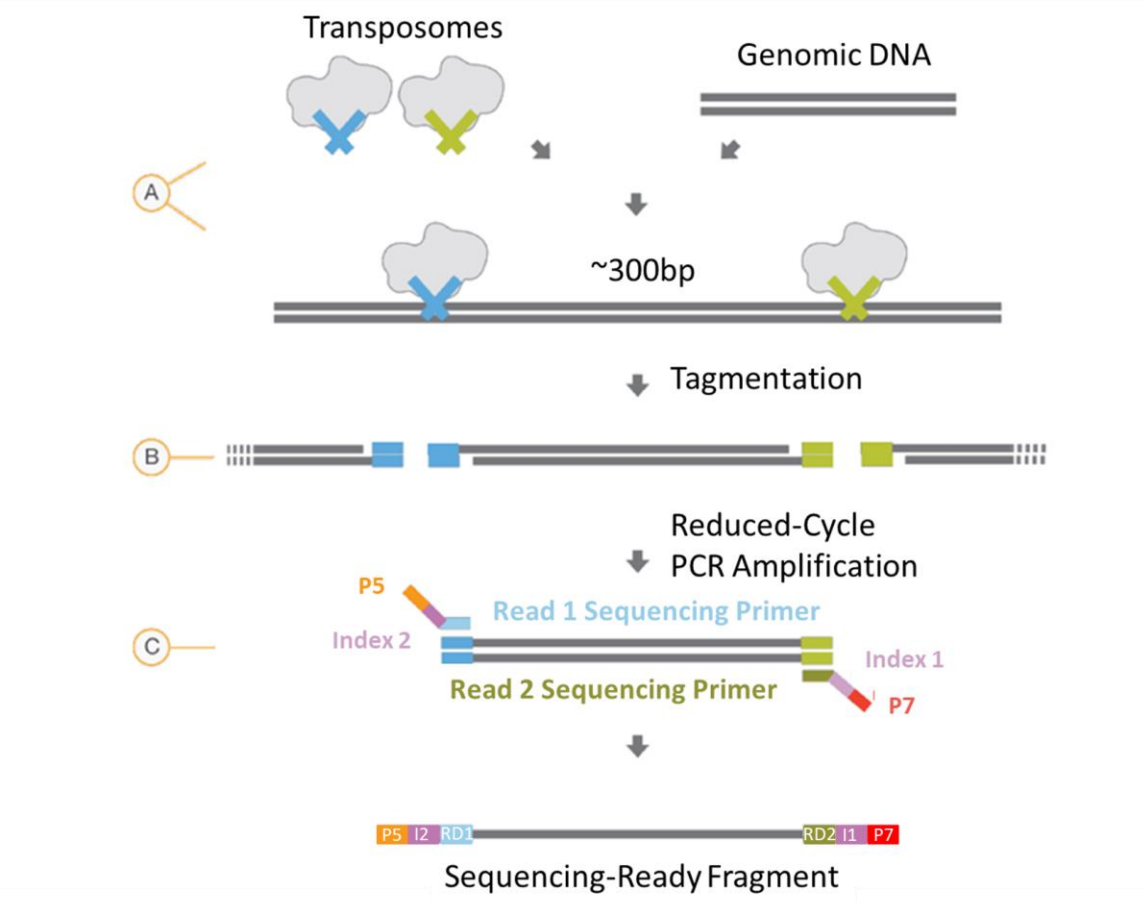

Figure 9. Nextera XT Assay Workflow[31]. (A) Transposomes are combined with the sample DNA. (B) Tagmentation, fragmenting and tagging, of the sample DNA is performed. (C) limited-cycle PCR adding index adapter sequences is done producing a sequencing ready fragment. Figure from Illumina [31].

Assembly of Consensus Sequences from Illumina

Paired-end reads were obtained as fastq files and demultiplexed into individual fastq files associated with each Nextera XT index combination using bcl2fastq[32]. Each fastq set was then trimmed using Trimmomatic[33] to remove low-quality base calls (under Q20) and primer and adapter sequences. The trimmed reads were then mapped to the human genome (hg19) as well as 3 consensus HIV sequences via Burrows-Wheeler transform implemented in BWA[34]. Reads that were mapped to any of the 3 consensus HIV sequences, unmapped, or with a MAPQ-score of 20 or

less against the human genome were extracted using picard (https://broadinstitute.github.io/picard/) and carried over into the final assembly steps. A first pass assembly of cleaned reads was performed in SPAdes[35] and then refined in an HIV-specific reference-guided assembler, SHIVER[36], using *de novo* assembled contigs (at least 250 nucleotides (nt)) to generate the final consensus sequence. Final consensus sequences (at least 250nt) were then scanned for 80% supermajority base-call purity in the mapped reads to omit PCR and sequencing errors. Bases called in the consensus sequences were required to have a minimum of 15 reads, nucleotide positions having between 1 and 14 reads are designated N. Code and environment for assembling and performing QC on the MiSeq reads are available at https://github.com/michaelbale/NGSTools/.

Sequence Analysis

Sequences obtained by NFL amplification and sequencing were first assembled and aligned using the ClustalW alignment tool[37] which aligns sequences based on their pairwise genetic distances. The online software, MEGA7[38], was used to construct neighbor joining, p-distance, phylogenetic trees to visualize rakes of identical sequences.

RESULTS AND DISCUSSION

Patient 1 is a donor with previous integration site and plasma and proviral sequence characterization [1, 4, 13, 24, 29]. Thirty-four populations (or rakes) of identical sub-genomic proviral sequences, called "possible cell clones" in Musick, et al. were also previously characterized for their RNA expression levels[1]. Here, I determined if the proviral sequences identical in the P6-PR-RT region resulted from clonal expansion of singly infected T cells or from infection of multiple T cells by a common viral ancestor prior to ART initiation or during ART interruption. To screen for the 34 P6-PR-RT proviral sequences within potential cell clones,[29] MDA-SGS was performed on EM and CTM T cells collected at two timepoints on ART (**Table 1**). Phylogenetic trees were constructed from 132 P6-PR-RT proviral sequences obtained from the MDA-SGS analyses (**Figure 10**). Sequences matching 20 of the 34 rakes previously described in Musick et al.[1] were identified by MDA-SGS, including those reported to be replication-competent[24] (AMBI-1, Outgrowth-1 (OG-1), and Outgrowth-2 (OG-2)). The AMBI-1 rake was the largest detected in the analysis with 30 total P6-PR-RT sequences, 27 in EM and 3 in CTM. Five of the rakes matching P6-PR-RT sequences in Musick, et al. are known to be defective, containing hypermutations or stop codons in P6-PR-RT (**Figure 10**, orange numbers). The remaining 12 rakes detected by MDA-SGS and matching the previously reported rakes are not known to be defective but did not replicate in culture (**Figure 10**, blue numbers).

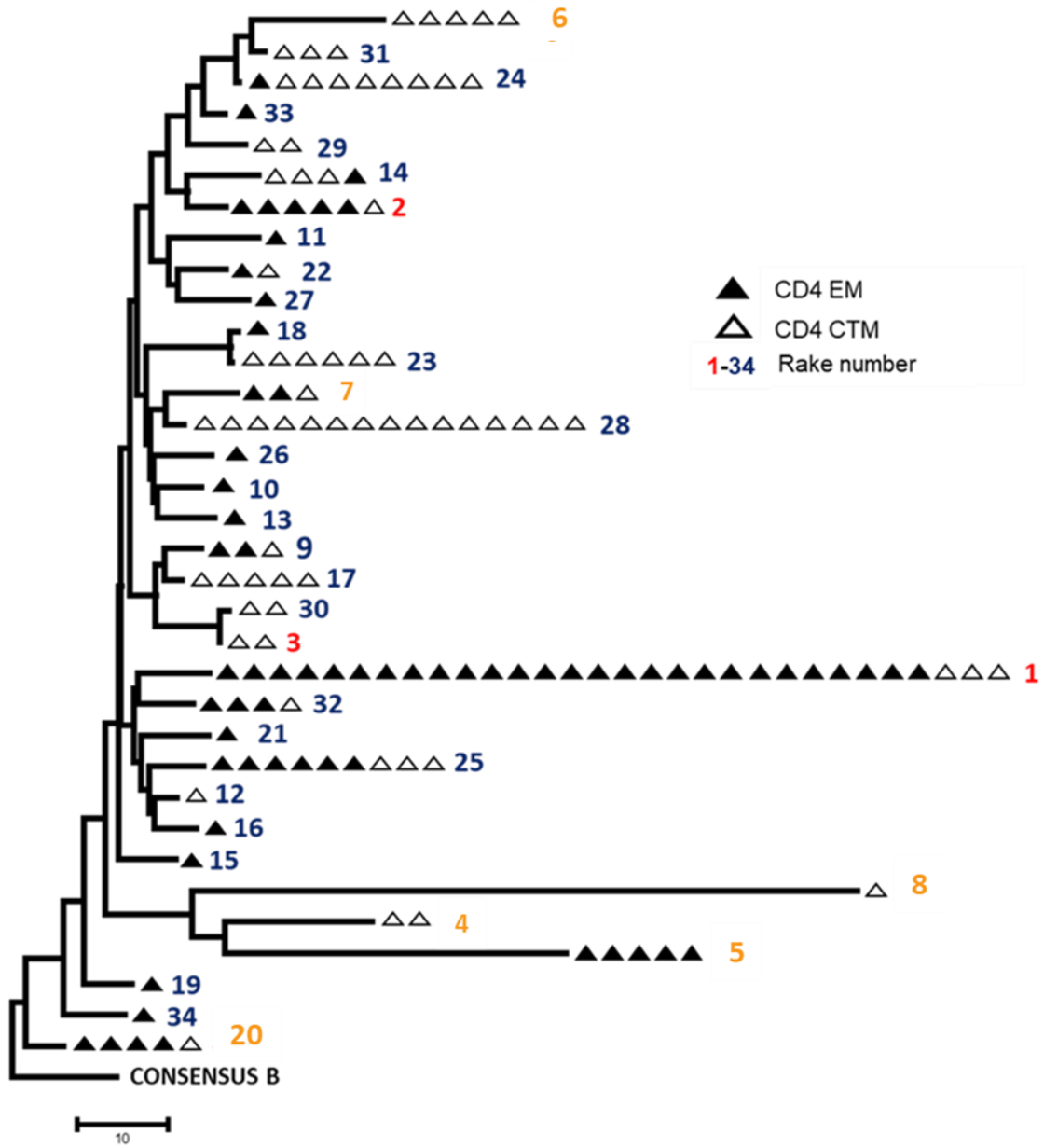

Figure 10. Patient 1 MDA-SGS Phylogenetic Tree of the 34 Rakes of P6-PR-RT Identical Sequences Matching Those in Musick, et al[1]. Sequences shown were obtained from CD4+ EM T cells (black triangles) and CD4+ CTM T cells (open white triangles). Rake numbers shown in different colors are those with P6-PR-RT sequences matching proviruses that are replication-competent proviruses (red numbers), those that are known to be defective (orange numbers), and those that are unknown to be defective but were not recovered in culture (blue numbers).

Integration Site Analyses Confirm Infected Cell Clones in Patient 1

MDA wells from the 20 possible infected cell clones described above were analyzed by ISA to determine if the identical P6-PR-RT sequences resulted from cell proliferation or from multiple cells that were infected by a common viral ancestor prior to ART or during an ART interruption. Identical proviral sequences with the same integration site in the host genome are confirmed as infected cell clones, as this event is highly statistically improbable to have occurred from distinct infection events. Each rake of identical P6-PR-RT sequences had 3 possible outcomes based on our level of sampling (**Figure 11**, hypothetical example): *Outcome 1*- They contained only identical integration sites at our level of sampling, indicating clonal expansion of an infected cell (**Table 3**), *Outcome 2*- They contained both identical and different integration sites, indicating both a cell clone and cells infected with a common ancestor (**Table 4**), and *Outcome 3*- They contained only different integration sites, indicating that the rakes were formed from multiple cells infected with common or closely related ancestors (**Table 5**).

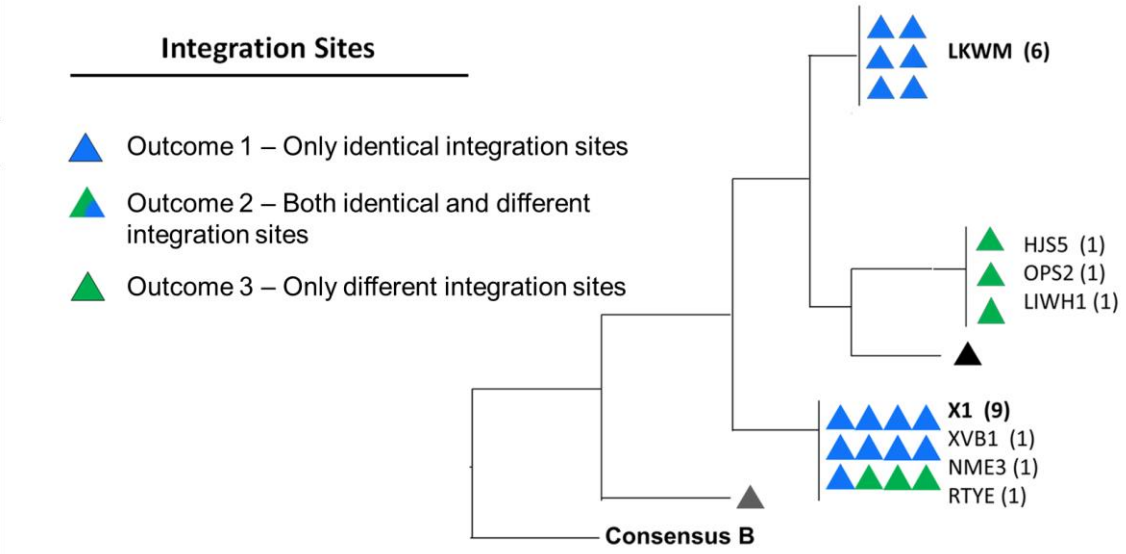

Figure 11. Hypothetical Tree of Three Possible Outcomes from ISA Results. Identical sub-genomic sequences within a rake can contain identical integration sites (blue triangles), different integration sites (green triangles), and both identical and different integration sites (both blue and green triangles).

The ISA results for 9 of the rakes shown in Figure 10 confirmed the proviruses as originating from cell clones as they all had the same site of integration into the host genome at my level of sampling (*Outcome 1*) (**Table 3**). Although the sampling level was shallow, these results confirm the assumption of Musick, et al. that identical P6-PR-RT sequences often result from the proliferation of an infected T cell. Of the 9 rakes observed in outcome 1, 2 were found in only EM, 4 in only CTM, and 3 in both. These results suggest that infected cells can expand and differentiate despite infection. Although Musick, et al. made this suggestion it was not confirmed by integration site analysis. My data demonstrates that expanded cell clones can include members across various cell subsets, indicating that activation in response to a cognate antigen is often not enough to induce expression of integrated proviruses to levels that are cytopathic or to subject the infected cell to CTL killing. Musick et al. detected 3 replication-competent intact proviruses, 5 defective proviruses due to hypermutations, and 26 proviruses that were not known to be intact or defective. The

defective proviruses, rakes 4-8, previously found by Musick et al. were also detected here to contain identical integration sites (**Table 3**), confirming the proviruses as originating from cell clones. This finding was expected since the defects were identical in the subgenomic region sequenced for each provirus. The fraction of cells within these clones that had HIV RNA was reported in Musick, et al. to range from 2-65%, demonstrating that most cell clones have proviruses that are transcriptionally silent despite their activation and differentiation. The genes (or nearest gene) for the integration sites in these rakes are shown in Table 3. All were previously reported in Maldarelli et al.[4] as genes known to be associated with cell growth, mitosis, ATP synthesis, and T cell development or clonal expansion. Whether these integration events drive expansion is not known except for BACH2 and STAT5B (CROI 2020- John Coffin Theme Discussion). Further studies are needed to determine the role of specific sites of integration in cellular proliferation of infected cells.

Table 3. *Outcome 1*- Rakes of Identical Sub-Genomic Proviral Sequences Where Only Identical Integration Sites Were Observed

Rake ID	# of MDA wells with identical P6-PR-RT sequences	# of MDA wells with failed ISA	Cell subsets (EM, CTM, or both)	Gene of integration (# of times observed)	Within or between genes	Orientation relative to gene	Cell associated RNA expression (%)
4	2	0	CTM	*BACH2*-1, *FAM76B* and *RANBP9**	Within	With	0.9
5	5	1	EM	*ATP6V1G3* (4)	Within	With	1.8
6	5	1	CTM	*BACH2*-2 (4)	Within	With	3.5
7	3	1	Both	*MAP4* (2)	Within	Against	6.4
9	3	1	Both	*TTC37* (2)	Within	Against	1.3
17	5	3	CTM	*DDX6* (2)	Within	Against	2.9
20	4	2	EM	*STAT5B* (2)	Within	Against	3.5
22	2	0	Both	*WASF2* (2)	Within	With	7.0
29	2	0	CTM	*BACH2*-3 (2)	Within	With	17.5

*Multiple integration sites in a single MDA well with identical sub-genomic sequence

The ISA results in **Table 4** show the presence of both identical and different integration sites within rakes of identical P6-PR-RT sequences, confirming some of the proviruses as originating from cell clones and others as being from long-lived or small cell clones for which only one cell was sampled in my analysis (*Outcome 2*). Finding sequences identical in P6-PR-RT but with different integration sites implies that a common viral ancestor infected multiple cells either before ART or during ART interruptions. Of note, these results also warn that SGS can over-estimate the size of predicted infected cell clones and the fraction of cells within them that have

HIV RNA. Within *Outcome 2* were the proviruses with P6-PR-RT regions matching the replication competent AMBI-1[24]. Of the 30 MDA wells with P6-PR-RT that matched AMBI-1, 22 had successful ISA and 19 had the expected AMBI-1 integration site. These results suggest that, with smaller sampling, this rake would likely have fallen under outcome 1 rather than outcome 2, supporting the idea that integration sites detected only once are probably from a cell that is a member of a small T cell clone. This observation also highlights that our "outcomes" are arbitrarily based on our level of sampling and that, if a very deep sampling could be done, all rakes of identical SGS might be found to consist of both identical and single integration sites. In Musick, et al., it was shown that only about 2% of the cells in the AMBI-1 clone had HIV RNA and these cells were found to be of the CTM and EM subset. Here, the AMBI-1 integration site was observed 18 times in the EM subset and only once in the CTM subset. Similarly, Musick et al. reported 88% of AMBI in EM, 12% in CTM, and none in the naïve subset. These data support our hypothesis that infected T cells can expand and differentiate without inducing the expression of the integrated provirus, including those that are inducible to produce replication-competent virus.

The integration sites in *TAF6L, BACH2*[4]*, MKL2*[1], and *FAM76B* were the same integration sites reported by Maldarelli et al.[4]. These genes are associated with cell growth and division, specifically as it relates to T cell development. *IER3* and *RNF157,* also found in this set of data, are involved in the prevention of cellular apoptosis[39, 40]. Further studies are needed to determine if integration into any of these genes play a role in cell survival or proliferation. As it relates to curative interventions, immune (i.e. cytotoxic T lymphocytes (CTL)) and selection pressures favoring persistence of a sequence may be distinct from what precipitated expansion of the infected clone (i.e. T cell receptor (TCR) expansion). Therefore, proviral sequences and integration sites may favor persistence in some infected cell clones.

Table 4. *Outcome 2*- Rakes of Identical P6-PR-RT Sequences Where Both Identical and Different Integration Sites Were Observed

Rake ID	# of MDA wells with entical P6-PR-RT sequences	# MDA wells with failed ISA	Cell subsets (EM, CTM, or both)	Gene of integration (# of times observed)	Within or between genes	Orientation (with or against gene)	Cell associated RNA expression (%)
1 (AMBI-1)	30	8	Both	AMBI-1 (19) *RAPH1* (1) *LCP1* (1) *MECP2* (1)	Unmapped Within Within Within	Unmapped With With Against	2.3
14	4	1	Both	*TAF6L* (2) *BACH2*-4 (1)	Within Within	Against With	1.8
23	5	2	CTM	*BACH2*-5 (2) *FMR1* (1)	Within Within	With With	7.5
24	9	5	Both	*MKL2*-1 (2) *ZNF567* (1) *BACH2*-6 (1)	Within Within Within	With Against With	8
28	15	8	CTM	*FAM76B* (5) *ABCA9* (1) *MKL2*-2 (1)	Between Between Within	With With With	16.7

ISA on 6 additional rakes of identical P6-PR-RT sequences yielded only different integration sites (**Table 5**), demonstrating that many cells can be infected by the same or by closely related ancestors. This finding emphasizes the importance of not assuming that identical sub-genomic sequences detected by SGS must result from clonal expansion of a single infected cell even when the genetic diversity of the HIV population is very high, as in this donor. Of these 6, 2 have P6-PR-RT matching the replication-competent proviruses OG-1 and OG-2. These results suggest that the cell clones that harbor the intact OG-1 and OG-2 proviruses must be very small. To determine if any of the integration sites here are of the real replication-competent/intact OG-1 or OG-2 proviruses, we attempted NFL amplification and sequencing of each (described in the next section on NFL).

Table 5. *Outcome 3*- Rakes of Identical P6-PR-RT Sequences Where Only Different Integration Sites Were Observed

Rake ID	MDA wells with identical P6-PR-RT sequences	# MDA wells with failed ISA	Cell subsets (EM, CTM, or both)	Gene of integration (# of times observed)	Within or between genes	Orientation (with or against the gene)	Cell associated RNA expression (%)
2 (OG-1)	6	2	Both	*FTSJD2* (1) *CCDC91* (1) *IER3* (1) *RNF157* (1)	Within Within Between Within	Against Against Against Against	1.2
3 (OG-2)	2	0	CTM	*DOCK8* (1) *ARHGEF6* (1)	Within Within	With Against	8.8
25	9	3	Both	*OR4E2* (1) *ANKRD13D* (1) *L3MBTL3* (1) *MAML2* (1) *FAM168A* (1), NR2C2 (1)	Between Within Within Within Within	Against Against Against Against With	11.4
31	3	2	CTM	*MKL2*-3 (1)	Within	With	22.8
30	2	0	CTM	*ATL1* (1) and *CDC27** (1)	Within	Against and with	19.3
32	4	1	Both	*MEDL13L* (1) *TNRC6C* (1)	Within Within	Against With	33.3

Multiple integration sites in a single MDA well with identical sub-genomic sequence

ISA on 20 rakes demonstrated that most rakes of identical P6-PR-RT also have identical integration sites (13/20 rakes of identical P6-PR-RT sequences had integration site matches). However, 5 rakes of identical P6-PR-RT sequences also had different integration sites and 6 had only different integration sites, indicating that their origin is from the infection of multiple T cells by a common (or very closely related) viral ancestor prior to ART initiation or during ART interruption. A summary of our ISA analyses described previously in Tables 3-5 are in **Figure 12**.

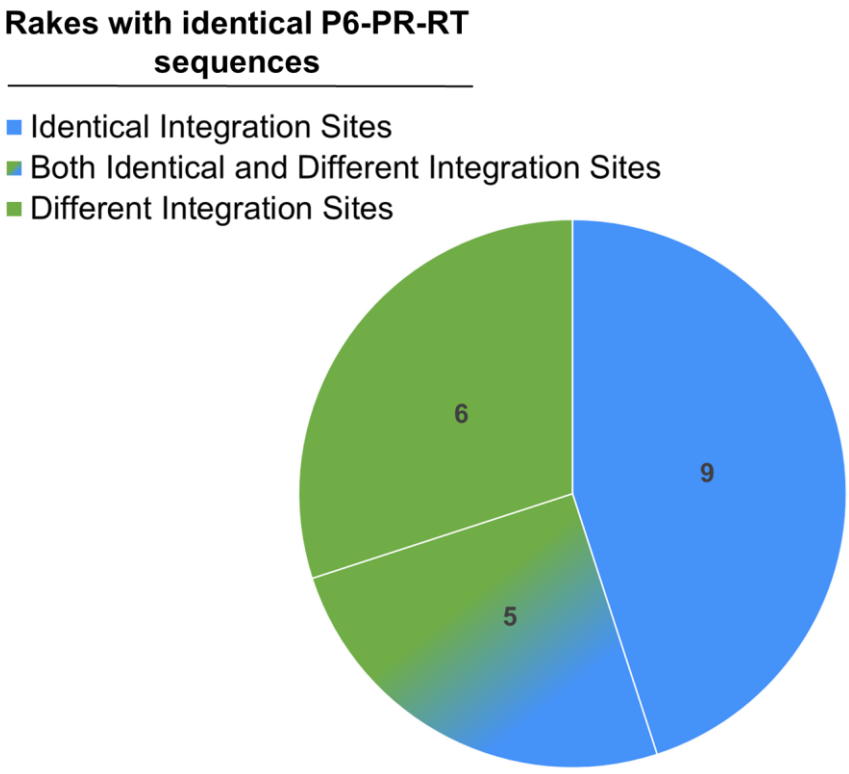

Figure 12. Number of Rakes with Identical, Different, or Both Identical and Different Integration Sites. Nine of the 20 integration sites were observed to have only identical integration sites, 5 were observed to have both identical and different sites, and 6 were observed to have only different sites.

NFL Amplification and Sequencing Reveals Genetically Identical Proviruses in Cell Clones and Infection of Multiple Cells with Common Viral Ancestors

NFL amplicons were generated and sequenced for 8 rakes of identical P6-PR-RT sequences obtained by MDA-SGS to determine the genetic sequence and proviral structure of those that were confirmed to be in cell clones and of those that appeared to result from infection of multiple cells with common viral ancestors **(Figure 13-15)**. Since integration sites were already determined for sequences within these rakes, the proviral sequences were grouped in Figure 13 by their outcomes (described previously). Overall, the NFL results showed 6 intact sequences, 11 with large internal deletions, 2 with small packaging signal deletions, and 2 hypermutants. While the sampling

presented here was not deep, these genomes have similar proviral landscapes reported in literature[27, 29, 41].

Figure 13a. shows the genetic structures of proviruses in the MDA wells with the same integration sites. For example, rakes 4 and 6 each yielded proviruses with identical large internal deletions in the 3' end of the genome and identical integration sites, supporting our previous conclusion that they resulted from proliferation of single infected cells (**Figure 13a**). Both rakes were previously reported to have defective proviruses due to hypermutation in the P6-PR-RT region as well. One MDA well in rake 4 contained multiple integration sites (*BACH2*, *FAM65B, and RANBP9*), indicating that multiple proviruses were present in the well as expected to occur in some wells in limiting dilution-based experiments. The proviral structure obtained from that well is most likely from the provirus integrated in the *BACH2* gene since it matches the proviral structure in another MDA well with the same integration site (**Figure 13**). The other two proviruses in the well may contain deletions in the NFL primer binding regions or could be solo LTRs. Both proviruses in rakes 4 and 6 included complete deletions of *tat*. As reported in Musick, et al., few cells in these clones had HIV RNA (0.9% in rake 4, 3.5% in rake 6). Defective genomes lacking essential RNA transcription/export elements may promote transcriptional silence, evasion of CTL immunity, and persistence/maintenance[42-44].

Proviruses in *Outcome 2* include those that resulted from cell proliferation but also from infection of cells with the same viral ancestor (**Figure 13b.**). Although the P6-PR-RT sequences are identical in these outcomes, the internal deletions and integration sites may vary. Proviruses in rake 1 contained both identical and different proviral structures, consistent with the integration site data that showed both identical (AMBI-1) and different integration sites (e.g. *RAPH1*). AMBI-1 proviral structures were intact and identical in 4 MDA wells as expected. A fifth MDA well was

determined to have a different site of integration, *RAPH1*, and had a provirus that was identical to the AMBI-1 provirus but for a deletion at the 3' end. This finding implies that proviruses containing identical sequences, but different deletions, can originate from a common viral ancestor that infected multiple cells. The AMBI-1 sequence in the MDA wells is consistent with the proviral sequence that was published in Simonetti et al.[24]. These data further demonstrate that SGS alone is not sufficient to establish cellular clonality and can over-estimate the true size of the clone.

Proviruses from rakes in *Outcome 3* were expected to have the same sequence but different deletions since the integration site and P6-PR-RT SGS data suggested that they resulted from the infection of multiple cells with a common ancestor (**Figure 13c.**). NFL on rake 24 in this category revealed both intact and defective proviruses with genetically identical sequences but some containing deletions. The provirus integrated in the *ZNF567* gene was intact according to the Proviral Sequence Annotation & Intactness Test (ProSeq-IT) (https://psd.cancer.gov/tools/pvs_annot.php). However, while genetically intact, it has not been confirmed to be replication competent as it has not been assayed for viral replication *in vitro*. Proviruses in two other MDA wells in rake 24 were identical in sequence and internal deletions while having, apparently, different integration sites. Because multiple integration sites were observed in these wells, indicating that multiple proviruses were present in the wells, it is possible that the integration site of the identical proviruses observed was not detected by ISA in one or both MDA wells. MDA products were screened using P6-PR-RT amplification and sequencing. Some of the MDA wells may contain multiple proviruses that are deleted in this region and, therefore, were not detected in the original screen. When ISA was performed, a fraction of the MDA product was used. If multiple proviruses were present in the well, it is possible that the integration sites of some proviruses in the well were detected, while others were not. To overcome this challenge,

future studies may screen MDA wells for LTR or for multiple regions of the HIV genome. Together, the NFL data from all three outcomes demonstrate that rakes of identical sub-genomic sequences result from clonal expansion of single cells and from multiple cells being infected with the same viral ancestor prior to ART or during an ART interruption.

The virogram of NFL sequences from rake 30 showed two completely identical proviruses (**Figure 13d.**). However, their integration site is unresolved, since one MDA well generated two integration sites and the other had a failed ISA. The NFL sequencing showed nearly intact proviruses with the exception of a small (28bp) deletion in the packaging signal, a previously reported common region for deletions[29] (**Figure 13c**). The combined NFL and ISA results inform of the overall landscape of proviruses within rakes of identical sub-genomic sequences in Patient 1.

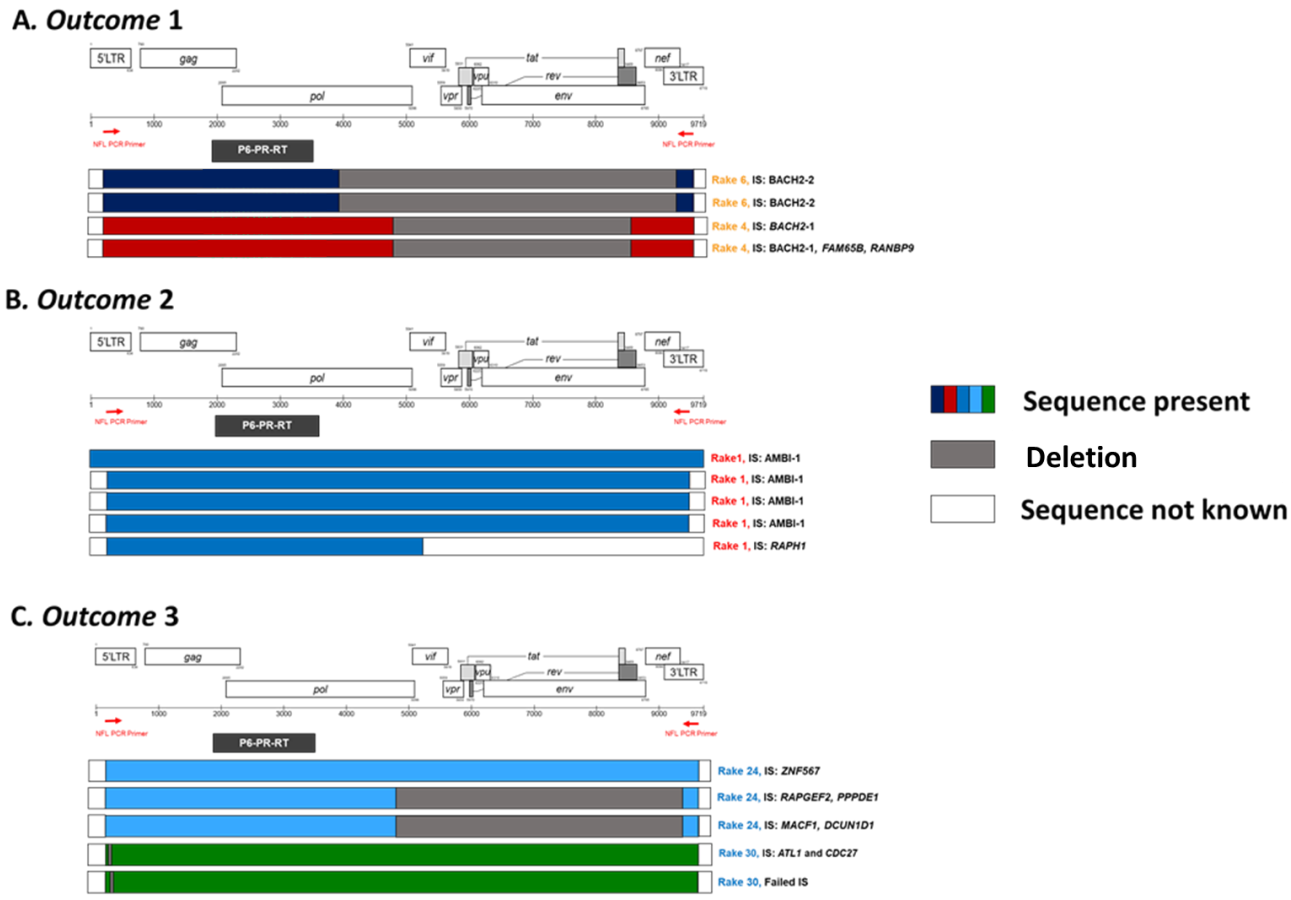

Figure 13. Virogram of Selected Rakes Found in Patient 1. A) Rakes containing identical integration sites and proviral structures. B) Rakes containing both identical and different integration sites and proviral structures. C) Rakes containing different integration sites but identical proviral structures. The 10kb HIV-1 viral genome is shown at the top. Identical sequences are represented by same bar color (i.e. identical sequences in rake 1 are shown in dark blue bar). Unknown sequences (outside of primer region) are represented by white bar. Deletions are represented by gray bar. Near-full length (NFL) PCR primers are represented by red arrows. Rake numbers and their corresponding integration sites are indicated next to their NFL sequences. The sub-genomic region used to screen, P6-PR-RT, is indicated.

Although all the NFL sequences obtained from rakes in Figure 13 were found to be genetically identical, demonstrating their source as clonal proliferation or infection with a common ancestor, we did find some NFL proviruses in outcomes 2 and 3 to have similar, but different proviral sequences (**Figure 14**). In some examples, the sequences were different by only a few nucleotides which could be from amplification or sequencing errors. However, others were more genetically distant, indicating that they resulted from infection by closely related, but not identical,

ancestors. An example is rake 2 (OG-1) (**Figure 14a**). NFL sequencing of MDA wells in rake 2 revealed different proviral sequences and structures. The first sequence shown in the virogram is the reference sequence for OG-1 obtained from the supernatant of viral outgrowth in culture[24]. The second sequence was obtained from an MDA well with the *CCDC91* integration site. The sequence is a close match to the OG-1 reference and is intact by ProSeq-IT test, making it the likely source of OG-1 outgrowth. The nucleotide differences (n=4) may have resulted from MDA, PCR, and/or sequencing errors or from mutations that occurred in the reference upon passaging in culture. These results strongly imply that the integration site for the replication-competent provirus OG-1[24] is in the CCDC91 gene. The third NFL from rake 2 was from an MDA well with a *RNF157* integration site and had a total of 15 nucleotide differences compared to the OG-1 reference. This provirus probably originated from a related, but not identical, ancestor (or descendant) of the OG-1 virus.

Rake 25 is a good example of infection of multiple cells with a common ancestor (**Figure 14b.**). Although these MDA wells were found to have identical sequences but for a few nucleotide differences that likely resulted from PCR and/or sequencing errors, NFL revealed differential deletions at the 3' end. This finding suggests that 3 different cells were infected by the same ancestor, that different errors in reverse transcription occurred in each resulting in different deletions, and that each infected cell persisted and clonally expanded. Such a finding implies that the seeding of cells that comprise the infected cell clones that persist on ART is a common occurrence. To my knowledge, this is the first report showing evidence that clonally expanded cells are seeded frequently.

Rake 3 had P6-PR-RT sequences that matched OG-2, the third replication-competent provirus reported in Patient 1[24] (**Figure 14c.**). The full-length sequence of OG-2 is not known

because NFL was not successful on the supernatant of viral outgrowth experiments. Although I identified two MDA wells with P6-PR-RT matching OG-2, neither yielded intact proviral sequences, suggesting that the OG-2 cell clone is probably very small. Deeper sampling is needed to identify the integration site for the OG-2 provirus.

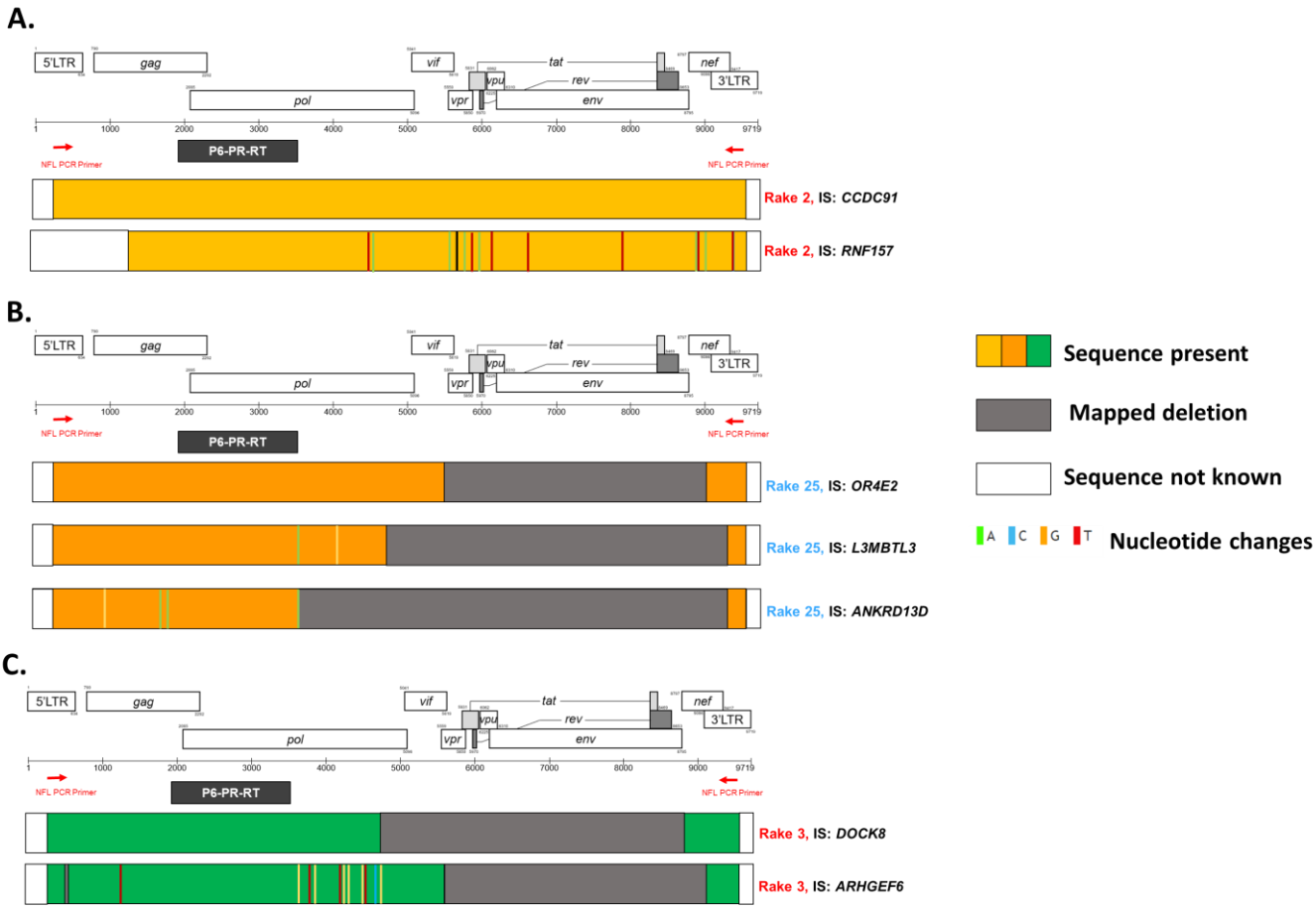

Figure 14. Virogram of Proviruses in Rake 2, 25, and 3. (A) Rake 2 (OG-1) provirus structure where the 9.7kb HIV-1 viral genome is shown at the top. Unknown sequences are represented by white bars. NFL PCR primers are represented by red arrows. The sub-genomic region used to screen, P6-PR-RT, is indicated. Nucleotide differences are represented by colored bars. (B) Rake 25 provirus structure where the sequence present is represented by the orange bars and mapped deletions are represented by gray bars. (C) Rake 3 (OG-2) provirus structure where the sequence present is represented by green bars and mapped deletions are represented by gray bars.

CONCLUSIONS

Using the MDA-SGS[29] approach, which informs both the proviral site of integration and proviral structure, I aimed to investigate the 34 possible infected cell clones reported in Musick et al.[1] to determine if they are, indeed, infected cell clones or if they originated instead from the infection of multiple cells with a common or genetically related ancestor. In total, 20 of the 34 rakes of single-genome sequences identical in the P6-PR-RT region were detected by MDA-SGS. Nine were found to have identical integration sites only and therefore, resulted from the clonal expansion of a single infected cell, as hypothesized. However, 5 rakes had both identical and different integration sites and an additional 6 had only different integration sites detected. These results confirmed Musick et al's study in that at least 14 of the 20 rakes (70%) contained proviruses within at least one infected cell clone and support those of Patro et al.[29], in which we concluded that the use of identical sub-genomic proviral sequences to identify infected cell clones is not sufficient and integration sites are required to draw this conclusion. Limitations of this study include the depth of sampling, which was limited by the number of cells that were available in each T cell subset.

The proviral structures determined by NFL sequencing revealed mostly identical sequences within the defined phylogenetic rakes regardless of their integration site. The identical proviruses with different integration sites and different internal deletions likely result from the infection of multiple cells by a common viral ancestor and implies that infected cell clones are seeded frequently prior to ART or upon its interruption. Defective proviruses, those containing large internal deletions, packaging signal deletions, or hypermutation, were most common. This finding is not unexpected as it was previously described by multiple groups[45-47]. However, 3 proviruses were found to be intact, two that were known to be replication-competent, AMBI-1 and OG-1, and one that was newly discovered in this study. The intact AMBI-1 and OG-1 proviruses were

predominantly found in EM and the newly discovered intact provirus in rake 24 was found only in CTM. Although the TCR sequence and cognate antigen of these CD4 clones are unknown, it is possible that the large EM constituent of AMBI-1 is due to expansion in response to an antigen. Both the AMBI-1 rake and OG-1 rake were previously reported by Musick et al. to contain 2.3% and 1.2% of the cells with detectable HIV RNA. Rake 24 was found to have more cells with HIV RNA (8%) than the AMBI-1 and OG-1 rakes. The reason for the higher fraction of cells with HIV RNA in rake 24 is not known but could be due to the cell subset that harbors to the location of the integration in the host genome. The integration sites for the 3 intact proviruses were observed between genes (AMBI-1) or within introns and oriented against the genes for both the CCDC91 integration site from OG-1 and the ZNF567 integration site from rake 24. These findings do not directly suggest that intact HIV has a preference for integration into introns or to be oriented against the host gene, but that integration events into these locations may be selected for persistence as suggested previously[28]. Integration sites in many of the confirmed clones with defective proviruses were in genes associated with cellular proliferation, apoptosis, and mitosis, suggesting that the integration events may have contributed to cellular proliferation. For example, Dr. John Coffin's theme discussion at CROI 2020 reported that proviral integration in BACH2 and STAT5B genes drive clonal expansion.

Rebound viremia when ART is interrupted is established by viral replication and spreading infection, possibly from as few as a single replication-competent provirus[26, 48]. Recent studies have attempted to determine the "age" of these proviruses (i.e. whether the source of rebound viremia is from cells that were infected just before ART initiation or long before) using the quantitative viral outgrowth assay (QVOA)[49] and phylogeny. Other studies have focused on the effect of ART treatment interruptions on the size or diversity of the reservoir[26]. Future studies using MDA-SGS

may contribute to addressing these questions by using phylogenetic analyses to estimate the "age" of intact proviruses in cell clones and determining if new cell clones are seeded during treatment interruptions. MDA-SGS may also be used to determine the viral sequence of the common ancestors that gave rise to cell clones with defective proviruses. Determining, predicting, and eliminating the source of rebound viruses is important for the development of future potential curative interventions. Distinguishing the sources of replication-competent proviruses that constitute the HIV reservoir on ART is crucial to revealing new targets, evaluating the effects of cure strategies for HIV, and guiding future directions of HIV research.

We, and others, have used SGS to identify potential infected cell clones[1, 4, 24, 27, 50]. Here, we emphasize that proviruses containing identical sub-genomic sequences can have different integration sites, suggesting that these proviruses result from the infection of multiple cells by a common ancestor. Thus, previous findings have underestimated the contribution of genetic bottlenecks in shaping the identical sub-genomic sequences in the proviral landscape (**Figure 15**). Better understanding the establishment and composition of the HIV reservoir may help better predict and target virus that rebounds upon cessation of ART.

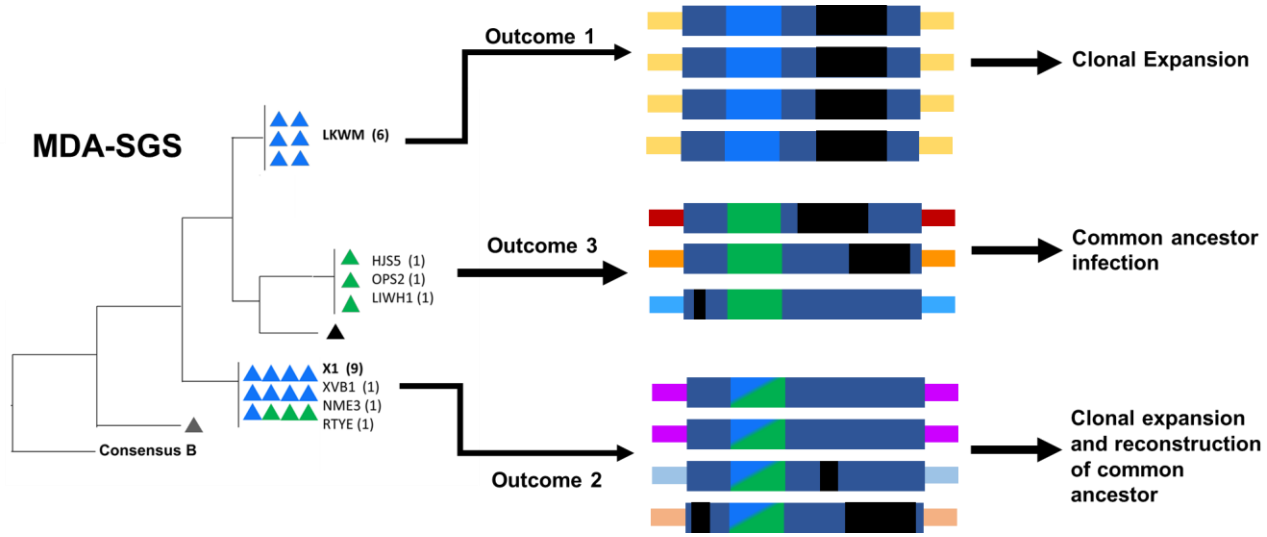

Figure 15. Model for the origin of identical sequences that persist on ART. Rakes of identical sub-genomic sequences are shown to contain identical integration sites (blue triangles), different integration sites (green triangles), and both identical and different integration sites (both blue and green triangles) leading to 3 outcomes. Integration sites are represented by identical colors (yellow) for outcome 1, different colors (red, orange, and blue) for outcome 3, and both identical and different colors (purple, light blue, and light orange) for outcome 2. Virogram models where sequence is present is represented by dark blue bars, subgenomic sequences matching the triangles in the phylogenetic tree are represented by blue, green and both blue and green bars, and deletions are shown by black bars. Identical sub-genomic sequences either originate from clonal expansion or from multiply infected cells by a common viral ancestor.

REFERENCES

1. Musick A, Spindler J, Boritz E, Pérez L, Crespo-Vélez D, Patro SC, Sobolewski MD, Bale MJ, Reid C, Keele BF, Shao W, Wiegand A, Simonetti FR, Mellors JW, Hughes SH, Coffin JM, Maldarelli F, Kearney MF. HIV Infected T Cells Can Proliferate in vivo Without Inducing Expression of the Integrated Provirus. Frontiers. 2019, submitted.
2. UNAIDS. Global HIV & AIDS statistics — 2019 fact sheet 2019 [cited 2019 7/27/2019]. Available from: https://www.unaids.org/en/resources/fact-sheet.
3. Burdick RC, Delviks-Frankenberry KA, Chen J, Janaka SK, Sastri J, Hu WS, Pathak VK. Dynamics and regulation of nuclear import and nuclear movements of HIV-1 complexes. PLoS Pathog. 2017;13(8):e1006570. Epub 2017/08/23. doi: 10.1371/journal.ppat.1006570. PubMed PMID: 28827840; PMCID: PMC5578721.
4. Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, Spindler J, Ferris AL, Mellors JW, Kearney MF, Coffin JM, Hughes SH. HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. Science. 2014;345(6193):179-83. Epub 2014/06/28. doi: 10.1126/science.1254194. PubMed PMID: 24968937; PMCID: PMC4262401.
5. Wagner TA, McLaughlin S, Garg K, Cheung CY, Larsen BB, Styrchak S, Huang HC, Edlefsen PT, Mullins JI, Frenkel LM. HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. Science. 2014;345(6196):570-3. Epub 2014/07/12. doi: 10.1126/science.1256304. PubMed PMID: 25011556; PMCID: PMC4230336.
6. Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature. 1995;373(6510):123-6. Epub 1995/01/12. doi: 10.1038/373123a0. PubMed PMID: 7816094.
7. Maldarelli F, Palmer S, King MS, Wiegand A, Polis MA, Mican J, Kovacs JA, Davey RT, Rock-Kress D, Dewar R, Liu S, Metcalf JA, Rehm C, Brun SC, Hanna GJ, Kempf DJ, Coffin JM, Mellors JW. ART suppresses plasma HIV-1 RNA to a stable set point predicted by pretherapy viremia. PLoS Pathog. 2007;3(4):e46. Epub 2007/04/07. doi: 10.1371/journal.ppat.0030046. PubMed PMID: 17411338; PMCID: PMC1847689.
8. Palmer S, Wiegand AP, Maldarelli F, Bazmi H, Mican JM, Polis M, Dewar RL, Planta A, Liu S, Metcalf JA, Mellors JW, Coffin JM. New real-time reverse transcriptase-initiated PCR assay with single-copy sensitivity for human immunodeficiency virus type 1 RNA in plasma. J Clin Microbiol. 2003;41(10):4531-6. Epub 2003/10/09. doi: 10.1128/jcm.41.10.4531-4536.2003. PubMed PMID: 14532178; PMCID: PMC254331.
9. (WHO) WHO. Available from: https://www.who.int/hiv/topics/drugresistance/en/.
10. Burdick RC, Li C, Munshi M, Rawson JMO, Nagashima K, Hu WS, Pathak VK. HIV-1 uncoats in the nucleus near sites of integration. Proc Natl Acad Sci U S A. 2020;117(10):5486-93. Epub 2020/02/26. doi: 10.1073/pnas.1920631117. PubMed PMID: 32094182; PMCID: PMC7071919.
11. Muesing MA, Smith DH, Cabradilla CD, Benton CV, Lasky LA, Capon DJ. Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus. Nature. 1985;313(6002):450-8. Epub 1985/02/07. doi: 10.1038/313450a0. PubMed PMID: 2982104.
12. Gallo R, Wong-Staal F, Montagnier L, Haseltine WA, Yoshida M. HIV/HTLV gene nomenclature. Nature. 1988;333(6173):504. Epub 1988/06/09. doi: 10.1038/333504a0. PubMed PMID: 2836736.
13. Kearney MF, Spindler J, Shao W, Yu S, Anderson EM, O'Shea A, Rehm C, Poethke C, Kovacs N, Mellors JW, Coffin JM, Maldarelli F. Lack of detectable HIV-1 molecular evolution

during suppressive antiretroviral therapy. PLoS Pathog. 2014;10(3):e1004010. Epub 2014/03/22. doi: 10.1371/journal.ppat.1004010. PubMed PMID: 24651464; PMCID: PMC3961343.
14. Joos B, Fischer M, Kuster H, Pillai SK, Wong JK, Boni J, Hirschel B, Weber R, Trkola A, Gunthard HF, Swiss HIVCS. HIV rebounds from latently infected cells, rather than from continuing low-level replication. Proc Natl Acad Sci U S A. 2008;105(43):16725-30. Epub 2008/10/22. doi: 10.1073/pnas.0804192105. PubMed PMID: 18936487; PMCID: PMC2575487.
15. Kearney MF, Wiegand A, Shao W, Coffin JM, Mellors JW, Lederman M, Gandhi RT, Keele BF, Li JZ. Origin of Rebound Plasma HIV Includes Cells with Identical Proviruses That Are Transcriptionally Active before Stopping of Antiretroviral Therapy. J Virol. 2016;90(3):1369-76. Epub 2015/11/20. doi: 10.1128/JVI.02139-15. PubMed PMID: 26581989; PMCID: PMC4719635.
16. McManus WR, Bale MJ, Spindler J, Wiegand A, Musick A, Patro SC, Sobolewski MD, Musick VK, Anderson EM, Cyktor JC, Halvas EK, Shao W, Wells D, Wu X, Keele BF, Milush JM, Hoh R, Mellors JW, Hughes SH, Deeks SG, Coffin JM, Kearney MF. HIV-1 in lymph nodes is maintained by cellular proliferation during antiretroviral therapy. J Clin Invest. 2019;130. Epub 2019/07/31. doi: 10.1172/JCI126714. PubMed PMID: 31361603.
17. von Stockenstrom S, Odevall L, Lee E, Sinclair E, Bacchetti P, Killian M, Epling L, Shao W, Hoh R, Ho T, Faria NR, Lemey P, Albert J, Hunt P, Loeb L, Pilcher C, Poole L, Hatano H, Somsouk M, Douek D, Boritz E, Deeks SG, Hecht FM, Palmer S. Longitudinal Genetic Characterization Reveals That Cell Proliferation Maintains a Persistent HIV Type 1 DNA Pool During Effective HIV Therapy. J Infect Dis. 2015;212(4):596-607. Epub 2015/02/26. doi: 10.1093/infdis/jiv092. PubMed PMID: 25712966; PMCID: PMC4539896.
18. Brodin J, Zanini F, Thebo L, Lanz C, Bratt G, Neher RA, Albert J. Establishment and stability of the latent HIV-1 DNA reservoir. Elife. 2016;5. Epub 2016/11/18. doi: 10.7554/eLife.18889. PubMed PMID: 27855060; PMCID: PMC5201419.
19. Bailey JR, Sedaghat AR, Kieffer T, Brennan T, Lee PK, Wind-Rotolo M, Haggerty CM, Kamireddi AR, Liu Y, Lee J, Persaud D, Gallant JE, Cofrancesco J, Jr., Quinn TC, Wilke CO, Ray SC, Siliciano JD, Nettles RE, Siliciano RF. Residual human immunodeficiency virus type 1 viremia in some patients on antiretroviral therapy is dominated by a small number of invariant clones rarely found in circulating CD4+ T cells. J Virol. 2006;80(13):6441-57. Epub 2006/06/16. doi: 10.1128/JVI.00591-06. PubMed PMID: 16775332; PMCID: PMC1488985.
20. Bui JK, Sobolewski MD, Keele BF, Spindler J, Musick A, Wiegand A, Luke BT, Shao W, Hughes SH, Coffin JM, Kearney MF, Mellors JW. Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir. PLoS Pathog. 2017;13(3):e1006283. Epub 2017/03/23. doi: 10.1371/journal.ppat.1006283. PubMed PMID: 28328934; PMCID: PMC5378418.
21. Cohn LB, da Silva IT, Valieris R, Huang AS, Lorenzi JCC, Cohen YZ, Pai JA, Butler AL, Caskey M, Jankovic M, Nussenzweig MC. Clonal CD4(+) T cells in the HIV-1 latent reservoir display a distinct gene profile upon reactivation. Nat Med. 2018;24(5):604-9. Epub 2018/04/25. doi: 10.1038/s41591-018-0017-7. PubMed PMID: 29686423; PMCID: PMC5972543.
22. Hosmane NN, Kwon KJ, Bruner KM, Capoferri AA, Beg S, Rosenbloom DI, Keele BF, Ho YC, Siliciano JD, Siliciano RF. Proliferation of latently infected CD4(+) T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics. J Exp Med. 2017;214(4):959-72. Epub 2017/03/28. doi: 10.1084/jem.20170193. PubMed PMID: 28341641; PMCID: PMC5379987.

23. Mansky LM, Temin HM. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J Virol. 1995;69(8):5087-94. Epub 1995/08/01. PubMed PMID: 7541846; PMCID: PMC189326.
24. Simonetti FR, Sobolewski MD, Fyne E, Shao W, Spindler J, Hattori J, Anderson EM, Watters SA, Hill S, Wu X, Wells D, Su L, Luke BT, Halvas EK, Besson G, Penrose KJ, Yang Z, Kwan RW, Van Waes C, Uldrick T, Citrin DE, Kovacs J, Polis MA, Rehm CA, Gorelick R, Piatak M, Keele BF, Kearney MF, Coffin JM, Hughes SH, Mellors JW, Maldarelli F. Clonally expanded CD4+ T cells can produce infectious HIV-1 in vivo. Proc Natl Acad Sci U S A. 2016;113(7):1883-8. Epub 2016/02/10. doi: 10.1073/pnas.1522675113. PubMed PMID: 26858442; PMCID: PMC4763755.
25. Laird GM, Rosenbloom DI, Lai J, Siliciano RF, Siliciano JD. Measuring the Frequency of Latent HIV-1 in Resting CD4(+) T Cells Using a Limiting Dilution Coculture Assay. Methods Mol Biol. 2016;1354:239-53. Epub 2015/12/31. doi: 10.1007/978-1-4939-3046-3_16. PubMed PMID: 26714716.
26. Salantes DB, Zheng Y, Mampe F, Srivastava T, Beg S, Lai J, Li JZ, Tressler RL, Koup RA, Hoxie J, Abdel-Mohsen M, Sherrill-Mix S, McCormick K, Overton ET, Bushman FD, Learn GH, Siliciano RF, Siliciano JM, Tebas P, Bar KJ. HIV-1 latent reservoir size and diversity are stable following brief treatment interruption. J Clin Invest. 2018;128(7):3102-15. Epub 2018/06/19. doi: 10.1172/JCI120194. PubMed PMID: 29911997; PMCID: PMC6026010.
27. Palmer S, Kearney M, Maldarelli F, Halvas EK, Bixby CJ, Bazmi H, Rock D, Falloon J, Davey RT, Jr., Dewar RL, Metcalf JA, Hammer S, Mellors JW, Coffin JM. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. J Clin Microbiol. 2005;43(1):406-13. Epub 2005/01/07. doi: 10.1128/JCM.43.1.406-413.2005. PubMed PMID: 15635002; PMCID: PMC540111.
28. Einkauf KB, Lee GQ, Gao C, Sharaf R, Sun X, Hua S, Chen SM, Jiang C, Lian X, Chowdhury FZ, Rosenberg ES, Chun TW, Li JZ, Yu XG, Lichterfeld M. Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy. J Clin Invest. 2019;129(3):988-98. Epub 2019/01/29. doi: 10.1172/JCI124291. PubMed PMID: 30688658; PMCID: PMC6391088.
29. Patro SC, Brandt LD, Bale MJ, Halvas EK, Joseph KW, Shao W, Wu X, Guo S, Murrell B, Wiegand A, Spindler J, Raley C, Hautman C, Sobolewski M, Fennessey CM, Hu W, Luke B, M. HJ, Niyongabo A, Keele BF, Milush J, Hoh R, Deeks SG, Maldarelli F, Hughes SH, Coffin JM, Rausch JW, Mellors JW, Kearney MF. Combined Full-Length HIV-1 Sequencing and Integration Site Analysis Informs Intra-Patient Viral Dynamics and Reconstruction of Replication-Competent Viral Ancestors. PNAS. 2019.
30. Daria W. Wells SG, Wei Shao, John M. Coffin, Stephen H. Hughes, Xiaolin Wu. An Analytical Pipeline for Identifying and Mapping the Integration Sites of HIV and other Retroviruses. BMC Genomics submitted. 2020.
31. Illumina I. Nextera XT DNA Library Prep. Available from: https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_nextera/nextera-xt/nextera-xt-library-prep-reference-guide-15031942-05.pdf.
32. Illumina I. bcl2fastq and bcl2fastq2 conversion software 2020. Available from: https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html.

33. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20. Epub 2014/04/04. doi: 10.1093/bioinformatics/btu170. PubMed PMID: 24695404; PMCID: PMC4103590.
34. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60. Epub 2009/05/20. doi: 10.1093/bioinformatics/btp324. PubMed PMID: 19451168; PMCID: PMC2705234.
35. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-77. Epub 2012/04/18. doi: 10.1089/cmb.2012.0021. PubMed PMID: 22506599; PMCID: PMC3342519.
36. Wymant C, Blanquart F, Golubchik T, Gall A, Bakker M, Bezemer D, Croucher NJ, Hall M, Hillebregt M, Ong SH, Ratmann O, Albert J, Bannert N, Fellay J, Fransen K, Gourlay A, Grabowski MK, Gunsenheimer-Bartmeyer B, Gunthard HF, Kivela P, Kouyos R, Laeyendecker O, Liitsola K, Meyer L, Porter K, Ristola M, van Sighem A, Berkhout B, Cornelissen M, Kellam P, Reiss P, Fraser C, Collaboration B. Easy and accurate reconstruction of whole HIV genomes from short-read sequence data with shiver. Virus Evol. 2018;4(1):vey007. Epub 2018/06/08. doi: 10.1093/ve/vey007. PubMed PMID: 29876136; PMCID: PMC5961307.
37. EMBL-EBI. ClustalW2 2020. Available from: http://www.ebi.ac.uk/Tools/msa/clustalw2/.
38. Molecular Evolutionary Genetics Analysis (MEGA) Available from: https://www.megasoftware.net/.
39. Arlt A, Schafer H. Role of the immediate early response 3 (IER3) gene in cellular stress response, inflammation and tumorigenesis. Eur J Cell Biol. 2011;90(6-7):545-52. Epub 2010/11/30. doi: 10.1016/j.ejcb.2010.10.002. PubMed PMID: 21112119.
40. Dogan T, Gnad F, Chan J, Phu L, Young A, Chen MJ, Doll S, Stokes MP, Belvin M, Friedman LS, Kirkpatrick DS, Hoeflich KP, Hatzivassiliou G. Role of the E3 ubiquitin ligase RNF157 as a novel downstream effector linking PI3K and MAPK signaling pathways to the cell cycle. J Biol Chem. 2017;292(35):14311-24. Epub 2017/06/29. doi: 10.1074/jbc.M117.792754. PubMed PMID: 28655764; PMCID: PMC5582827.
41. Siliciano JD, Siliciano RF. Recent developments in the search for a cure for HIV-1 infection: targeting the latent reservoir for HIV-1. J Allergy Clin Immunol. 2014;134(1):12-9. Epub 2014/08/15. doi: 10.1016/j.jaci.2014.05.026. PubMed PMID: 25117799.
42. Imamichi H, Dewar RL, Adelsberger JW, Rehm CA, O'Doherty U, Paxinos EE, Fauci AS, Lane HC. 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(31):8783-8. Epub 2016/07/20. doi: 10.1073/pnas.1609057113. PubMed PMID: 27432972; PMCID: PMC4978246.
43. Imamichi H, Smith M, Adelsberger JW, Izumi T, Scrimieri F, Sherman BT, Rehm CA, Imamichi T, Pau A, Catalfamo M, Fauci AS, Lane HC. Defective HIV-1 proviruses produce viral proteins. Proc Natl Acad Sci U S A. 2020;117(7):3704-10. Epub 2020/02/08. doi: 10.1073/pnas.1917876117. PubMed PMID: 32029589; PMCID: PMC7035625.
44. Manzoni TB, Lopez CB. Defective (interfering) viral genomes re-explored: impact on antiviral immunity and virus persistence. Future Virol. 2018;13(7):493-503. Epub 2018/09/25. doi: 10.2217/fvl-2018-0021. PubMed PMID: 30245734; PMCID: PMC6136085.

45. Ho YC, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DI, Lai J, Blankson JN, Siliciano JD, Siliciano RF. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. Cell. 2013;155(3):540-51. Epub 2013/11/19. doi: 10.1016/j.cell.2013.09.020. PubMed PMID: 24243014; PMCID: PMC3896327.
46. Hiener B, Horsburgh BA, Eden JS, Barton K, Schlub TE, Lee E, von Stockenstrom S, Odevall L, Milush JM, Liegler T, Sinclair E, Hoh R, Boritz EA, Douek D, Fromentin R, Chomont N, Deeks SG, Hecht FM, Palmer S. Identification of Genetically Intact HIV-1 Proviruses in Specific CD4(+) T Cells from Effectively Treated Participants. Cell Rep. 2017;21(3):813-22. Epub 2017/10/19. doi: 10.1016/j.celrep.2017.09.081. PubMed PMID: 29045846; PMCID: PMC5960642.
47. Van Zyl GU, Katusiime MG, Wiegand A, McManus WR, Bale MJ, Halvas EK, Luke B, Boltz VF, Spindler J, Laughton B, Engelbrecht S, Coffin JM, Cotton MF, Shao W, Mellors JW, Kearney MF. No evidence of HIV replication in children on antiretroviral therapy. J Clin Invest. 2017;127(10):3827-34. Epub 2017/09/12. doi: 10.1172/JCI94582. PubMed PMID: 28891813; PMCID: PMC5617669.
48. Goswami R, Nelson AN, Tu JJ, Dennis M, Feng L, Kumar A, Mangold J, Mangan RJ, Mattingly C, Curtis AD, 2nd, Obregon-Perko V, Mavigner M, Pollara J, Shaw GM, Bar KJ, Chahroudi A, De Paris K, Chan C, Van Rompay KKA, Permar SR. Analytical Treatment Interruption after Short-Term Antiretroviral Therapy in a Postnatally Simian-Human Immunodeficiency Virus-Infected Infant Rhesus Macaque Model. mBio. 2019;10(5). Epub 2019/09/07. doi: 10.1128/mBio.01971-19. PubMed PMID: 31488511; PMCID: PMC6945967.
49. Abrahams MR, Joseph SB, Garrett N, Tyers L, Moeser M, Archin N, Council OD, Matten D, Zhou S, Doolabh D, Anthony C, Goonetilleke N, Karim SA, Margolis DM, Pond SK, Williamson C, Swanstrom R. The replication-competent HIV-1 latent reservoir is primarily established near the time of therapy initiation. Sci Transl Med. 2019;11(513). Epub 2019/10/11. doi: 10.1126/scitranslmed.aaw5589. PubMed PMID: 31597754.
50. Siliciano JD, Siliciano RF. Enhanced culture assay for detection and quantitation of latently infected, resting CD4+ T-cells carrying replication-competent virus in HIV-1-infected individuals. Methods Mol Biol. 2005;304:3-15. Epub 2005/08/03. doi: 10.1385/1-59259-907-9:003. PubMed PMID: 16061962.