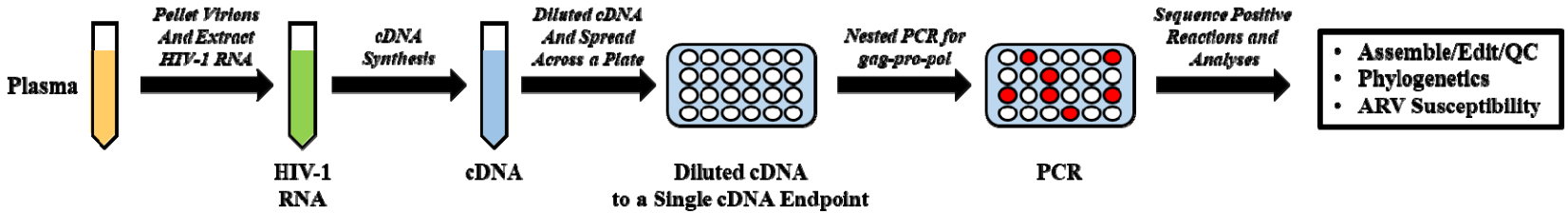
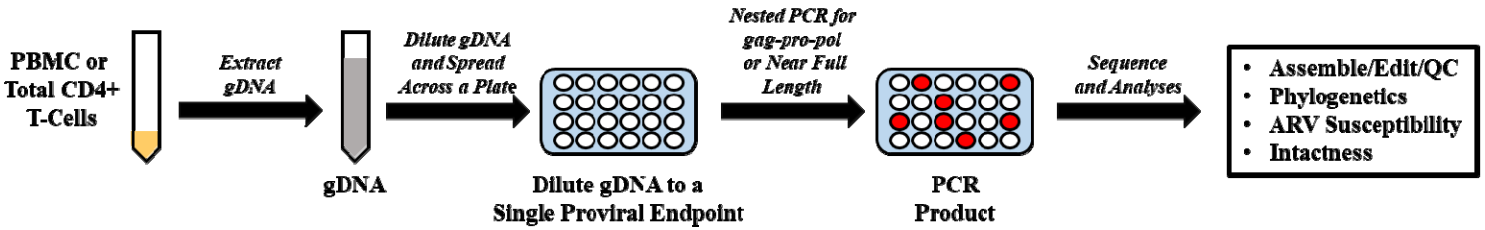
Supplementary Materials

Approaches to Evaluating Non-Suppressible Viremia

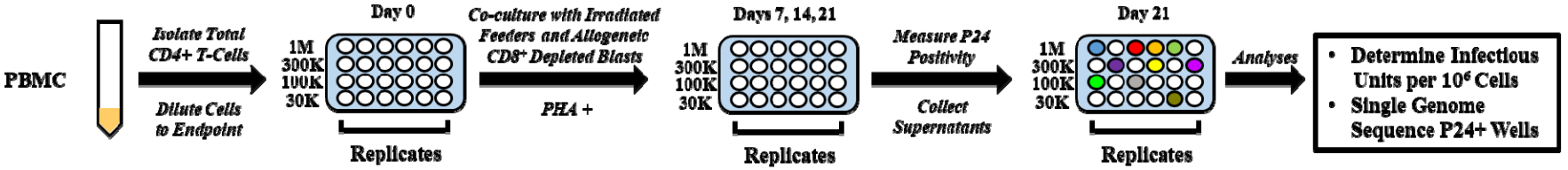
A. **Single Genome Sequencing of HIV-1 Viral RNA (*gag-pro-pol*) from Plasma**



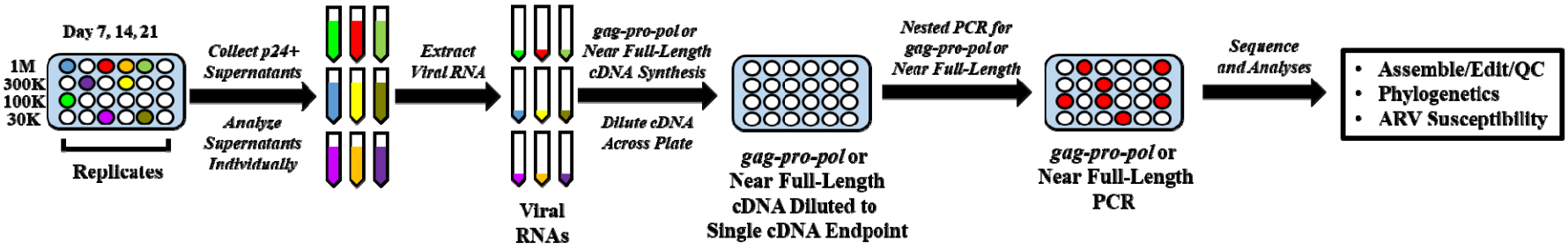
B. **Single Genome Sequencing of gDNA from HIV-1 Infected T-Cells**

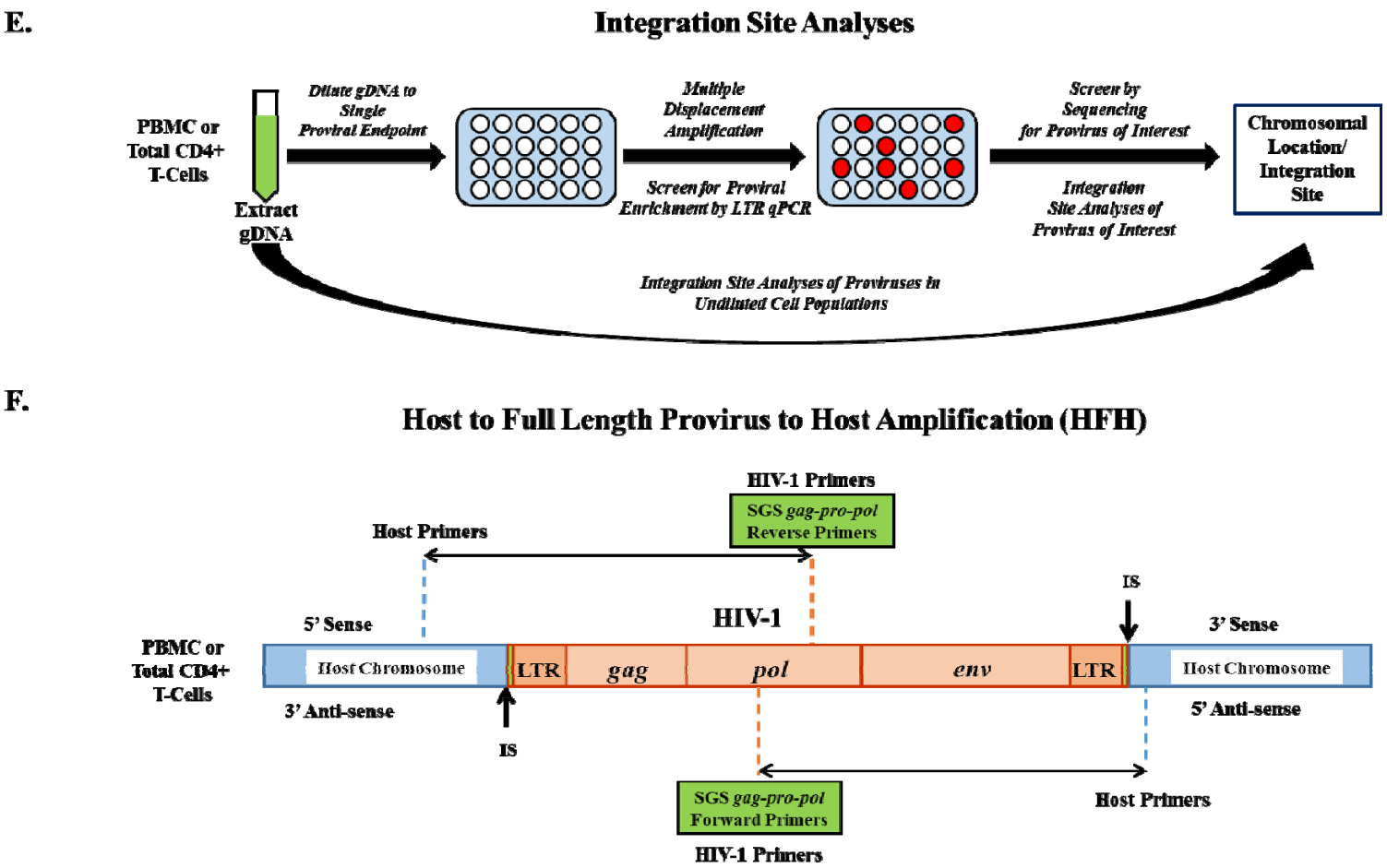


C. Quantitative Viral Outgrowth Assay

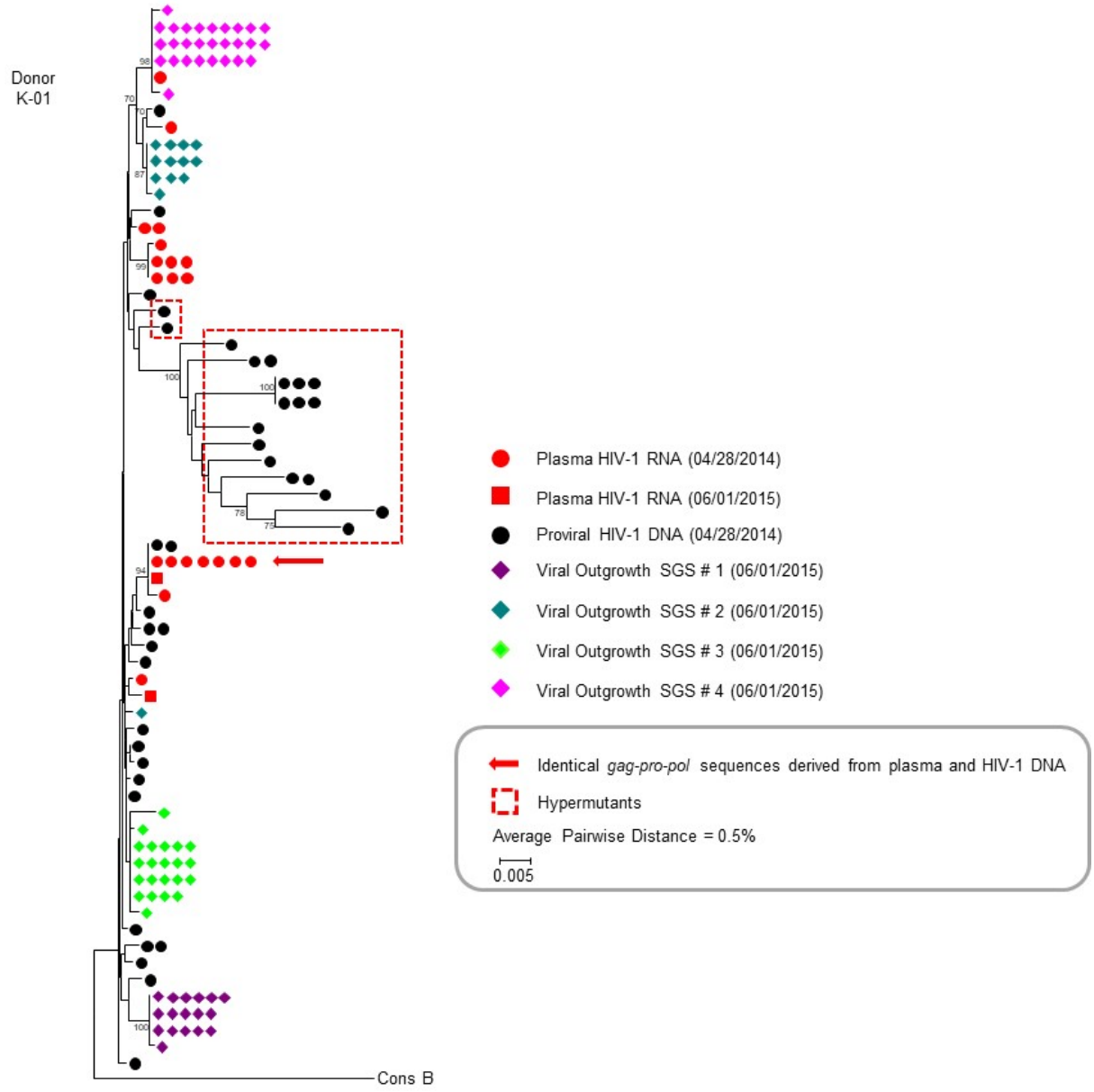


D. Single Genome or Population Sequencing of p24+ Quantitative Viral Outgrowth Assay Wells



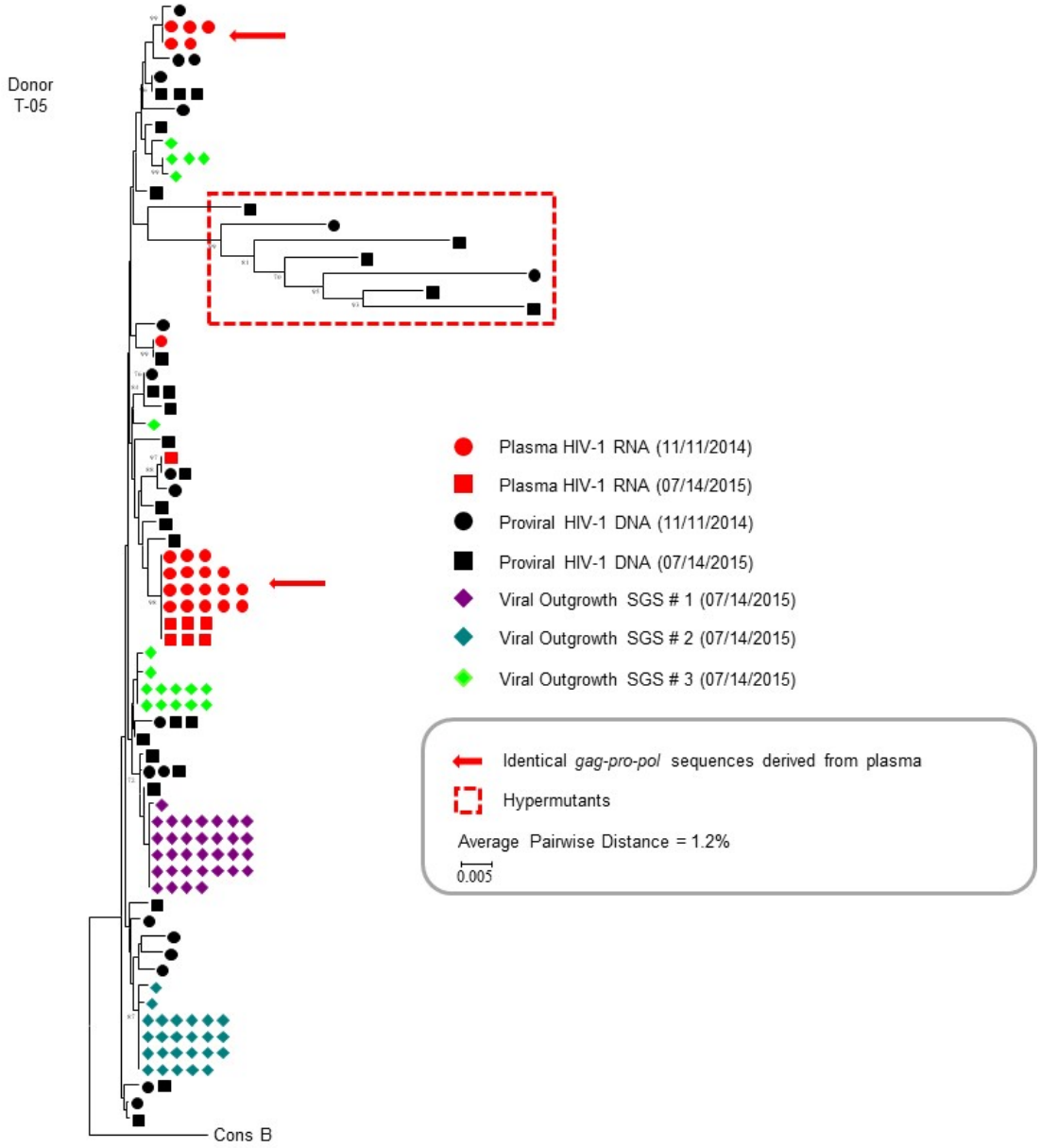


Supplementary Figure 1. Schematic of experimental approaches to evaluate non-suppressible viremia. Donor plasma and PBMC or total $CD4^+$ T-cells originated from large volume blood draws or leukaphereses. (**A**) Single genome sequencing (SGS) of HIV-1 RNA from plasma generated an amplicon containing a portion of *gag (p6)*, all of *pro*, and, from *pol* the portion encoding the 1st 300 amino acids of reverse transcriptase (*gag-pro-pol*) (*35*). (**B**) Proviral HIV-1 DNA SGS of *gag-pro-pol* or near full-length amplicons from PBMC or total $CD4^+$ T-cells (*8, 35*). (**C**) Quantitative viral outgrowth assays performed using total $CD4^+$ T- cells (*23, 24*). (**D**) SGS of *gag-pro-pol* or near full-length HIV-1 RNA genome from quantitative viral outgrowth assay $p24^+$ wells (*8, 36*). (**E**) Integration site analyses (ISA) were preformed directly on undiluted cell populations or at a single proviral endpoint using multiple displacement amplification of DNA originating from PBMC or total $CD4^+$ T-cells (*34, 37*). (**F**) For proviruses of interest, host to full-length provirus to host amplification and sequencing (HFH) was used to confirm the sequence identity of the clonally expanded provirus.

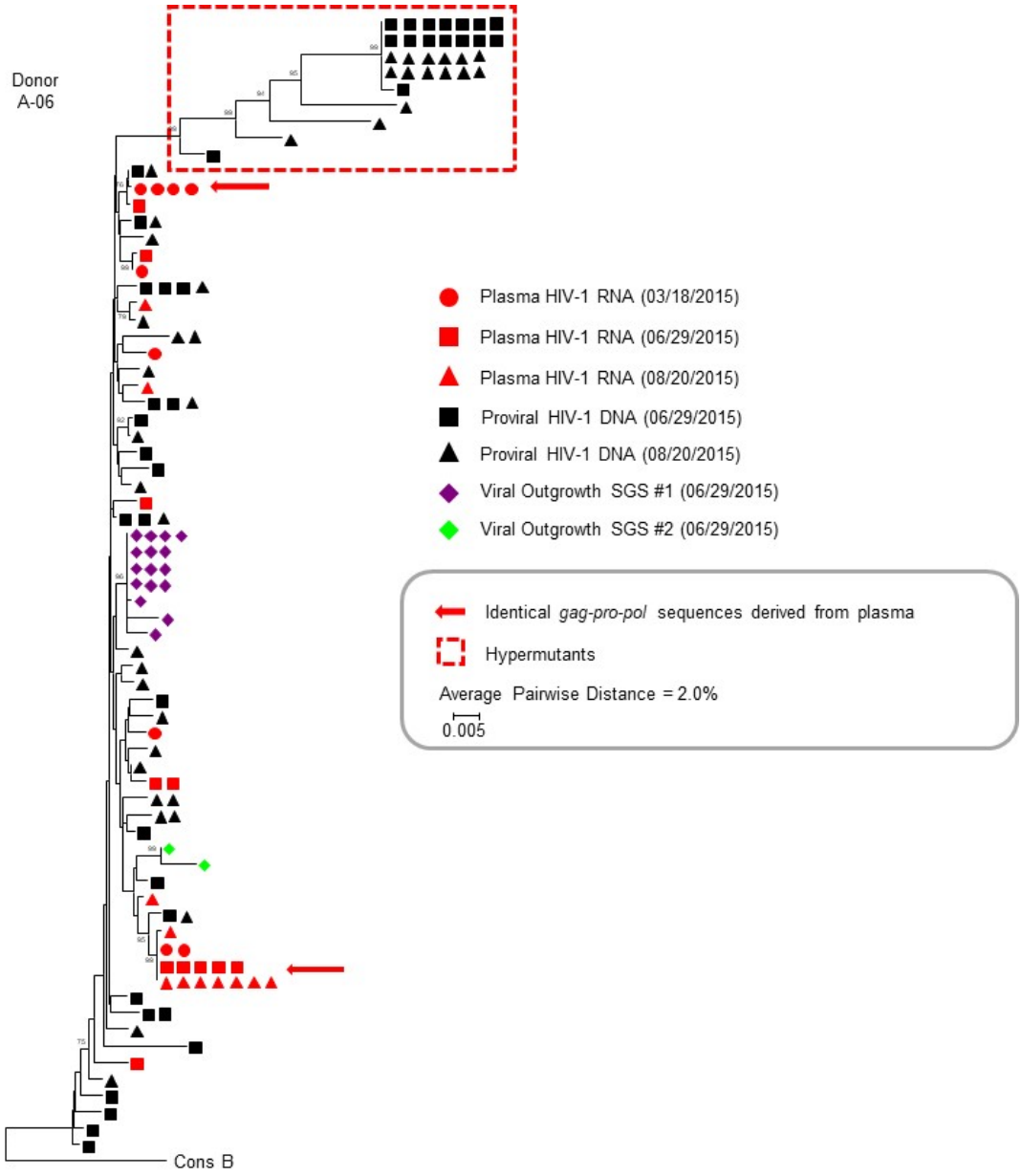


Supplemental Figure 2. Representative neighbor-joining *p*-distance phylogenetic trees of plasma HIV-1 RNA-, HIV-1 DNA-, and quantitative viral outgrowth assay (qVOA)-derived sequences for donor K-01 with non-suppressible viremia. The tree was rooted to a subtype B consensus sequence. Single genome sequences (SGS) of a portion of *gag (p6)*, all of *pro*, and the portion of *pol* encoding the 1st 300 amino acids of reverse transcriptase (*gag-pro-pol*) were obtained from plasma HIV-1 RNA, HIV-1 DNA from peripheral blood mononuclear cells (PBMC), and culture supernatants from p24^{+} qVOA wells for donor K-01 (*35*). Red circles and squares represent plasma-derived sequences from two different time points. Black circles represent HIV-1 DNA-derived sequences from one time point. Different colored diamonds represent viral outgrowth assay-derived sequences from independent p24^{+} wells. A red arrow shows identical *gag-pro-pol* sequences for plasma HIV-1 RNA- and HIV-1 DNA-derived sequences for donor K-01. HIV-1 DNA sequences with G to A hypermutations are enclosed in red-hashed boxes. The viral outgrowth sequence variants that differ by 1-2 nucleotides can be attributed to either ex vivo replication or errors introduced during cDNA synthesis. Average pairwise distances (APD) calculated by MEGA v6.0 using HIV-1 DNA sequences and excluding hyper-mutated sequences.

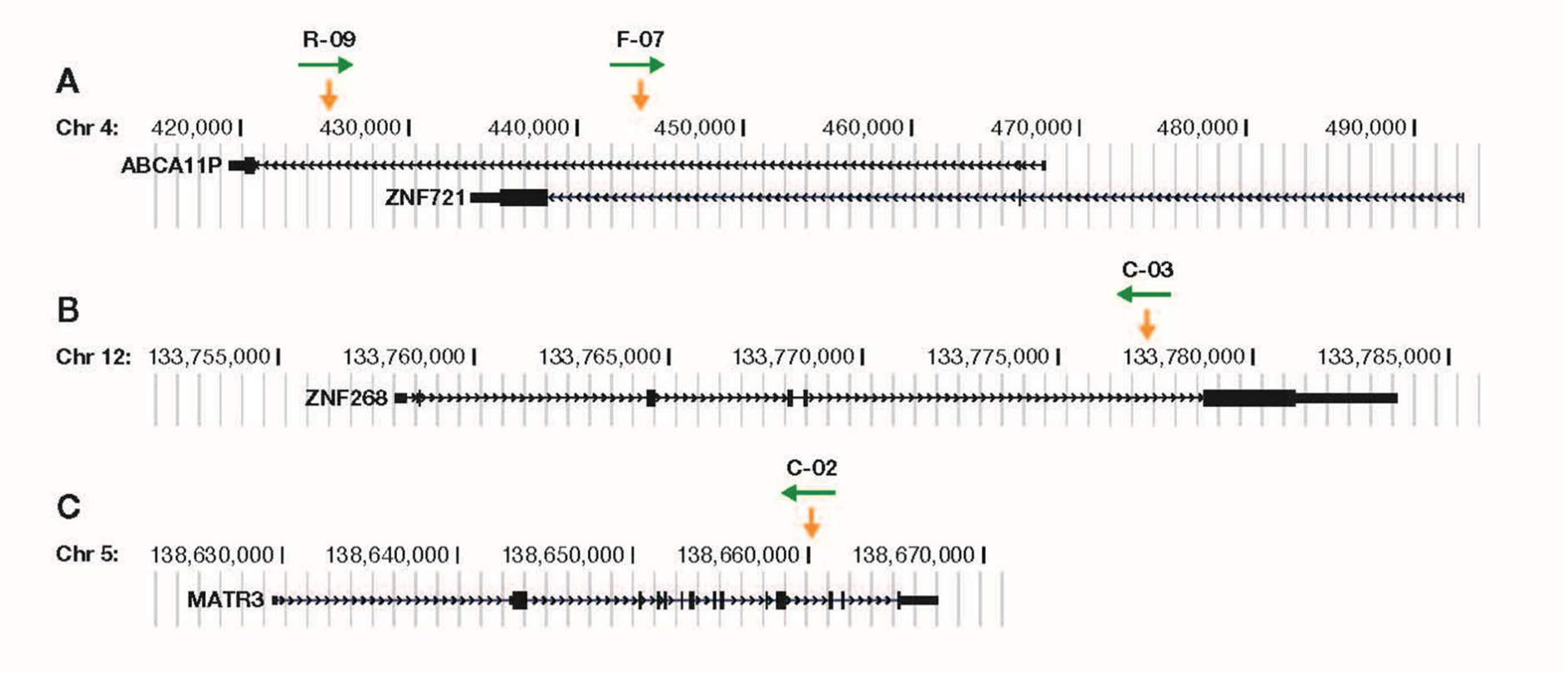
Supplemental Figure 3. Representative neighbor-joining *p*-distance phylogenetic trees of plasma HIV-1 RNA-, HIV-1 DNA-, and quantitative viral outgrowth assay (qVOA)-derived sequences for donor P-08 with non-suppressible viremia. The tree was rooted to a subtype B consensus sequence. Single genome sequences (SGS) of a portion of *gag (p6)*, all of *pro*, and the portion of *pol* encoding the 1st 300 amino acids of reverse transcriptase (*gag-pro-pol*) (*35*) were obtained from plasma HIV-1 RNA, HIV-1 DNA from peripheral blood mononuclear cells (PBMC), and culture supernatants from $p24^+$ qVOA wells for donor P-08. Red circles and squares represent plasma-derived sequences from two different time points. Black circles represent HIV-1 DNA-derived sequences from one time point. Different colored diamonds represent viral outgrowth assay-derived sequences from independent $p24^+$ wells. A red arrow shows identical *gag-pro-pol* sequences for plasma HIV-1 RNA- and HIV-1 DNA-derived sequences for donor P-08. HIV-1 DNA sequences with G to A hypermutations are enclosed in red-hashed boxes. The viral outgrowth sequence variants that differ by 1-2 nucleotides can be attributed to either ex vivo replication or errors introduced during cDNA synthesis. Average pairwise distances (APD) calculated by MEGA v6.0 using HIV-1 DNA sequences and excluding hyper-mutated sequences.



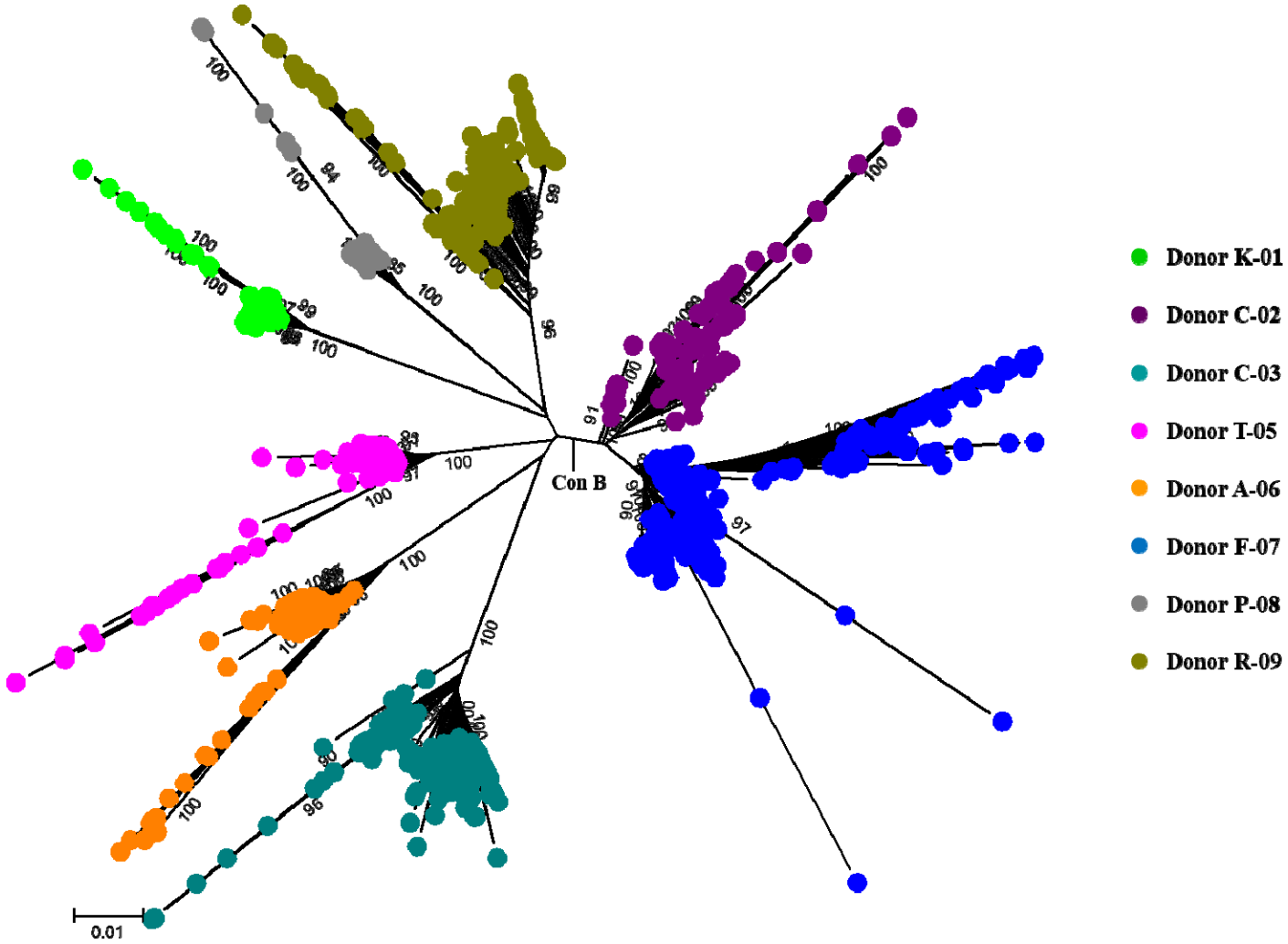
Supplemental Figure 4. Representative neighbor-joining *p*-distance phylogenetic trees of plasma HIV-1 RNA-, HIV-1 DNA-, and quantitative viral outgrowth assay (qVOA)-derived sequences for donor T-05 with non-suppressible viremia. The tree was rooted to a subtype B consensus sequence. Single genome sequences (SGS) of a portion of *gag (p6)*, all of *pro*, and the portion of *pol* encoding the 1st 300 amino acids of reverse transcriptase (*gag-pro-pol*) (*35*) were obtained from plasma HIV-1 RNA, HIV-1 DNA from peripheral blood mononuclear cells (PBMC), and culture supernatants from $p24^+$ qVOA wells for donor T-05. Red circles and squares represent plasma-derived sequences from two different time points. Black circles and squares represent HIV-1 DNA-derived sequences from two different time points. Different colored diamonds represent viral outgrowth assay-derived sequences from independent $p24^+$ wells. A red arrow shows identical *gag-pro-pol* sequences for plasma HIV-1 RNA-derived sequences for donor T-05. HIV-1 DNA sequences with G to A hypermutations are enclosed in red-hashed boxes. The viral outgrowth sequence variants that differ by 1-2 nucleotides can be attributed to either ex vivo replication or errors introduced during cDNA synthesis. Average pairwise distances (APD) calculated by MEGA v6.0 using HIV-1 DNA sequences and excluding hyper-mutated sequences.



Supplementary Figure 5. Representative neighbor-joining *p*-distance phylogenetic trees of plasma HIV-1 RNA-, HIV-1 DNA-, and quantitative viral outgrowth assay (qVOA)-derived sequences for donor A-06 with non-suppressible viremia. The tree was rooted to a subtype B consensus sequence. Single genome sequences (SGS) of a portion of *gag (p6)*, all of *pro*, and the portion of *pol* encoding the 1st 300 amino acids of reverse transcriptase (*gag-pro-pol*) (*35*) were obtained from plasma HIV-1 RNA, HIV-1 DNA from peripheral blood mononuclear cells (PBMC), and culture supernatants from $p24^+$ qVOA wells for donor A-06. Red circles, squares, and triangles represent plasma-derived sequences from three different time points. Black squares and triangles represent HIV-1 DNA-derived sequences from two different time points. Different colored diamonds represent viral outgrowth assay-derived sequences from independent $p24^+$ wells. A red arrow shows identical *gag-pro-pol* sequences for plasma HIV-1 RNA-derived sequences for donor A-06. HIV-1 DNA sequences with G to A hypermutations are enclosed in red-hashed boxes. The viral outgrowth sequence variants that differ by 1-2 nucleotides can be attributed to either ex vivo replication or errors introduced during cDNA synthesis. Average pairwise distances (APD) calculated by MEGA v6.0 using HIV-1 DNA sequences and excluding hyper-mutated sequences.



Supplementary Figure 6. Gene diagrams with integrated, intact, infectious proviruses. The gene diagrams are based on the hg19 reference of the human genome and were prepared using the UCSC genome browser (https://genome.ucsc.edu/). There are, for each of the four genes shown in the diagram, additional transcripts that are not shown. The small black arrowheads along the black lines in the gene diagrams denote the direction in which the genes are transcribed. Exons are shown as bars within the gene. The narrow portions of the exons at the ends of the genes are non-translated regions; the wider exons/portions of exons are coding. The yellow vertical arrows mark the sites where the proviruses were integrated; the horizontal green arrows show the direction in which the proviruses were oriented. (**A**) The organization of the overlapping genes *ABCA11P* and *ZNF721* and the sites of integration for the repliclone proviruses from donors R-09 and F-07. (**B**) The organization of the *ZNF268* gene and integration site of the repliclone provirus from donor C-03. (**C**) The organization of the *MATR3* gene and integration site of the repliclone provirus from donor C-02.



Supplementary Figure 7. Neighbor-joining *p*-distance radial tree of all sequences from all donors. The tree was rooted to a subtype B consensus sequence. Single genome sequences (SGS) of a portion of *gag (p6)*, all of *pro*, and the portion of *pol* encoding the 1st 300 amino acids of reverse transcriptase (*gag-pro-pol*) (*35*) were obtained from plasma HIV-1 RNA, HIV-1 DNA in peripheral blood mononuclear cells (PBMC), and culture supernatants from $p24^+$ qVOA wells. All sequences from each donor are represented as different colored circles; Donor K-01 (Green), C-02 (Purple), C-03 (Teal), T-05 (Pink), A-06 (Orange), F-07 (Blue), P-08 (Grey), and R-09 (Gold).

Supplementary Table 1. HIV drug susceptibilities, diversity, and coreceptor tropisms

Donor ID	Mutations Causing Resistance to Donor's Current ART Regimen[a]	Average Pairwise Distance of All Proviruses(%)[b]	Geno2pheno Coreceptor Tropisms[c,d]
R-09	0/30	2.4	R5-Tropic (13.2%FPR)
C-03	0/49, 0/13*	1.9	R5-Tropic (62.8%FPR)
C-02	0/47	2.3	R5-Tropic (48.7%FPR)
F-07	0/39[†]	2.3	R5-Tropic (33%FPR)
K-01	0/24	0.5	-
P-08	0/21	0.6	-
T-05	0/34	1.2	-
A-06	0/31	2.0	-

[a] Drug-resistance mutations in plasma were identified using the Stanford HIV Drug Resistance Database v8.7; genotype covered protease, reverse transcriptase and integrase in individuals receiving an integrase inhibitor* (*38*)

[b] Average pairwise distance of all proviral sequences calculated using MEGA v6.0 (*57*)

[c] Coreceptor tropisms predicted by Geno2pheno for X4/R5 tropism using near full-length sequences (*41*)

[d] R5, CCR5-tropic; X4, CXCR4-tropic; FPR, false-positive rate for incorrectly identifying the sequence as X4-tropic was determined using cutoffs either from the German (5% to 15% cutoff), European (10% cutoff) and MOTIVATE (2% to 5.75% cutoff) guidelines. False positive rate (FPR) results represented are using European Guideline cutoffs of which were identical to those obtained by German and MOTIVATE guidelines

[†] 21 of 39 plasma-derived sequences contained D67N/K70R/K219Q conferring low-level resistance by the Stanford HIVdb algorithm to Abacavir (*38*)

Supplementary Table 2. Antiretroviral drug levels in individuals with non-suppressible viremia

Donor	Sample Type	Sample Draw Date	Patient Regimen[a]	DRV (ng/mL)	RTV (ng/mL)	ATV (ng/mL)	EFV (ng/mL)	DTG (ng/mL)
R-09	Plasma	6/4/2015	TDF/FTC/**EFV**	-	-	-	1636	-
R-09	Plasma	8/6/2015	TDF/FTC/**EFV**	-	-	-	1334	-
C-03	Plasma	9/30/2014	**DRV/r**/ETV/**DLG**	4522	441	-	-	2033
C-02	Plasma	9/24/2014	TDF/FTC/**EFV**	-	-	-	3321	-
F-07	Plasma	6/3/2015	ABC/3TC/**EFV**	-	-	-	6766	-
K-01	Plasma	4/28/2014	TDF/FTC/**DRV/r**	2277	115	-	-	-
P-08	Plasma	9/29/2015	TDF/FTC/**EFV**	-	-	-	975	-
T-05	Plasma	11/11/2014	TDF/FTC/**ATV/r**	-	36.5	655	-	-
A-06	Plasma	1/29/2015	TDF/FTC/**ATV/r**	-	177	1636	-	-
		Target Trough Range ng/mL		1000-8000	<50-2500	150-850	1000-4000	800-1000

[a] Antiretroviral drug ABC (abacavir); ATV/r (atazanavir/ritonavir; DRV/c (darunavir/cobicistat); DRV/r (darunavir/ritonavir; DTG (dolutegravir); EFV (efavirenz); ETV (etravirine); 3TC (lamivudine); FTC (emtricitabine); TDF (tenofovir disoproxil fumarate)

Antiretroviral drugs tested are shown in **Bold** (*39, 40*)

DRV, EFV assay range (200-15,000 ng/mL); ATV, RTV assay range (10-4,000 ng/mL); DTG assay range (20-10,000 ng/mL)

Supplementary Table 3. Immunophenotyping of PBMC from donors referred for non-suppressible viremia

	Frequency (%) of Lymphocytes						**Cellular Activation Markers**				
							Frequency (%) of $CD4^+$ T-Cells				**Frequency (%) of $CD8^+$ T-Cells**
Donor ID	***$CD3^+$ $CD4^+$***	***$CD3^+$ $CD8^+$***	***$CD3^+$ $CD4^+$ $CD8^+$***	***$CD3^-$ $CD19^+$***	***$CD3^-$ $CD56^+$***		***$CD25^+$***	***$CD69^+$***	***HLA-DR^+***	***$CD38^+$ HLA-DR^+***	***$CD107a^+$***
R-09	31.8	12.5	0.87	27.8	4.7		9.8	12.4	22.2	7.9	2.6
C-03	31.7	31.7	0.72	15.8	2.1		9.3	6.4	9.0	4.2	0.7
C-02	33.5	26.4	0.42	6.5	8.5		8.7	19.6	16.4	8.0	3.9
F-07	26.1	33.0	1.08	8.1	1.9		5.2	3.2	10.3	5.3	0.8
K-01	30.9	29.3	1.05	10.4	5.2		13.1	3.2	18.4	5.4	0.4
P-08	17.0	37.7	0.75	9.4	11.2		6.5	9.4	31.7	8.7	2.6
T-05	48.5	22.9	1.26	5.9	4.1		8.7	1.5	5.9	3.0	1.0
A-06	32.0	32.0	0.69	5.9	1.1		10.5	4.1	19.7	4.1	1.5
Median	31.8	30.5	0.75	8.8	4.4		9.0	5.3	17.4	5.4	1.3
Healthy Donor	32.5-68.3[a]	11.5-38.6[a]	0.25-6[b]	4.7-22.5[a]	6.3-10[c]		0.3-10.7[a]	1-9[d]	0.8-4.4[a]	0.3-1.35[a]	2-20[e]

[a] Range (%) of lymphocyte subsets $CD3^+/CD4^+$, $CD3^+/CD8^+$, $CD3^-/CD19^+$, $CD4^+/CD25^+$, $CD4^+$/HLA-DR^+, and $CD4^+/CD38^+$/HLA-DR^+ in healthy human adults (*25, 26*)

[b] Range (%) of lymphocyte subset $CD3^+/CD4^+/CD8^+$ in healthy human adults (*27, 28*)

[c] Range (%) of lymphocyte subset $CD3^-/CD56^+$ in healthy human adults (*29, 30*)

[d] Range (%) of lymphocyte subset $CD4^+/CD69^+$ in healthy human adults (*31*)

[e] Range (%) of lymphocyte subset $CD8^+/CD107a^+$ in healthy human adults (*32*)

Supplementary Table 4. Host to full-length provirus to host amplification primer sets

Donor (Provirus)	HIV-Specific Primers	HIV-Specific Primer Sequences (5'-3')	Host-Specific Primer	Host-Specific Primer Sequences (5'-3')
R-09 (ABCA1P)	R-09_F1	GATGACAGCATGCCAGGGAG	ABCA11P_R1	TGGGATTACAGGCTGGGATAATG
	R-09_F2	GAGTCTTAGCTGAAGCAATGAG	ABCA11P_R2	GTATAACGTAAAATGAATACATCCTTGTC
	R-09_R1	CTATTAAGTCTTTTGATGGGTCATAG	ABCA11P_F1	TGGTTGTTCCCTATACATTTTAATC
	R-09_R2	CTGTTAGTGGTATTACTTCTGTTAGTGCTT	ABCA11P_F2	ATTCTCAGTGTAGAGCGTGGTTACC
C-03 (ZNF268)	C-03_F1	GATGACCGCATGTCAGGGAG	ZNF268_F1	CACAAAGCTGTTTGCCTACCC
	C-03_F2	GAGTCTTGGCTGAAGCAATGAG	ZNF268_F2	TTCTTTTCCATGCCTGCTAGAG
	C-03_R1	CTATTAAATCTTTTGATGGGTCATAA	ZNF268_R1	GCAGAGAACAATGCAGATTACT
	C-03_R2	CTGTCAGTGGTACTACATCTGTTAGTGCTT	ZNF268_R2	CAGGATAAAAATTGCACAGCAGGC
C-02 (MATR3)	C-02_F1	GATGACCGCATGTCAGGGAG	MATR3_F1	CCCAACATAGTTAAAACTTTTGCCACTCATTC
	C-02_F2	GAGTCTTGGCTGAAGCAATGAG	MATR3_F2	ATTTGAACAAGTAAGTCATTTAGAAGCC
	C-02_R1	CTATTAAATCTTTTGATGGGTCATAA	MATR3_R1	GGAAACGGATAGCGTCTTTG
	C-02_R2	CTGTCAGTGGTACTACATCTGTTAGTGCTT	MATR3_R2	GGAAGGTCTGCCTCACACAAAG
F-07 (ZNF721/ ABCA11P)	F-07_F1	GATGACAGCATGTCAGGGAG	ZNF721_R1	CTCTTAAAGTCTCTTGCCTATATTCAAATTG
	F-07_F2	GGGTTTTGGCGGAAGCAATGAG	ZNF721_R2	CTGCCTGTGCTTTTGAGGTCTTAAG
	F-07_R1	CTATTAAGTCTTTTGATGGGTCATAA	ZNF721_F1	GGTGCTAGGAAAATTATCTACAAG
	F-07_R2	TTGTTAGTGGTACTACTTCTGTTAGTGCCT	ZNF721_F2	GAAACCCAAATAAAGCTATTTGAAGTAAACATAC

Supplementary Table 5. Sequencing primers used for near full-length proviral sequencing

Sequencing Primer Name	Primer Sequence	Sequencing Primer Name	Primer Sequence
A	5'-CTTTCGCTTTCAAGTCCCT-3'	**M**	5'-AAGAGATATAGCACACAAGTAG-3'
B	5'-AAATCTCTAGCAGTGGCG-3'	**M_R-09**	5'-AGGAACTATAGCACACAAGTAG-3'
SGS_C	5'-TTCTTCTGTCAATGGCCATTGTTTAAC-3'	**N**	5'-ATTGGGTGCCAACATAGCAGAATA-3'
SGS_D	5'-TTGCCCAATTCAATTTTCCCACTAA-3'	**O**	5'-CTATGGCAGGAAGAAGCGG-3'
C	5'-CAAAGGATAGAGGTAAAAGAC-3'	**P**	5'-GTGGAAAAATAACATGGTAG-3'
C_R-09	5'-CAAAGGATAGATGTAAAAGAT-3'	**Q**	5'-CAGTAGTATCAACTCAACTG-3'
C_R-09_RC	5'-ATCTTTTACATCTATCCTTTG-3'	**Q_R-09**	5'-CAGTGGTGTCAACTCAATTG-3'
C_F-07_RC	5'-GTCTTTTACCTCTATCTTTTG-3'	**R**	5'-TGCACAGTTTTAATTGTG-3'
D	5'-AGTGACATAGCAGGAACTACTAG-3'	**S**	5'-CAATGTATGCCCCTCCCATC-3'
E	5'-GAAAAAGGGCTGTTGGAAATG-3'	**S_R-09**	5'-CAATATATGCCCCTCCTATC-3'
F	5'-GGAATTGGAGGTTTTATCAAAG-3'	**T**	5'-GCAGGAAGCACTATGGGC-3'
G	5'-GTAACAGTACTAGATGTGGGTG-3'	**U**	5'-ATAGAGTTAGGCAGGGATACTC-3'
H	5'-GCTGGACTGTCAATGACATAC-3'	**V**	5'-GAGCCTGTGCCTCTTCAGC-3'
I	5'-CCACAGAAAGCATAGTAATAT-3'	**W**	5'-GTGGCAAGTGGTCAAAAAGTAG-3'
J	5'-CACACAAAGGGATTGGAGGAAATGA-3'	**W_R-09**	5'-GTGGCAAGTGGTCAAAACCCAG-3'
K	5'-AATTAGCAGGAAGATGGC-3'	**X**	5'-GGGACTGGAAGGGCTAATTTAC-3'
K_RC	5'-GCCATCTTCCTGCTAATT-3'	**Y_R-09**	5'-GCTGCTCTTTGCTTGTACTGGG-3'
L	5'-CAGCAGTACAAATGGCAGTATTC-3'		

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **Abstract, Methods, Page 2:** The study was small (N = 8), cross-sectional, and referral-based from clinical practice. Donors were referred from local HIV care providers to evaluate viremia that was not suppressible with antiretroviral therapy for more than 6 months. One or more samples were collected from donors. Single templates of HIV-1 RNA obtained from plasma and viral outgrowth of cultured cells and from proviral DNA were PCR-amplified and sequenced for evidence of clones of cells that produced infectious viruses. Clones were confirmed by host-proviral integration site analysis.
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **Abstract, Results, Page 2:** HIV-1 genomic RNA with identical sequences were identified in plasma samples from all eight donors. The identical viral RNA sequences did not change over time and did not evolve resistance to the ART regimen. In four of the donors, viral RNA sequences obtained from plasma matched those sequences from viral outgrowth cultures, indicating that the viruses were replication-competent. Integration sites for infectious proviruses from those four donors were mapped to the introns of the *MATR3*, *ZNF268*, *ZNF721/ABCA11P*, and *ABCA11P* genes. The sizes of the clones were estimated to be from 50 million to 350 million cells.
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **Introduction, Paragraph 1, Page 4:** Despite clinically-effective ART, infected cells persist for the life of the individual and a small subset of the infected cells carry intact proviruses capable of producing infectious virus that can fuel viral rebound when ART is stopped (*2-6*). Recent work has shown that this reservoir of HIV-1 is sustained by the proliferation of clones of cells (*7-11*). **Introduction, Paragraph 2, Page 4:** The current study arose from requests by local HIV-1 care providers to investigate the cause of persistent viremia (detectable >6 months at >20 copies/mL) above the limit of detection of FDA-approved HIV-1 RNA assays in individuals who reported consistent adherence to their medication.
Objectives	3	State specific objectives, including any prespecified hypotheses **Introduction, Paragraph 1, Page 4:** We previously reported that one donor with advanced malignancy had clinically detected levels of virus in blood (>40 copies/mL) on ART that was comprised of a mixture of drug-resistant virus and wild-type virus, the latter produced by a large clone of cells carrying an intact, infectious provirus (*10*). Similar instances have not been published.
Methods		
Study design	4	Present key elements of study design early in the paper **Methods, Paragraph 1, Page 14:** *Study design.* The primary objective of the study was to evaluate non-suppressible viremia and confirm its clonal cellular origin. The

		schematic in Supplementary Figure 1 provides an overview of the multistep methodological approach to evaluating non-suppressible viremia and confirming its clonal cellular origin. This included single genome sequencing (SGS) of amplicons containing HIV-1 gag-pro-pol and/or near full-length genomes derived from i) plasma, ii) PBMCs or total $CD4^+$ T-cells, and/or iii) $p24^+$ qVOA wells in all eight participants.
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **Methods, Paragraph 2, Page 15:** *Participants and sample collection.* Study participants were referred from the University of Pittsburgh Medical Center for HIV-AIDS Program or from the Allegheny Health Network Positive Health Clinic and enrolled into the study at the University of Pittsburgh Clinical Trials Unit between April of 2014 and June of 2015. **Figures 1-4 and Supplementary Figure 2-5:** Specified draw dates of samples. **Methods, Paragraph 2, Page 15:** Follow up of study participants after evaluation for non-suppressible viremia included continuation of primary care with immunologic and virologic monitoring.
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants **Methods, Paragraph 2, Page 15:** Inclusion criteria for the study were: 1) initial suppression of plasma HIV-1 RNA to below the limit of detection of commercial assays (<20 or <40 copies/mL) on a DHHS-recommended ART regimen; 2) followed by clinically-detectable plasma viremia (HIV-1 RNA >20 copies/mL) for > 6 months, as measured by the COBAS Ampliprep/COBAS Taqman, v2.0 assay (TMv2.0) (Roche) or the m2000sp/RealTime HIV-1 Assay (Abbott); and 3) assessment by the referring physician that the patient was fully compliant with their ART regimen. Switches in ART regimen or intensification with another antiretroviral did not exclude participants from the study.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **Methods, Paragraph 2, Page 17:** *Assessment of cell clonality.* A clone of HIV-1 infected cells was defined by identifying multiple cells that had identical proviral sequences integrated into the identical position in the human genome. **Methods, Paragraph 4, Page 17:** *Integration site analyses.* Integration site analyses (ISA) were conducted as previously described (*34*). **Methods, Paragraph 3, Page 18:** *Sequence alignments, quality control, and phylogenetic analyses.* Sequence alignments and phylogenetic analyses of the *gag-pro-pol* sequences were performed as reported previously (*8*). This included alignments, exclusion of mixtures, and quality control of sequences in Sequencher v5.0 (Gene Codes) (Supplementary Figure 7). **Methods, Paragraph 2 Page 19:** *HIV-1 subtype, drug susceptibility analyses, drug concentration determination and coreceptor tropism.* HIV-1 subtype and genotype-predicted susceptibilities to antiretroviral drugs were determined by the HIVdb program (Stanford University HIV Drug Resistance Database) (*38*). Drug level concentrations in human plasma for the non-nucleoside reverse transcriptase inhibitor

		efavirenz and the protease inhibitors darunavir, atazanavir, and ritonavir, were measured by a gradient separation through Ultra Performance Liquid Chromatography with Electrospray Ionization Tandem Mass Spectrometry for detection (*39*). Drug level concentration in human plasma for the integrase inhibitor dolutegravir was measured using a protein precipitation method with isocratic separation by Liquid Chromatography with tandem mass spectrometry for detection (*40*).
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **Methods, Paragraph 3, Page 15 (Sources of Data):** Longitudinal samples were collected at two or more time points as large-volume phlebotomy (100-180mL) or plasmapheresis and leukapheresis between April of 2014 and April of 2019. Plasma, peripheral blood mononuclear cells (PBMC), and total $CD4^+$ T-cells were isolated and stored as previously reported from all donors (*22*). **Methods, Paragraph 2, Page 16:** Quantification of plasma HIV-1 RNA and cellular HIV-1 DNA and RNA. HIV-1 DNA and RNA in PBMC or total $CD4^+$ T-cells was quantified by qPCR targeting the 3' end of integrase as previously reported (*22*). **Methods, Paragraph 3, Page 16:** *Quantitative viral outgrowth assay.* Quantitative viral outgrowth assays (qVOA) were performed using total $CD4^+$ T-cells. The frequency of HIV-1 infected cells carrying an inducible infectious provirus (IUPM) was determined by a maximum likelihood method, as reported (*8, 23, 24*). **Methods, Paragraph 4, Page 16:** *Single genome amplification and sequencing (SGS).* SGS of *gag* (*p6*), *pro*, and the first 300 amino acids of *pol* (*gag-pro-pol*) was performed using HIV-1 RNA from plasma or from $p24^+$ qVOA wells or from cellular HIV-1 DNA as follows: i) endpoint dilution of extracted nucleic acid to single HIV-1 template per PCR reaction as determined by Poisson distribution statistics; ii) generation of a ~1.56 kb RT-PCR or PCR amplicon; and, iii) bi-directional sequencing of the amplicon by the Sanger method (*8, 35*). **Methods, Paragraph 5 Page 16:** *Near full-length (NFL) single proviral and viral genome amplification and sequencing.* Near full-length (NFL) HIV-1 DNA was amplified by nested PCR from genomic DNA extracted from PBMC or total $CD4^+$ T-cell at a proviral endpoint of a single template per PCR reaction as determined by Poisson distribution statistics. PCR amplifications were performed using the 2x RANGER DNA Polymerase Mix according to the manufacturer's recommendations (Bioline) and previously reported primers (*56*). The sizes of the NFL amplicons were confirmed with the Perkin Elmer GX Touch 24 LabChip bioanalyzer using the 12K DNA module. NFL amplicons containing 16bp symmetrical barcodes were size selected by BluePippin (Sage Science) and libraries were constructed using the PacBio SMRTbell Template Prep Kit prior to PacBio sequencing (Pacific Biosciences). Alternatively, NFL amplicons were sequenced by Illumina MiSeq using the KAPA HyperPlus kit (KAPA Biosystems) for library construction and the MiSeq nano v2 500 cycle, 2x250run kit with dual index (Illumina) for sequencing. SGS of overlapping half genomes from virion-associated HIV-1 RNA in $p24^+$ qVOA wells was performed as reported (*8, 36*). **Methods, Paragraph 4 Page 17:** *Integration site analyses.* Integration site analyses (ISA) were conducted as previously described (*34*). For integration site analysis of proviruses with specific viral sequences of interest (i.e., matching plasma HIV-1 RNA

		sequences), ISA was performed with the following modifications: the starting template was 0.8× SPRI purified MDA material that contained the *gag-pro-pol* sequence that matched the plasma HIV-1 RNA or $p24^+$ qVOA HIV-1 RNA sequence of interest. ISA was performed using an in-house workflow utilizing multiple displacement amplification and a specificity-enhancing linker-mediated PCR that amplifies across the 5'LTR host/virus junction (*37*). **Methods, Paragraph 2 Page 19:** *HIV-1 subtype, drug susceptibility analyses, drug concentration determination and coreceptor tropism.* HIV-1 subtype and genotype-predicted susceptibilities to antiretroviral drugs were determined by the HIVdb program (Stanford University HIV Drug Resistance Database) (*38*). Drug level concentrations in human plasma for the non-nucleoside reverse transcriptase inhibitor efavirenz and the protease inhibitors darunavir, atazanavir, and ritonavir, were measured by a gradient separation through Ultra Performance Liquid Chromatography with Electrospray Ionization Tandem Mass Spectrometry for detection (*39*). Drug level concentration in human plasma for the integrase inhibitor dolutegravir was measured using a protein precipitation method with isocratic separation by Liquid Chromatography with tandem mass spectrometry for detection (*40*). Co-receptor tropism was determined by the Geno2Pheno bioinformatics software using both established and individually selected cutoffs (*41*). **Methods, Paragraph 3 Page 19:** *Immunophenotyping.* Surface flow cytometric staining of CD3, CD4, CD8, CD19, CD56, CD25, CD69, CD38, HLA-DR, and CD107a (BD Biosciences) was performed on a BD LSRII cytometer according to standard published methods (*58*), analyzed using FlowJo, and results compared to published results of healthy adults (*25-32)*. **Methods, Paragraph 4 Page 19:** *Calculating the size of the repliclones.* The calculations used to estimate the size of each repliclone were based on i) the estimated total number of lymphocytes found in the human body, a value of 2×10^{12} cells *(46)*, ii) the HIV-1 DNA copies/million $CD4^+$ T-cells (*22*), and iii) the fraction of proviruses that matched the relevant repliclone as determined by single genome sequencing. Repliclone copies/10^6 $CD4^+$ cells was determined by multiplying HIV-1 DNA/10^6 $CD4^+$ cells by the frequency of the repliclone. The number of $CD4^+$ cells per donor was determined by multiplying the percent of lymphocytes $CD4^+$ T-cells by 2×10^{12} lymphocytes in the human body *(46)*. The size each repliclone was then determined by multiplying the number of repliclone copies/million $CD4^+$ T-cells by the estimated total body CD4+ T-cells. This calculation does not account for skewed distribution of CD4+ T-cells in non-lymphoid tissues. **Methods, Paragraph 2 Page 20:** *Statistical analyses.* Average pairwise distances were calculated by MEGA6 through the exclusion of hypermutants and identification of hypermutants was performed using the HIVdb algorithm (Stanford University HIV Drug Resistance Database) (38, 57).
Bias	9	Describe any efforts to address potential sources of bias **Methods, Paragraph 2 Page 15:** *Participants and sample collection.* Study participants were referred from the University of Pittsburgh Medical Center HIV-AIDS Program or from the Allegheny Health Network Positive Health Clinic and enrolled into the study at the University of Pittsburgh Clinical Trials Unit between April of 2014 and June of 2015. Inclusion criteria for the study were: 1) initial suppression of plasma HIV-1 RNA to below the limit of detection of commercial

		assays (<20 or <40 copies/mL) on a DHHS-recommended ART regimen; 2) followed by clinically-detectable plasma viremia (HIV-1 RNA >20 copies/mL) for > 6 months, as measured by the COBAS Ampliprep/COBAS Taqman, v2.0 assay (TMv2.0) (Roche) or the m2000sp/RealTime HIV-1 Assay (Abbott); and 3) assessment by the referring physician that the patient was fully compliant with their ART regimen. Switches in ART regimen or intensification with another antiretroviral did not exclude participants from the study. Eligibility was confirmed through source documentation. **Methods, Paragraph 2 Page 18:** *Host to full-length provirus to host amplification and sequencing (HFH).* For specific clones of interest, full-length sequences (host to full-length provirus to host [HFH]) of the proviruses integrated at the same site were assessed for identity. This procedure eliminated bias that could have resulted in identification of the wrong clone that was the origin of non-suppressible viremia. **Methods, Paragraph 2 Page 20:** *Statistical analyses.* Average pairwise distances were calculated by MEGA6 through the exclusion of hypermutants and identification of hypermutants was performed using the HIVdb algorithm (Stanford University HIV Drug Resistance Database) (38, 57). By defining the overall diversity of the donors' sequences, this analysis eliminated potential bias that identical sequences were from chance alone.
Study size	10	Explain how the study size was arrived at **Introduction, Paragraph 2, Page 4:** The current study arose from requests by local HIV-1 care providers to investigate the cause of persistent viremia (detectable >6 months at >20 copies/mL) above the limit of detection of FDA-approved HIV-1 RNA assays in individuals who reported consistent adherence to their medication. The study size was limited by the number of donors who were referred and met eligibility criterion described below. **Methods, Paragraph 2, Page 15:** Inclusion criteria for the study were: 1) initial suppression of plasma HIV-1 RNA to below the limit of detection of commercial assays (<20 or <40 copies/mL) on a DHHS-recommended ART regimen; 2) followed by clinically-detectable plasma viremia (HIV-1 RNA >20 copies/mL) for > 6 months, as measured by the COBAS Ampliprep/COBAS Taqman, v2.0 assay (TMv2.0) (Roche) or the m2000sp/RealTime HIV-1 Assay (Abbott); and 3) assessment by the referring physician that the patient was fully compliant with their ART regimen. Switches in ART regimen or intensification with another antiretroviral did not exclude participants from the study.
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **Methods, Paragraph 2 Page 16:** Quantification of plasma HIV-1 RNA and cellular HIV-1 DNA and RNA. HIV-1 DNA and RNA in PBMC or total $CD4^+$ T-cells was quantified by qPCR targeting the 3' end of integrase as previously reported (*22*). **Methods, Paragraph 3 Page 16:** *Quantitative viral outgrowth assay.* Quantitative viral outgrowth assays (qVOA) were performed using total $CD4^+$ T-cells. The frequency of HIV-1 infected cells carrying an inducible infectious provirus (IUPM) was determined by a maximum likelihood method, as reported (*8, 23, 24*). **Methods, Paragraph 4 Page 17:** *Integration site analyses.* Integration site analyses (ISA) were conducted as previously described (*34*). **Methods, Paragraph 2 Page 19:** *HIV-1 subtype, drug susceptibility analyses, drug*

		concentration determination and coreceptor tropism. HIV-1 subtype and genotype-predicted susceptibilities to antiretroviral drugs were determined by the HIVdb program (Stanford University HIV Drug Resistance Database) (*38*). Drug level concentrations in human plasma for the non-nucleoside reverse transcriptase inhibitor efavirenz and the protease inhibitors darunavir, atazanavir, and ritonavir, were measured by a gradient separation through Ultra Performance Liquid Chromatography with Electrospray Ionization Tandem Mass Spectrometry for detection (*39*). Drug level concentration in human plasma for the integrase inhibitor dolutegravir was measured using a protein precipitation method with isocratic separation by Liquid Chromatography with tandem mass spectrometry for detection (*40*). Co-receptor tropism was determined by the Geno2Pheno bioinformatics software using both established and individually selected cutoffs (*41*). **Methods, Paragraph 4 Page 19:** *Calculating the size of the repliclones.* The calculations used to estimate the size of each repliclone were based on i) the estimated total number of lymphocytes found in the human body, a value of 2×10^{12} cells *(46)*, ii) the HIV-1 DNA copies/million $CD4^+$ T-cells (*22*), and iii) the fraction of proviruses that matched the relevant repliclone as determined by single genome sequencing. Repliclone copies/10^6 $CD4^+$ cells was determined by multiplying HIV-1 DNA/10^6 $CD4^+$ cells by the frequency of the repliclone. The number of $CD4^+$ cells per donor was determined by multiplying the percent of lymphocytes $CD4^+$ T-cells by 2×10^{12} lymphocytes in the human body *(46)*. The size each repliclone was then determined by multiplying the number of repliclone copies/million $CD4^+$ T-cells by the estimated total body CD4+ T-cells. This calculation does not account for skewed distribution of CD4+ T-cells in non-lymphoid tissues.
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding **Methods, Paragraph 1 Page 19:** *Phylogenetic analyses.* Neighbour-joining *p*-distance phylogenetic trees were rooted to subtype B with bootstrapping at 1,000 replicates per tree using MEGA6 (*57*). MEGA6 was also used to calculate average pairwise distance (APD) by including only non-hypermutated HIV-1 DNA sequences in the analyses. This procedure was used to identify clonal sequences and exclude cross-contamination of sequences from different donors. Supplementary Figure 7. **Methods, Paragraph 2 Page 20:** *Statistical analyses.* Average pairwise distances were calculated by MEGA6 through the exclusion of hypermutants and identification of hypermutants was performed using the HIVdb algorithm (Stanford University HIV Drug Resistance Database) (38, 57). By defining the overall diversity of the donors' sequences, this analysis eliminated potential bias that identical sequences were from chance alone.
		(*b*) Describe any methods used to examine subgroups and interactions Not applicable
		(*c*) Explain how missing data were addressed **Results, Table 3:** For donor missing proviral sequences matching the dominant plasma HIV-1 RNA sequences, % was designated as <1/total proviral sequences detected by single genome sequencing. For donors missing qVOA-derived sequences matching plasma-derived sequences, % designated as 0. For proviral sequences not matching qVOA-derived sequences, % was designated as <1/total proviral sequences

		detected by single genome sequencing. **Result, Table 4:** Integration site analyses only performed on donors who had matching sequencing plasma, proviral, and qVOA sequences. For the other donors it would not be possible to identify the origin of non-suppressible viremia
		(*d*) If applicable, describe analytical methods taking account of sampling strategy Not applicable
		(*e*) Describe any sensitivity analyses Not applicable
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **Results, Paragraph 1, Page 5:** We evaluated and analysed eight individuals referred for non-suppressible viremia, plasma HIV-1 RNA above 20 copies/mL and detectable for >6 months, despite reported adherence to ART and absence of drug resistance to the current regimen. **Results, Paragraph 2, Page 6:** *Identical plasma-derived HIV-1 sequences contribute to detectable viremia.* All eight donors were evaluated using a number of different approaches (Supplementary Figure 1A-F) (*23, 24, 34-37*). In all eight donors, groups of identical HIV-1 RNA *gag-pro-pol* sequences derived from single HIV-1 RNA templates (single genome sequencing [SGS]) (Supplementary Figure 1A) were found in plasma samples. Drug concentrations for all donors were within the expected therapeutic range based upon target trough concentrations (Supplementary Table 2) (*39, 40*). **Results, Paragraph 1, Page 7:** *Identical plasma-, proviral-, and viral outgrowth-derived HIV-1 sequences.* Genomic DNA (gDNA) from PBMC or purified $CD4^+$ T-cells and HIV-1 RNA from $p24^+$ qVOA wells were analyzed by *gag-pro-pol* single genome sequencing (SGS) (Supplementary Figure 1B-1D). Matching identical *gag-pro-pol* sequences were found in proviral DNA and plasma HIV-1 RNA in six of the eight donors (K-01, C-02, C-03, F-07, P-08, and R-09) (Figures 1-4, Supplementary Figures 2-3). In three of these donors (C-02, C-03, and R-09), the identical viral sequences in plasma also matched multiple qVOA-derived sequences from $p24^+$ wells (Table 3, Figures 1-3). In one donor (F-07), there were identical proviral and qVOA-derived sequences, but no sequences were found in plasma that matched (0 of 39). A different provirus from this donor matched plasma viral sequences, but these sequences were not found in $p24^+$ qVOA wells (Table 3, Figure 4). **Results, Paragraph 2, Page 7:** *Persistent viremia from intact, replication-competent HIV-1 proviruses.* Near full-length (NFL) viral sequence analyses were performed on samples from the four donors (R-09, C-03, C-02, and F-07) for which there were identical *gag-pro-pol* sequence matches between plasma and $p24^+$ qVOA cultures. **Results, Paragraph 2, Page 8:** *Proof that non-suppressible viremia can originate from clones of infected cells.* We performed integration site analysis (ISA) (Supplementary Figure 1E) on genomic DNA extracted from purified $CD4^+$ T-cells from the donors (R-09, C-03, C-02, and F-07) with plasma and/or proviral sequences that matched viruses obtained from qVOAs using a multiple displacement

		amplification (MDA) approach (*34, 42, 43*). **Discussion, Paragraph 2, Page 10:** Although the frequency of non-suppressible viremia of clonal origin is unknown, a total of 14 cases referred to us for evaluation, eight of which are described here, were from a group of approximately 2,000 patients cared for at the referring centers. The other 6 donors were beyond the scope of this initial evaluation.
		(b) Give reasons for non-participation at each stage As noted above, 6 additional patients have been referred with non-suppressible viremia since this initial study was competed. Resources were not available to evaluate these subsequent individuals.
		(c) Consider use of a flow diagram Not needed because of small N (8). Characteristics of the participants are reported in Tables 1 and 2 and a workflow of testing in Supplementary Figure 1.
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders The characteristics of the 8 study participants are summarized in **Table 1**.Other virologic and immunologic characteristics of the donors are provided in **Table 2.**
		(b) Indicate number of participants with missing data for each variable of interest **Results, Paragraph 1, Page 7:** *Identical plasma-, proviral-, and viral outgrowth-derived HIV-1 sequences.* Genomic DNA (gDNA) from PBMC or purified $CD4^+$ T-cells and HIV-1 RNA from $p24^+$ qVOA wells were analyzed by *gag-pro-pol* single genome sequencing (SGS) (Supplementary Figure 1B-1D). Matching identical *gag-pro-pol* sequences were found in proviral DNA and plasma HIV-1 RNA in six of the eight donors (K-01, C-02, C-03, F-07, P-08, and R-09) (Figures 1-4, Supplementary Figures 2-3). **Results, Paragraph 1, Page 7:** In three of these donors (C-02, C-03, and R-09), the identical viral sequences in plasma also matched multiple qVOA-derived sequences from $p24^+$ wells (Table 3, Figures 1-3). In one donor (F-07), there were identical proviral and qVOA-derived sequences, but no sequences were found in plasma that matched (0 of 39). A different provirus from this donor matched plasma viral sequences, but these sequences were not found in $p24^+$ qVOA wells (Table 3, Fig. 4). **Discussion, Paragraph 1, Page 11:** In the remaining four donors, there were viruses with identical sequences found in the plasma, but these sequences did not match any of the sequences obtained by qVOA (Supplementary Figures 2-5). In two of these four donors (K-01 and P-08), proviruses were found that matched the HIV-1 RNA in plasma, but did not match viral RNA sequences from $p24^+$ qVOA wells (Supplementary Figures 2-3 and Table 3). **Discussion, Paragraph 1, Page 11:** In two donors, proviruses were not found with sequences that matched HIV-1 RNA in plasma (T-05, 0 of 114 sequences; A-06, 0 of 77 sequences), suggesting that the virus-producing cells were either present at a low frequency in the periphery or absent from the blood and likely residing in lymphoid tissues (Supplementary Figures 3-4). **Results, Paragraph 2, Page 7:** *Persistent viremia from intact, replication-competent HIV-1 proviruses.* Near full-length (NFL) viral sequence analyses were performed on

		samples from the four donors (R-09, C-03, C-02, and F-07) for which there were identical *gag-pro-pol* sequence matches between plasma and $p24^+$ qVOA cultures. **Results, Paragraph 2, Page 8:** *Proof that non-suppressible viremia can originate from clones of infected cells.* We performed integration site analysis (ISA) (Supplementary Figure 1E) on genomic DNA extracted from purified $CD4^+$ T-cells from the donors (R-09, C-03, C-02, and F-07) with plasma and/or proviral sequences that matched viruses obtained from qVOAs using a multiple displacement amplification (MDA) approach (*34, 42, 43*). **Results, Paragraph 1, Page 9:** We determined the frequency of these integrated intact proviruses compared to all other proviruses (Table 4) and found them to be a minor fraction of all infected cells (0.03-1.1%) (*34*).
Outcome data	15*	Report numbers of summary measures **Results, Paragraph 1, Page 5:** Medians and ranges for donor characteristics as well as donor virologic and immunologic characteristics (Table 1-2). **Results, Paragraph 2, Page 6:** Medians and ranges for identical plasma-derive HIV-1 sequences (Table 3 data).
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included Non-applicable.
		(*b*) Report category boundaries when continuous variables were categorized Not applicable
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period Not applicable
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses Not applicable
Discussion		
Key results	18	Summarise key results with reference to study objectives **Discussion, Paragraph 1, Page 9:** Here we describe a new cause of HIV-1 plasma viremia observed in standard clinical practice that is not suppressible by ART, and we show that the viremia can arise from large clones of HIV-1 infected cells. **Discussion, Paragraph 1, Page 9:** Although each of the eight individuals have unique features, the consistent finding, in all donors was that non-suppressible plasma viremia consisted of one or more large groups of identical viral sequences. The largest group of identical viral sequences in plasma comprised 37.5%-100% (median 58.6%) of all HIV-1 RNA in the plasma (Table 3). **Discussion, Paragraph 1, Page 10:** In the current study, drug resistance to the donors' ART regimen was not evident by sequence analysis nor was medication non-adherence, since antiretroviral drug concentrations measured in random plasma samples were within the therapeutic range (Supplementary Tables 1 and 2).

		Longitudinal analyses of the viral sequences in all donors showed no evidence of virus evolution over time, indicating that these viruses originated from a stable, non-evolving reservoir of infected cells. For all three donors with infectious clonal viremia, identical sequence matches were found between plasma HIV-1 RNA, proviral DNA and HIV-1 RNA in outgrowth cultures across the 1.5 kb *p6/protease/reverse transcriptase* amplicon and the full-length 2.6 kb *envelope* amplicon. **Discussion, Paragraph 1, Page 11:** In three of the eight donors, identical sequences were found in plasma, proviral DNA, and qVOA, indicating that the viruses were replication-competent and potentially transmissible by exposure to blood (Figures 1-3 and Table 3). In a fourth donor, identical sequences were obtained from proviral DNA and qVOA that did not match any viral sequences in plasma, indicating for that donor, the viruses in the plasma had a different origin (Figure 4 and Table 3). **Discussion, Paragraph 2, Page 11:** In the four donors with clones containing intact proviruses that produced infectious virus, we identified their corresponding integration sites (Tables 3-4 and Supplementary Figure 6). **Discussion, Paragraph 2, Page 12:** Integration site analyses showed that the clones carrying intact proviruses were a small fraction (0.03-1.1%) of all of the infected cells. However, given the large number of $CD4^+$ T-cells in the human immune system, the clones were estimated to comprise large numbers of infected cells (50 million-350 million cells) based upon the percentage of matching proviruses and estimated total body $CD4^+$ T-cells (Table 4) (*46*).
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **Discussion, Paragraph 2, Page 12:** The very low frequency of cells infected with the specific intact proviruses makes analyses of the $CD4^+$ T-cell subsets and proviruses involved very challenging given the limitations of cell sampling and current technologies to detect specific proviruses, although such work is in progress. **Discussion, Paragraph 3, Page 12:** The reason non-suppressible viremia developed after years of ART-mediated suppression is not known. The mechanisms that underlie infected cell expansion and proviral expression are undefined. **Discussion, Paragraph 1, Page 14:** The non-suppressible viremia in our study participants could result in immune activating or inflammatory effects, although recent work from Gandhi *et al* has shown no association between measures of HIV-1 persistence and inflammation/activation in over 300 well-suppressed individuals (*33*). Additional studies with matched controls without non-suppressible viremia are needed to answer the role of immune activation and inflammation in clonal expansion and virus production from clones.
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **Discussion, Paragraph 1, Page 9:** He were describe a new cause of HIV-1 plasma viremia observed in standard clinical practice that is not suppressible by ART, and we show that the viremia can arise from large clones of HIV-1 infected cells. **Discussion, Paragraph 1, Page 9:** Although each of the eight individuals have unique features, the consistent finding, in all donors was that non-suppressible plasma viremia consisted of one or more large groups of identical viral sequences. The largest

		group of identical viral sequences in plasma comprised 37.5%-100% (median 58.6%) of all HIV-1 RNA in the plasma (Table 3). **Discussion, Paragraph 2, Page 13:** Our findings have important implications for efforts to develop a cure for HIV-1 infection. It is widely believed that the latent HIV-1 reservoir is the main obstacle to a cure. Some have questioned the relevance of persistent viremia on ART, but our findings show that some of the viruses that persist in plasma can be infectious and could quickly initiate viral rebound if ART is discontinued (*48, 52, 53)*.
Generalisability	21	Discuss the generalisability (external validity) of the study results **Discussion, Paragraph 2, Page 10:** The important implications of these findings for clinical management are that changes in the antiretroviral drug regimen or efforts to enhance drug adherence are unlikely to change the viremia that is being produced by infected cell clones, which are not affected by current antiretroviral drugs. Drugs that block virion production from cells that are already infected would be expected to lower viremia, but such agents are not currently available. HIV-1 protease inhibitors that are available block virion maturation into infectious particles, but do not reduce virion production from cells that are already infected (*49*). Although the frequency of non-suppressible viremia of clonal origin is unknown, a total of 14 cases referred to us for evaluation, eight of which are described here, were from a group of approximately 2,000 patients cared for at the referring centers. **Discussion, Paragraph 2, Page 13:** Our findings have important implications for efforts to develop a cure for HIV-1 infection. It is widely believed that the latent HIV-1 reservoir is the main obstacle to a cure. Some have questioned the relevance of persistent viremia on ART, but our findings show that some of the viruses that persist in plasma can be infectious and could quickly initiate viral rebound if ART is discontinued (*48, 52, 53)*. This "active" HIV-1 reservoir is thus also a key barrier to curing HIV-1 infection. It is possible, in cases in which there are much lower levels of viremia, that there are smaller clones of infected cells releasing infectious virus at levels that are below detection in plasma, but are still capable of rekindling viral replication if ART is interrupted.
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **Funding, Page 3:** The National Cancer Institute, National Institutes of Health (NIH); the Howard Hughes Medical Research Fellows Program, Howard Hughes Medical Institute; the Bill and Melinda Gates Foundation; the Office of AIDS Research; The American Cancer Society; the National Cancer Institute through a Leidos subcontract; the National Institute for Allergy and Infectious Diseases, NIH, to the I4C Martin Delaney Collaboratory; the University of Rochester Center for AIDS Research and the University of Rochester HIV/AIDS Clinical Trials Unit.

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely

available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.