Near-Full-Length Single-Genome HIV-1 DNA Sequencing

Guinevere Q. Lee[1], **Mathias Lichterfeld**[2,3,4,5]

[1]Weill Cornell Medicine, New York, NY, USA.

[2]Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, USA.

[3]Brigham and Women's Hospital, Boston, MA, USA.

[4]Harvard Medical School, Boston, MA, USA.

[5]Broad Institute of MIT and Harvard, Cambridge, MA, USA.

Abstract

HIV-1 integrates into human chromosomes to establish a lifelong reservoir of virally infected cells. However, the majority of integrated viral DNA shows lethal defects, likely due to errors introduced during reverse transcription of viral RNA. Identifying and quantifying HIV-1 DNA sequences that are genome-intact and can give rise to rebound viremia during antiretroviral treatment interruption are critical steps for understanding the complexity and evolutionary dynamics of HIV-1 reservoir cells. Here, we describe FLIP-Seq, (**F**ull-**L**ength **I**ndividual **P**roviral **S**equencing) a near full-length, single-genome next-generation sequencing approach for analyzing HIV-1 DNA in human cells. Briefly, this technique involves sequential dilution of proviral DNA to single genomes, amplification of near full-length viral DNA, deep sequencing of amplification products, and a biocomputational analysis designed to distinguish genome-intact HIV-1 DNA from defective viral DNA species. This procedure can be performed with small numbers of cells from highly purified CD4 T cell subsets, allows to generate an absolute quantification of viral sequences present in a given cell population, provides insight into phylogenetic associations of intact proviruses, and can identify proportions of sequence-identical proviruses likely derived from clonally expanded reservoir cells.

Keywords

Deep sequencing; Provirus; Viral DNA; Latency; Reservoir; HIV-1; Persistence

1 Introduction

HIV-1 can persist lifelong in the human body by integrating its viral DNA into the host genome. Current antiretroviral therapies effectively suppress further viral replication cycles, but do not eliminate integrated copies of viral DNA. Traditionally, these integrated copies of viral DNA have been referred to as "proviruses"—upon therapy cessation, these integrated viral DNA genomes are responsible for the production of new viral progenies, and for

MLICHTERFELD@partners.org.

fueling rebound viremia. Quantifying proviruses has represented a considerable challenge in the past, specifically since it became evident that the majority of HIV-1 DNA sequences in genomic DNA of patients treated with antiretroviral therapy (ART) display lethal defects that preclude viral replication. These defective proviruses, sometimes colloquially referred to as "junk DNA", include proviruses with large deletions (resulting from defects during viral reverse transcription), with APOBEC3G/3F-mediated hypermutations or with small sequences variations that nevertheless can cause premature stop codons, internal inversions or other deletions at structurally important locations in the proviral genome. The small proportion of HIV-1 proviruses that do not exhibit any of these defects are likely to be fully replication-competent; however, they cannot be easily distinguished from defective proviruses using standard PCR assays and are difficult to quantify. Traditionally, replication-competent HIV-1 DNA has been identified using functional viral outgrowth assays, during which patient-derived PBMC are stimulated with viral reactivating agents, followed by biochemical assays for viral antigen detection. However, such complex tissue culture assays frequently do not allow for quantitative evaluations, particularly when small cell input numbers are analyzed.

As an assay that can reliable identify and quantify genome-intact HIV-1 proviruses in patient-derived cells, we here describe a single-genome, near full-length next-generation assay that we termed "FLIP-Seq" ("full-length individual proviral sequencing"). When first developed [1], full-length viral genome analysis was based on single-genome amplification and sequencing of four overlapping amplicons spanning the entire HIV-1 genome, using Sanger sequencing technology. In comparison to this original assay, our protocol has been optimized and streamlined to include the following modifications: First, instead of using serial dilutions to determine the dilution range for single-genome amplification, we begin the assay by using droplet digital PCR to quantify the total number of HIV *gag* DNA copies per microliter of nucleic acid extract, and subsequently using this concentration value as a guide to dilute the nucleic acid extract to single proviral genomes, thereby improving through-put. Second, instead of a four-amplicon amplification strategy, we amplify the viral genome using a single-amplicon approach, thereby reducing enzyme cost. Third, instead of using Sanger sequencing, we use Illumina deep sequencing which improves resolution and controls for non–single-template amplification events. Finally, instead of manual bioinformatics analyses, we automated the classification of "intact" versus "defective" proviruses with a R-language based bioinformatics pipeline. The bioinformatics pipeline is optimized for subtype B and C HIV-1 [2, 3]. Please cite this protocol using references [2, 3].

It is important to recognize that the viral 5-LTR promoter region is not captured with our protocol; this omission must be considered as a limitation of the FLIP-Seq technique, given that the 5-LTR promoter can critically influence proviral replication fitness [4]. A modification of the FLIP-Seq protocol that includes a prior whole-genome amplification step of single proviral species allows to analyze the complete proviral sequence, including the 5-LTR region, and also offers opportunities to characterize the corresponding chromosomal integration site [5, 6].

2   Materials

Prepare the following supplies and equipment.

1. PCR-grade water.

2. QIAGEN DNeasy Blood and Tissue Kit (*see* Note 1).

3. Droplet digital PCR (ddPCR) reader and reagents.

   a. ddPCR Supermix for Probes,

   b. LTRgag primers (F and R) and probe (P).

   c. 20 μM LTRgagF—5′-TCTCGACGCAGGACTCG-3′.

      • 20 μM LTRgagR—5′-TACTGACGCTCTCGCACC-3′

      • 20 μM LTRgagP—5′–/56-FAM/ CTCTCTCCT / ZEN/ TCTAGCCTC/31ABkFQ/–3′.

4. Invitrogen Platinum Taq DNA Polymerase High Fidelity (*see* Note 1).

5. Invitrogen 10 mM dNTP Mix PCR Grade (*see* Note 1).

6. HIV-specific primers; [7] with modifications.

1stPCR Forward	U5-623F	5′-AAATCTCTAGCAGTGGCGCCCGAACAG-3′
1stPCR reverse	U5-601R	5′-TGAGGGATCTCTAGTTACCAGAGTC-3′
2ndPCR forward	U5-638F	5′-GCGCCCGAACAGGGACYTGAAARCGAAAG-3′
2ndPCR reverse	U5-547R	5′-GCACTCAAGGCAAGCTTTATTGAGGCTTA-3′

7. PCR-grade 96-well plates.

8. BIO-RAD C1000 thermocycler (*see* Note 1).

9. Agencourt AMPure XP—PCRPurification (*see* Note 1).

10. Illumina library preparation kit.

11. Illumina MiSeq Sequencer.

12. Installation of HIVSeqinR [2].

3   Methods

A schematic overview of the full protocol is summarized in prior publication [2]. All steps below should be performed with filtered pipette tips.

3.1 Nucleic Acid Extraction

The objective of this step is to extract total nucleic acids from cells infected with HIV.

1. Load 1–5 million cells containing HIV DNA into a QIAGEN DNeasy Blood and Tissue kit column.

2. Follow the manufacturer's protocol.

3. Elute the nucleic acid from the column filter using 50–200 μL DEPC-treated water. Label the tube "EXT." This will be the starting material for Subheadings 3.2 and 3.3. Store at −80 °C when not in use. Minimize freeze–thaw.

3.2 Droplet Digital PCR (ddPCR) Quantification of Total HIV DNA

The objective of this step is to obtain an estimate of HIV DNA concentration in EXT by ddPCR amplification of a short fragment in HIV LTR-gag (HXB2 684-810).

1. In a PCR-clean room, prepare ddPCR master mix for the amplification of HIV LTR-gag.

2. Per reaction planned, use 10 μL of ddPCR Supermix for Probes, 1 μL of 20 μM forward primer LTRgagF, 1 μL of 20 μM reverse primer LTRgagR, and 0.35 μL of 20 μM ddPCR probe LTRgagP. Add 1–3 μL of EXT and DEPC-treated water to a final volume of 22 μL (*see* Note 2).

3. Generate ddPCR amplicons under the following thermocycler condition using the BIO-RAD C1000 thermocycler: 10 min at 95 °C, 46 cycles [30 s at 95 °C, 1 min at 60 °C], 10 min at 98 °C and an infinite hold at 4 °C.

4. Follow standard ddPCR operating procedure to obtain an estimate of total HIV LTR-*gag* copies per microliter in EXT.

3.3 PCR Amplification of Near Full-Genome HIV DNA

The objective of this step is to single-genome amplify (SGA) HIV-1 DNA genomes from EXT in a single-amplicon-fashion covering HXB2 coordinates 638-9632. To achieve single-genome amplification, HIV DNA templates in EXT are diluted to single genomes according to the total HIV DNA concentration obtained in Subheading 3.2 by ddPCR, followed by PCR amplification. According to Poisson distribution statistics, to achieve a probability of ~86% single-template in each PCR-positive reaction, EXT should be diluted so that less than 33% of the PCR reactions should be positive for HIV amplification. The protocol below outlines steps to set up a single 96-well plate by adding 30 absolute copies of HIV DNA templates and spreading them into 90 wells to achieve limiting dilution.

1. Dilute EXT to 30 copies of HIV total DNA per 90 μL DEPC-treated H_2O (i.e., 0.33 copies/μL HIV total DNA).

2. In a PCR-clean room, prepare 90-well worth of PCR master mixes for both first- and second-round PCR as follows: 1332 μL DEPC-treated H_2O, 180 μL Invitrogen Platinum Taq DNA Polymerase High Fidelity 10 × PCR buffer (final concentration 1 ×), 72 μL of 50 mM $MgSO_4$ (provided in-kit; final concentration 2 mM), 36 μL of 10 mM dNTP (final concentration 0.2 mM), 36 μL of each of 20 μM forward and reverse primers (*see* Subheading 2; final concentration 0.4 μM), and 18 μL of 5 units/μL of Invitrogen Platinum Taq DNA Polymerase High Fidelity (final concentration 1 unit/μL). The total volume of each of the first- and second-round master mix should be 1710 μL.

3.   To the first-round PCR master mix, add 90 μL of diluted EXT. Mix by inverting the tube 20 times. Label this tube "SGA mix."

4.   Store the second-round PCR mix at −20 °C until **step 7**.

5.   In a 96-well plate, distribute the SGA mix into 90 wells at 20 μL per well. Label this plate "1stPCR."

6.   Incubate the 1^{st}PCR 96-well plate under the following thermocycler program using the Bio-Rad C1000 thermocycler: 2 min at 92 °C, 10 cycles [10 s at 92 °C, 30 s at 60 °C, 10 min at 68 °C], 20 cycles [10 s at 92 °C, 30 s at 55 °C, 10 min at 68 °C], 10 min at 68 °C, and an infinite hold at 4 °C. This program should run for approximately 6.5 h.

7.   At the end of the first-round PCR, thaw the second-round PCR mixes (**step 4**). In a 96-well plate, distribute 19 μL of the second-round PCR mix to 90 wells. Label this plate "2^{nd}PCR."

8.   Transfer 1 μL of the first-round PCR products to the 2^{nd}PCR plate.

9.   Incubate the 2nd PCR 96-well plate under the same thermocycler conditions as in **step 5**.

10.   At the end of the second-round PCR, subject 3 μL of the product to DNA gel electrophoresis. Note, to ensure limiting dilution to single genomes was achieved, confirm at this point that less than 30% of the PCR reactions are HIV-positive (*see* Notes 3 and 4).

11.   Transfer the contents of only the PCR-positive wells to a clean 96-well plate. Label the plate "2^{nd}PCRTransfer." These wells contain amplified HIV DNA genomes and will be beads-purified and sequenced in the next step.

3.4   Beads Purification of PCR Amplicons

The objective of this step is to purify the PCR products for Illumina library preparation and sequencing.

1.   Vortex Agencourt AMPure XP beads to ensure the magnetic beads are thoroughly mixed.

2.   Add 36 μL AMPure XP beads into each well in the 2^{nd}PCRTransfer plate. Pipet up and down to mix.

3.   Leave at room temperature for 10 min.

4.   Put the plate on a 96-well format magnet for 2 min.

5.   While leaving the plate on-magnet, aspirate and discard supernatant.

6.   While leaving the plate on-magnet, wash pellet twice with 70% ethanol.

7.   Aspirate and discard supernatant.

8.   Air-dry.

9.   Remove plate from magnet and add 40 μL of H_2O. Mix well by pipetting up and down.

10.   The eluted DNA should now be in the aqueous phase. Put the plate on a 96-well format magnet for 2 min. Remove 35 μL of the eluant and transfer to a fresh 96-well plate. Label this plate "AMPure."

3.5 Illumina Library Preparation and Sequencing

Submit the AMPure plate to a DNA core facility for Illumina library preparation and MiSeq deep sequencing, according to local standards and protocols. We recommend using a physical shearing, blunt-end adaptor ligation approach (as opposed to Nextera XT), and a sequencing kit that provides read length of at least 150 base pairs with a minimum number of 1000 reads per base. Note that the Illumina library preparation kit Nextera XT can also be used as an alternative, but the bioinformatics pipeline described in Subheading 3.6 is not fully compatible with the viral genome sequences generated by Nextera XT; manual validation may be required.

3.6 Viral Genome Bioinformatics Analysis

1.   De novo assemble short reads in FASTQ file. If using the sequencing service provided by the Massachusetts General Hospital (MGH) Center for Computational & Integrative Biology (CCIB) DNA Core (https://dnacore.mgh.harvard.edu/new-cgi-bin/site/pages/index.jsp; open to international submission), this step is included in the service package. Otherwise, use SPAdes at default settings to complete de novo assembly.

2.   Obtain consensus sequence of the de novo assembled sequences in FASTA format. This is the input file for HIVSeqinR.

3.   Download HIVSeqinR from http://www.github.com/guineverelee.

4.   HIVSeqinR is a R-based script and can be run locally in a PC or a Mac. The purpose of the script is to bioinformatically analyze an HIV DNA genome sequence and to categorize the genome as either intact or defective [2, 3]. For instructions on how to use this script, please see documentation files at http://www.github.com/guineverelee. For additional sequence data quality control steps, *see* Notes 5 and 6.

4   Notes

1.   This protocol is incompatible or suffers reduced sensitivity if performed with certain nucleic acid extraction methods, Taq polymerase, thermocycler models and PCR amplicon purification methods. We have validated the protocol to be used with the reagents and equipment listed in Subheading 2.

2.   In cases where over 3000 copies of HIV-1 DNAper μL in EXT is expected, diluted appropriately to below 3000 copies. This is to ensure less than 30% of the ddPCR droplets are positive to achieve a probability of ~86% single-template per droplet according to Poisson distribution.

3. Maintain a positive control such as the 8E5/LAV cell line to monitor the sensitivity of the assay.

4. Maintain a HIV-amplicon contamination-free work environment. Ideally, EXT, 1stPCR, and 2ndPCR should be performed in separate rooms. Routinely swab bench and equipment surfaces for HIV amplification to detect contamination. Always include both positive (e.g., bulk 8E5/LAV extracted DNA) and negative (no template) controls in all PCRruns.

5. Since single-genome amplification work is extremely sensitive to contamination by other high concentration HIV DNA sources, it is important to routinely use phylogenetic methods to validate sequencing results. HIV DNA sequences derived from a single patient regardless of sampling time point should almost always group into a monophyletic cluster, with exception in cases of superinfections.

6. In this protocol, limiting dilution is achieved by plating 0.33 HIV genomes per PCR reaction. Poisson distribution estimates that ~86% of the PCR-positive reactions under this dilution factor are true single genome. In other words, about 14% of the PCR-positive reactions are not single genome and should be excluded for downstream analyses. Samples that contain a higher-than-usual amount of nucleotide mixtures are suspects of non–single-genome-amplified products. This quality control step is included in the HIVSeqinR pipeline when used with MGH CCIB DNA core-generated output files. Samples that yield multiple contigs after de novo assembly are also suspects and should be discarded from downstream analysis.

References

1. Ho YC, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DI, Lai J, Blankson JN, Siliciano JD, Siliciano RF (2013) Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. Cell 155(3):540–551. 10.1016/j.cell.2013.09.020 [PubMed: 24243014]

2. Lee GQ, Reddy K, Einkauf KB, Gounder K, Chevalier JM, Dong KL, Walker BD, Yu XG, Ndung'u T, Lichterfeld M (2019) HIV-1 DNA sequence diversity and evolution during acute subtype C infection. Nat Commun 10(1):2737. 10.1038/s41467-019-10659-2 [PubMed: 31227699]

3. Lee GQ, Orlova-Fink N, Einkauf K, Chowdhury FZ, Sun X, Harrington S, Kuo Hh, Hua S, Chen HR, Ouyang Z, Reddy K, Dong K, Ndung'u T, Walker BD, Rosenberg ES, Yu XG, Lichterfeld M (2017) Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4+ T cells. J Clin Invest. 127 (7):2689–2696. 10.1172/JCI93289 [PubMed: 28628034]

4. van Opijnen T, Jeeninga RE, Boerlijst MC, Pollakis GP, Zetterberg V, Salminen M, Berkhout B (2004) Human immunodeficiency virus type 1 subtypes have a distinct long terminal repeat that determines the replication rate in a host-cell-specific manner. J Virol 78(7):3675–3683. 10.1128/jvi.78.7.3675-3683.2004 [PubMed: 15016888]

5. 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 (2019) Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy. J Clin Invest 129(3):988–998. 10.1172/JCI124291 [PubMed: 30688658]

6. Garcia-Broncano P, Maddali S, Einkauf KB, Jiang C, Gao C, Chevalier J, Chowdhury FZ, Maswabi K, Ajibola G, Moyo S, Mohammed T, Ncube T, Makhema J, Jean-Philippe P, Yu XG, Powis KM, Lockman S, Kuritzkes DR, Shapiro R, Lichterfeld M (2019) Early antiretroviral therapy in neonates

with HIV-1 infection restricts viral reservoir size and induces a distinct innate immune profile. Sci Transl Med 11(520). 10.1126/scitranslmed.aax7350

7. Li B, Gladden AD, Altfeld M, Kaldor JM, Cooper DA, Kelleher AD, Allen TM (2007) Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. J Virol 81(1):193–201. 10.1128/JVI.01231-06 [PubMed: 17065207]