Journal of Visualized Experiments

Amplification of near full-length HIV-1 proviruses in preparation for Next Generation Sequencing

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE58016R2
Full Title:	Amplification of near full-length HIV-1 proviruses in preparation for Next Generation Sequencing
Keywords:	next generation sequencing; HIV-1; HIV; Proviruses; Single-proviral Sequencing; HIV DNA; Latency; Replication-competency; Antiretroviral therapy; HIV reservoir; Integrated HIV
Corresponding Author:	Bonnie Hiener The Westmead Institute for Medical Research Westmead, NSW AUSTRALIA
Corresponding Author's Institution:	The Westmead Institute for Medical Research
Corresponding Author E-Mail:	bonnie.hiener@sydney.edu.au
Order of Authors:	Bonnie Hiener
	John-Sebastian Eden
	Bethany A Horsburgh
	Sarah Palmer
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Open Access (US$4,200)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	The Westmead Institute for Medical Research. 176 Hawkesbury Rd Westmead NSW 2145 Sydney Australia

19 July 2018

To the Editors of JoVE

The manuscript entitled “Full-Length Individual Proviral Sequencing for the Identification of Intact HIV-1 Genomes” (JoVE58016) was recently reviewed for publication in JoVE. We greatly appreciate the constructive critique from the editor as this has improved the quality of our manuscript.

We have carefully addressed each comment within the manuscript file and updated the manuscript accordingly.

Kind regards,

Bonnie Hiener
Master’s Student
The Westmead Institute for Medical Research
Sydney Australia

TITLE:

Amplification of Near Full-Length HIV-1 Proviruses for Next-Generation Sequencing

AUTHORS & AFFILIATIONS:

Bonnie Hiener, John-Sebastian Eden, Bethany A Horsburgh, Sarah Palmer

Centre for Virus Research, The Westmead Institute for Medical Research, The University of Sydney, Sydney, New South Wales, Australia

Corresponding Author:

Bonnie Hiener

bonnie.hiener@sydney.edu.au

E-mail Addresses of Co-authors:

John-Sebastian Eden (js.eden@sydney.edu.au)

Bethany A Horsburgh (bethany.horsburgh@sydney.edu.au)

Sarah Palmer (sarah.palmer@sydney.edu.au)

KEYWORDS:

Next-generation sequencing, HIV-1, HIV, proviruses, single-proviral sequencing, HIV DNA, latency, replication-competency, antiretroviral therapy, HIV reservoir, integrated HIV

SUMMARY:

Full-length individual proviral sequencing (FLIPS) provides an efficient and high-throughput method for the amplification and sequencing of single, near full-length (intact and defective) HIV-1 proviruses and allows for determination of their potential replication-competency. FLIPS overcomes limitations of previous assays designed to sequence the latent HIV-1 reservoir.

ABSTRACT:

The Full-Length Individual Proviral Sequencing (FLIPS) assay is an efficient and high-throughput method designed to amplify and sequence single, near full-length (intact and defective), HIV-1 proviruses. FLIPS allows determination of the genetic composition of integrated HIV-1 within a cell population. Through identifying defects within HIV-1 proviral sequences that arise during reverse transcription, such as large internal deletions, deleterious stop codons/hypermutation, frameshift mutations, and mutations/deletions in cis acting elements required for virion maturation, FLIPS can identify integrated proviruses incapable of replication. The FLIPS assay can be utilized to identify HIV-1 proviruses that lack these defects and that are therefore potentially replication-competent. The FLIPS protocol involves: lysis of HIV-1 infected cells, nested PCR of near full-length HIV-1 proviruses (using primers targeted to the HIV-1 5' and 3' LTR), DNA purification and quantification, library preparation for Next-generation Sequencing (NGS), NGS, *de novo* assembly of proviral contigs, and a simple process of elimination for identifying replication-competent proviruses. FLIPS provides advantages over traditional methods designed to sequence integrated HIV-1 proviruses, such as single-proviral sequencing. FLIPS amplifies and sequences near full-length proviruses enabling replication competency to be determined, and

also uses fewer amplification primers, preventing the consequences of primer mismatches. FLIPS is a useful tool for understanding the genetic landscape of integrated HIV-1 proviruses, especially within the latent reservoir, however, its utilization can extend to any application in which the genetic composition of integrated HIV-1 is required.

INTRODUCTION:

Genetic characterization of the latent HIV-1 reservoir, which persists in individuals on long-term antiretroviral therapy (ART), has been vital to understanding that the majority of integrated proviruses are defective and replication-incompetent[1,2]. During the process of reverse transcription, errors are introduced into the integrated proviral sequence. Some mechanisms that generate defective proviral sequences include the error-prone HIV-1 reverse transcriptase enzyme[3], template switching[4], and/or APOBEC-induced hypermutation[5,6]. Two recent studies have found that approximately 5% of HIV-1 proviruses isolated from individuals on long-term ART are genetically intact, and potentially replication-competent, and may contribute to the rapid rebound in HIV-1 plasma levels upon cessation of ART[1,2,7]. Previous studies have identified that replication-competent HIV-1 proviruses persist in naïve and resting memory $CD4^+$ T cell subsets (including central, transitional and effector memory T cells), indicating the importance of targeting these cells in future eradication strategies[2,8,9].

Early insights into the distribution, dynamics and maintenance of the latent HIV-1 reservoir were achieved through utilization of single-proviral sequencing (SPS) methods that genetically characterize sub-genomic regions of the HIV-1 genome[10-13]. SPS is a versatile tool, able to sequence a single HIV-1 provirus from within a single infected cell. However, SPS is unable to determine the replication-competency of proviruses, since it only sequences sub-genomic regions and misses proviruses that contain large deletions within primer binding sites. A previous study has demonstrated that SPS overestimates the size of the replication-competent reservoir by 13- to 17-fold through selectively sequencing intact sub-genomic regions[2].

To address the limitations of SPS, Ho *et al.*[4] and Bruner *et al.*[1] developed assays to sequence near full-length HIV-1 proviruses. This allowed the frequency of genetically intact, and potentially replication-competent, HIV-1 proviruses in individuals on long-term ART to be determined. These assays amplified and sequenced (via Sanger sequencing) sub-genomic regions that were then assembled to obtain a sequence of the (intact or defective) HIV-1 provirus. Three limitations of this approach are: 1) the use of multiple sequencing primers increases the risk of unintentionally introducing defects into the proviral sequence, 2) primer mismatches may prevent amplification of particular proviruses, and 3) often the entire proviral sequence cannot be resolved due to the technicality of these methods.

To overcome the limitations of existing full-length HIV-1 proviral sequencing assays, we developed the full-length individual proviral sequencing (FLIPS) assay. FLIPS is a next-generation sequencing (NGS)-based assay which amplifies and sequences near full-length (intact or defective) HIV-1 proviruses in a high-throughput and efficient manner. FLIPS provides advantages over previous assays, as it limits the number of primers utilized; therefore, it decreases the chance of primer mismatches, which may limit the population of proviruses

captured or unintentionally introduce defects into a viral sequence. FLIPS is also less technically challenging than previous assays and involves 6 main steps: 1) lysis of HIV-1 infected cells, 2) amplification of single HIV-1 proviruses via nested PCR performed at limiting dilution using primers specific for the highly conserved HIV-1 5' and 3' U5 LTR region (**Figure 1A**), 3) purification and quantification of amplified products, 4) library preparation of amplified proviruses for NGS, 5) NGS, and 6) *de novo* assembly of sequenced proviruses to obtain contigs of each individual provirus.

Sequences generated by FLIPS can undergo a stringent process of elimination to identify those which are genetically intact and potentially replication-competent (**Figure 1C**)[2]. Genetically intact proviruses lack all known defects which result in generation of a replication-incompetent provirus. These defects include: inversion sequences, large internal deletions, hypermutation/deleterious stop codons, frameshifts, or mutations in the 5' packaging signal or major splice donor (MSD) site.

[place **Figure 1** here]

PROTOCOL:

All methods described here have been approved by the institutional review boards at the University of California San Francisco and the Western Sydney Local Health District, which includes The Westmead Institute for Medical Research.

1. Lysis of HIV-1-Infected Cells

Note: Cells may be isolated from peripheral blood, leukapheresis samples, bone marrow biopsy, or tissue biopsy. Cell populations may be sorted using fluorescence-activated cell sorting (FACS).

1.1. Prepare a lysis buffer containing 10 mM Tris-HCl, 0.5% Nonidet P-40**,** 0.5% Tween-20, and 0.3 mg/mL proteinase K. To a cell pellet, add 100 µL of lysis buffer per 1 x 10^6 cells. Pipette up and down to mix. Incubate at 55 °C for 1 hour followed by 85 °C for 15 min to lyse the cells and release genomic DNA for PCR amplification.

Note: Cell lysis is sufficient to obtain genomic DNA for amplification. No DNA isolation or purification is required. The protocol can be paused at this point and genomic DNA can be stored at -20 °C indefinitely.

2. Amplification of Single HIV-1 DNA Proviruses via Nested PCR

2.1. Mix the reagents for the first round PCR (PCR1) listed in **Table 1** (see **Table of Materials**). Add 38 µL of master mix to 85 wells (80 samples, 4 negative controls, 1 positive control) of a 96-well PCR plate (follow the layout in **Figure 1B**). Designate this plate 'PCR1'.

[place **Table 1** here]

Note: The primers used for PCR1 are:
BLOuterF: 5'-AAATCTCTAGCAGTGGCGCCCGAACAG-3' (HXB2 position 623-649)
BLOuterR: 5'-TGAGGGATCTCTAGTTACCAGAGTC-3' (HXB2 position 9662-9686)

2.2. After preparing the master mix, move the PCR1 plate to a clean area designated for the addition of genomic DNA.

2.2.1. Serially dilute genomic DNA from 1:3 to 1:81 with Tris-HCl (5 mM, pH 8), preparing 45 µL for each dilution (enough for 20 wells for each dilution). Add 2 µL of diluted genomic DNA to each sample well and 2 µL of Tris-HCl (5 mM, pH 8) to each negative control well (follow the layout in **Figure 1B**). Seal all the wells of the PCR1 plate, excluding the positive control well, using a clear adhesive seal (see **Table of Materials**).

Note: The above recommended dilutions serve only as a starting guide for determining the end-point dilution. Dilutions will depend on the concentration of integrated HIV-1 DNA within the samples.

2.3. Move to an area designated for the addition of positive control. Add 2 µL of positive control (pNL4-3 diluted to 10^5 copies/µl) to the positive control well of the PCR1 plate and seal the plate. Briefly spin the PCR1 plate in a PCR plate spinner or centrifuge (400 x g for 10 s at room temperature) to pull down any residual contents from the sides of the wells.

2.4. Run the PCR1 plate in a thermocycler: 94 °C for 2 min; then 94 °C for 30 s, 64 °C for 30 s, 68 °C for 10 min for 3 cycles; 94 °C for 30 s, 61 °C for 30 s, 68 °C for 10 min for 3 cycles; 94 °C for 30 s, 58 °C for 30 s, 68 °C for 10 min for 3 cycles; 94 °C for 30 s, 55 °C for 30 s, 68 °C for 10 min for 21 cycles; then 68 °C for 10 min (30 cycles total). Hold at 4 °C.

Note: The protocol can be paused here and the PCR1 plate kept at 4 °C for up to 2 days.

2.5. Mix the reagents for the second round of PCR (PCR2) listed in **Table 1**. Add 28 µL to 85 wells (80 samples, 4 negative controls, 1 positive control) of a new 96-well PCR plate (follow the layout in **Figure 1B**). Designate this plate 'PCR2'.

Note: The primers used for PCR2 are:
275F: 5'-ACAGGGACCTGAAAGCGAAAG-3' (HXB2 position 646-666)
280R: 5'-CTAGTTACCAGAGTCACACAACAGACG-3' (HXB2 position 9650-9676)

2.6. Briefly spin the PCR1 plate in a PCR plate spinner or centrifuge (400 x g for 10 s at room temperature) to pull down any residual contents from the sides of the wells. Add 80 µL of Tris-HCl (5 mM, pH 8) to each well of the PCR1 plate.

2.6.1. Transfer 2 µL of the PCR1 plate to the PCR2 plate using a multichannel pipette. Ensure samples are transferred well to well (*i.e.,* 2 µL from well A1 of PCR1 plate is transferred to well A1 of PCR2 plate). Seal the PCR2 plate using a clear adhesive seal (see **Table of Materials**).

2.6.2. Briefly spin the PCR2 plate in a PCR plate spinner or centrifuge (400 x g for 10 s at room temperature) to pull down any residual contents from the sides of the wells. Seal the PCR1 plate with a heat sealing film for long term storage at -20 °C (see **Table of Materials**).

2.7. Run the PCR2 plate in a thermocycler: 94 °C for 2 min; then 94 °C for 30 s, 64 °C for 30 s, 68 °C for 10 min for 3 cycles; 94 °C for 30 s, 61 °C for 30 s, 68 °C for 10 min for 3 cycles; 94 °C for 30 s, 58 °C for 30 s, 68 °C for 10 min for 3 cycles; 94 °C for 30 s, 55 °C for 30 s, 68 °C for 10 min for 31 cycles; then 68 °C for 10 min (40 cycles total). Hold at 4 °C.

Note: The protocol can be paused here and the PCR2 plate kept at 4 °C for up to 2 days.

2.8. Briefly spin the PCR2 plate in a PCR plate spinner or centrifuge to pull down any residual contents from the sides of the wells. Add 60 µL of Tris-HCl (5 mM, pH 8) to each well using a multichannel pipette.

2.8.1. Run 15 µL of each well on 2 x 48-well precast 1% agarose gels containing ethidium bromide (0.1–0.3 µg/mL, see **Table of Materials**). Use a ladder with a range up to 10 kb (see **Table of Materials**). Visualize to identify the wells containing the amplified product and their approximate sizes. Save the gel image.

Note: Ensure the positive control well contains amplified product. If the negative control wells contain amplified product, consider contamination and disregard plate.

2.9. Determine the dilution at which no more than 30% of wells are positive for amplified product.

Note: This is the end-point dilution in which a majority (80%) of wells contain amplified product from a single template. This dilution should be prepared and used in subsequent PCRs to obtain further proviral amplicons. Record the approximate size of each amplified product.

3. DNA Purification and Quantification

3.1. Briefly spin the PCR2 plate in a PCR plate spinner or centrifuge to pull down any residual contents from the sides of the wells. Transfer 40 µL from the wells containing amplified product (at or below the end-point dilution) to a new 96-well midi plate (with a well volume of 0.8 mL, see **Table of Materials**).

Note: Write down the original and new well position of each amplified product in a spreadsheet as a record of each amplified product to be sequenced.

3.2. Purify the amplified DNA products using a magnetic bead based PCR purification kit (see **Table of Materials**) to remove primers, nucleotides, enzymes, oils and salts. Before starting, bring magnetic beads to room temperature and prepare fresh 80% ethanol. Use new pipette tips when appropriate to avoid cross contamination of DNA samples.

3.2.1. Vortex magnetic beads to ensure they are thoroughly resuspended. Using a multichannel pipette, add 40 µL of beads to the 40 µL of amplified product in the 0.8 mL 96-well midi plate. Gently pipette up and down 10 times to mix. Alternatively, mix the solution by sealing the plate then shaking on a microplate shaker at 1,800 rpm for 2 min. Incubate at room temperature for 5 min.

3.2.2. Place the plate on a magnetic stand (see **Table of Materials**) for 2 min. Remove and discard supernatant.

3.2.3. Wash beads by adding 200 µL 80% of ethanol to each well with the plate on the stand. Incubate at room temperature for 30 s. Remove and discard the supernatant. Repeat once. Using a multichannel pipette with fine tips, remove any excess ethanol following the second wash. With the plate on the stand, allow the beads to air dry for 15 min.

3.2.4. Remove the plate from the stand. Add 30 µL of elution buffer (see **Table of Materials**) to each well. Gently pipette up and down 10 times to mix. Alternatively, mix the solution by sealing the plate then shaking on a microplate shaker at 1,800 rpm for 2 min. Incubate at room temperature for 2 min.

3.2.5. Place the plate on a magnetic stand for 2 min. Using a multichannel pipette, transfer 25 µL of supernatant (purified DNA) to a new 96-well PCR plate. Ensure that the samples are transferred well-to-well (*i.e.,* supernatant of well A1 in 0.8 mL plate is transferred to well A1 of new 96-well plate).

Note: 5 uL of the eluted sample is left behind to ensure no transfer of residual purification beads.

3.3. Determine and record the approximate concentration of each amplified DNA product following purification using a spectrophotometer (absorbance at a wavelength of 260 nm).

Note: Measuring the approximate concentration of each amplified product is necessary at this stage to ensure no samples are lost during the purification steps and that the approximate concentration is within the range of the standard curve used in the next stage (50 pg to 2 µg DNA in 100 µL of volume). The protocol can be paused here and cleaned samples can be stored at -20 °C for up to 6 months.

3.4. Quantify the concentration of DNA of each purified product using a dsDNA quantification kit (see **Table of Materials**).

Note: This involves using a standard curve to determine the concentration of dsDNA in a sample. The kit includes buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7.5), a lamda dsDNA standard, and a fluorescent dye. Keep all reagents on ice. Cover the tube containing dye with foil to avoid exposure to light. Measure the concentration of each amplified product in triplicate.

3.4.1. Dilute buffer to 1x in sterile DNase free H_2O. Add 99 µL of buffer to an appropriate number of empty wells (3 times the number of amplified products to be measured, see **Table 2** for an example layout) of a flat bottom tissue culture plate. Add 100 µL of buffer to 3 blank wells.

3.4.2 For each amplified product to be measured, add 1 µL of purified DNA (from step 3.2.5) in triplicate to each well containing buffer (see **Table 2** for an example layout).

[place **Table 2** here]

3.4.3. To prepare standards, dilute lamda dsDNA 10-fold from 2 ng/µL to 0.002 ng/µL. Add 100 µL of each dsDNA standard to 3 wells.

Note: Following step 3.4.3, the final concentration of the standards will range from 1 to 0.001 ng/µL

3.4.4. Dilute the fluorescent dye 1:200 with buffer. Quickly add 100 µL to each well containing sample, blanks, and standards. Mix up and down with a pipette. Cover the plate with foil to avoid contact with light.

3.4.5. Read fluorescence emission on a microplate reader (excitation at 480 nm, emission at 520 nm). Record results in a spreadsheet.

3.4.6. Determine the concentration of dsDNA in each sample using the fluorescence measurements recorded. Subtract the fluorescence measured in the blank wells from the sample and standards. Determine the average fluorescence for each sample and standard from the triplicates. Draw a standard curve based on the fluorescence measurements of the standards. Determine the concentration of the samples relative to the standards.

3.5. Dilute each purified product (obtained in step 3.2.5) with H_2O to 0.2 ng/µL.

4. Sequencing Library Preparation

4.1. Prepare amplified and purified proviral DNA for NGS using an NGS DNA library preparation kit (see **Table of Materials**). Follow the manufacturer's instructions for all tagmentation, PCR amplification, and clean-up steps [except that the reaction volumes including input DNA (from step 3.5) can be halved to extend the use of library preparation reagents].

4.2. Normalize libraries manually using a qPCR-based NGS library quantification kit (see **Table of Materials**) to determine the individual concentration of provirus. Combine individual provirus libraries in equimolar amounts to a final concentration of 4 nM, or as specified by the sequencing service provider.

Note: The protocol can be paused here and the pooled library stored at -20 °C.

4.3. Quantify the final pooled library using the same dsDNA quantification kit used in step 3.4. Follow the manufacturer's instructions. Determine the average fragment lengths by running 1 µL of the pooled library on an automated electrophoresis system using an appropriate kit (see the **Table of Materials**). Use the concentration and average fragment lengths to determine the molarity of the pooled library.

4.4. Combine 5 µL of the pooled library with 5 µL of 0.2 N NaOH to denature the library. Add 5 µL of 200 mM Tris-HCl to neutralize. Dilute the final library to 12.5 pM with chilled hybridization buffer (available with DNA library preparation kit) immediately prior to sequencing.

4.5. Perform 2 x 150 nucleotide (nt) paired-end sequencing on an appropriate NGS platform (see **Table of Materials**).

Note: When 96 proviral libraries are indexed per run, this yields approximately 20 million paired-end reads per run or 200,000 reads per individual provirus library for analysis following de-multiplexing. Steps 4.4 and 4.5 are typically performed by a sequencing facility.

5. *De Novo* Assembly of Sequenced HIV-1 Proviruses

Note: To obtain the genetic sequence of each amplified provirus, contigs are assembled *de novo* from the paired-end reads. Many platforms (*e.g.*, CLC Genomics Workbench[14]), allow the design of custom workflows for *de novo* assembly. Other open source software such as FastQC[15], Trimmomatic[16], Cutadapt[17], and FLASH[18] can also be utilized for processing reads, as well as tools such as Bowtie2[19] and SPAdes[20] for read mapping and *de novo* assembly. The steps for the *de novo* assembly of HIV-1 contigs using a specific commercial platform (see the **Table of Materials**) are outlined below (**Figure 2**). Customized workflow file is available upon request.

5.1. Import sequences and check quality: Import the paired sequence reads (in fastq.gz format) into the software, which will then be combined as single set of paired reads. Generate a sequence QC report and examine the quality of your data.

5.2. Quality control

5.2.1. Perform read trimming according to the QC report. Remove any adapter sequences and ambiguous nucleotides. Trim fifteen 5' and two 3' terminal nucleotides. Discard reads less than 50 nucleotides in length. Use a stringent quality limit of 0.001 corresponding to a QC phred score of 30.

5.3. Merging overlapping pairs

5.3.1. Form single extended reads by merging paired forward and reverse reads with overlapping regions.

5.4. *De novo* assembly

5.4.1. Use the native CLC genomics *de novo* assembler with a word (or k-mer) size of 30 nt and a maximum bubble size of 65 nt to assemble a random subsample of 10,000 non-overlapping paired reads. The expected coverage for the *de novo* assembled contig is ~200x for a ~9 kb sequence.

Note: This subsampling can reduce the computational burden such that most standard desktop computers can handle the assembly and analysis, and given the clonality of each provirus (one library is one provirus), this does not limit the diversity.

5.5. Re-mapping all reads to contigs

5.5.1. To obtain the final majority consensus sequence of each provirus, map the full read set to the *de novo* assembled contig. Accept only contigs with a minimum average coverage of 1,000X and ensure the final contig length corresponds to the size of the band on the original agarose gel (step 2.8.1). Save the final majority consensus sequence as a .fasta file.

Note: Most contigs can be assembled using the above steps. However, in some cases, for a single provirus, multiple contigs with a similar coverage are assembled. In these circumstances, contigs are aligned to a near full-length (~9 kb) consensus sequence from the same participant and scaffolded through manually assembly. All reads are then mapped to the manually assembled scaffold and the final consensus accepted if read coverage is even throughout the assembly and no single nucleotide polymorphisms (SNPs) > 40% are present.

5.6. Further quality control

5.6.1. To ensure each contig represents a single provirus, and is not due to the amplification of multiple proviruses within a single well, screen read coverage and variant calling of the final contig.

5.6.2. Multiple copy provirus templates present during PCR are often identified by very uneven read coverage (due to the co-amplification of proviruses of different sizes) when mapping to a full-length consensus from the same participant, or by the presence of SNPs with a frequency of > 40% (see representative results, **Figure 4**). Disregard mixed populations in subsequent analyses.

5.7. Alignment

5.7.1. Import the final consensus of each proviral sequence into sequence viewing software such as Molecular Evolutionary Genetics Analysis (MEGA) 7[21]. Align each sequence manually to the HXB2 reference sequence. Trim the 5' and 3' ends to positions 666-9650 of HXB2 to remove primer sequences.

5.7.2 Export the sequence list in fasta format and then align using MAFFT version 7[22], with manual editing where appropriate to obtain the final alignment.

Note: If any sequences do not align, first reverse complement the sequence. If the sequence still does not align perform a BLAST search to ensure the sequence is HIV-1. It is possible to amplify non-HIV-1 templates and these can be identified at this stage. If sequences are HIV-1 but do not align, consider the presence of inversions (see step 6.1).

6. Determining Potential Replication Competency of HIV-1 Proviral Sequences

Note: To identify sequences of genetically intact, and potentially replication-competent, HIV-1 proviruses a stringent process of elimination is followed (**Figure 1C**). Proviral sequences lacking inversions, large internal deletions, deleterious stop codons/ hypermutation, frameshift mutations, and/or defects in the MSD site or packing signal are considered genetically intact and potentially replication-competent.

6.1. Inversions

6.1.1. During the alignment stage, identify inversions. Inversions are regions where the sequence does not align to the reference HXB2 unless the region is reverse complemented.

Note: Depending on the application of the data, sequences containing inversions may need to be omitted from further analysis.

6.2. Large internal deletions

6.2.1. Identify contigs with large internal deletions (> 600 bp) in the alignment stage. Unless the deletion sits within *nef*, the sequence can be defined as defective.

Note: Any sequences with internal deletions < 600 bp will be identified by Gene Cutter (step 6.3.1) as having incomplete gene sequences.

6.3. Deleterious stop codons, frameshift mutations and deletions

6.3.1. Check all contigs of length > 8400 nucleotides for the presence of deleterious stop codons, frameshifts and incomplete gene sequences using the Los Alamos National Laboratory HIV Sequence Database Gene Cutter tool[23].

Note: The Gene Cutter tool divides the proviral sequence into the genes *gag, pol, vif, vpr, tat, rev, vpu, env,* and *nef*, and translates them to amino acids. Gene Cutter then screens for the presence of stop codons and frameshift mutations (due to insertions or deletions). Contigs containing stop codons or frameshifts in any gene, excluding *nef*[24], are classified as defective. Gene Cutter also identifies incomplete gene sequences and any proviruses with a deletion < 600 bp in a gene other than *nef* can be reclassified as defective due to a large internal deletion.

6.4. Hypermutation

6.4.1. Generate a consensus of the remaining full-length proviral sequences using the Los Alamos National Laboratory HIV Sequence Database Consensus Maker tool (simple consensus maker)[25]. Add the consensus sequence to the top of an alignment containing only the remaining full-length proviral sequences. Using this alignment and the Los Alamos National Laboratory HIV Sequence Database Hypermut tool[26] to identify APOBEC-induced G-A hypermutation in the remaining full-length proviral sequences.

6.5. Defects in the MSD and packaging signal

6.5.1. Inspect each of the remaining contigs for defects in the MSD and the packaging signal (HXB2 region 670-810) which render the proviral sequence defective[4]. Look for a point mutation in the MSD site (sequence GT, HXB2 744-745) or any deletion in the four stem loops of the packaging signal (SL1 (HXB2 691-734), SL2 (HXB2 736-754), SL3 (HXB2 766-779), and SL4 (HXB2 790-810)).

REPRESENTATIVE RESULTS:

The FLIPS assay amplifies and sequences single, near full-length HIV-1 proviruses. The protocol involves 6 steps to obtain near full-length proviral sequences. These steps include: lysis of infected cells, nested PCR of full-length (intact and defective) HIV-1 proviruses, DNA purification and quantification, sequencing library preparation, NGS, and *de novo* assembly of sequenced proviruses. The end of each step can be considered a checkpoint in which the quality of the product (*e.g.* amplified DNA, purified DNA, sequencing library or sequence) can be assessed prior to the next step. An overview of the assessment performed at the end of each step and the expected results is outlined below.

Following nested PCR, amplified products are run on a 1% agarose gel (**Figure 3**). The initial quality of the PCR can be determined by inspection of negative and positive controls. Negative control wells containing amplified product indicate contamination and positive control wells absent of amplified product indicate insufficient amplification. Next, wells containing amplified product are selected for sequencing. To avoid wells containing mixtures of multiple amplified proviruses, only wells containing amplified product run at end-point dilution are considered for sequencing. According to Poisson distribution, the end-point dilution is found when 30% of wells are positive for amplified product. At this dilution, there is an 80% chance these wells contain a single amplified provirus. Additionally, wells containing multiple amplified proviruses of different lengths can be visualized at this stage as multiple bands will appear on the gel. These wells should not be selected for sequencing (**Figure 3**).

Following purification of the amplified proviruses selected for sequencing, quantification ensures no proviral DNA is lost during the purification stage. If the DNA concentration of an amplified provirus is < 0.2 ng/µL, the remaining sample in the PCR2 plate can be purified. A similar checkpoint occurs following library preparation, in which each individual library is quantified. This ensures individual proviruses have been appropriately fragmented, tagged and amplified prior to sequencing. Individual proviral libraries are pooled in equimolar amounts to a final library

concentration of 4 nM (or as specified by the sequencing provider). If the concentration of an individual proviral library is too low, it can be excluded from the sequencing library pool, or the individual library prepared again. A final check of the concentration of the pooled library is performed prior to sequencing along with confirming the average fragment lengths.

Quality control steps before the *de novo* assembly stage ensures the quality of the reads used to assemble the final proviral contig. These steps include: removal of adapter sequences, trimming of 5' and 3' nucleotides, a stringent quality limit, and disregarding short reads. CLC Genomics Workbench can provide quality control reports that can be used before to assess the initial quality of the reads and guide trimming settings, and then after to determine if the trimming was sufficient to remove low quality regions. Additionally, for *de novo* assembly, the quality of the assembled contigs can be assessed for sufficient depth (> 1000X) and evenness of coverage (**Figure 4A**). Mixed populations can also be identified at this stage. Mixtures of multiple full-length (~9 kb) proviruses are identified through the presence of multiple SNPs with a frequency of greater than 40%, whereas mixtures of short (containing a large internal deletion) and full-length proviruses can be identified by uneven read coverage following mapping to a full-length reference from the same participant (**Figure 4B**).

Depending on the application, the final alignment can be visualized using tools such as ggtree available as a package in "R: A language and environment for statistical computing"[27]. In a recent study, FLIPS was utilized to sequence HIV-1 proviruses from naïve, central, transitional, and effector memory $CD4^+$ T cells isolated from individuals on long-term ART, with the aim to identify if particular cell subsets showed higher proportions of genetically intact and potentially replication-competent HIV-1[2]. Here, visual representation of the sequences isolated from one participant of this study (participant 2026) is presented (**Figure 5**). In this participant, the majority (97%) of sequences were defective, with intact sequences found in effector and transitional memory $CD4^+$ T cells. This visualization tool is useful for showing the number of sequences with large internal deletions and their position in the genome. It can be annotated further to indicate sequences with deleterious stop codons, frameshift mutations, and deletions/mutations in the MSD site and/or packaging signal.

One application of the FLIPS assay is the identification of genetically intact and potentially replication-competent HIV-1 proviruses. In a recent study of 531 sequences isolated from $CD4^+$ T cells from 6 participants on long-term ART, 26 (5%) genetically intact HIV-1 proviruses were identified[2]. The remaining defective proviruses included those with inversion sequences (6%), large internal deletions (68%), deleterious stop codons/hypermutation (9%), frameshift mutations (1%), and defects in the packaging signal and/or mutations in the MSD site (11%).

FIGURE AND TABLE LEGENDS:

Figure 1: Critical steps in the full-length individual proviral sequencing (FLIPS) assay. (**A**) HIV-1 DNA genome with primer binding sites in 5' and 3' U5 LTR regions used by FLIPS to amplify near full-length (defective and intact) HIV-1 proviruses via nested PCR. (**B**) Layout of a 96-well PCR plate containing 80 sample wells (20 wells for each dilution), 4 negative control wells, and 1

positive control well. (C) Process of elimination used to identify genetically intact, and potentially replication-competent, HIV-1 proviruses. This figure has been modified from Hiener *et al.*[2].

Figure 2: Overview of workflow for *de novo* assembly of HIV-1 proviruses. The major steps in the workflow include: 1) sequence read quality control, 2) merging overlapping pairs, 3) *de novo* assembly, and 4) remapping and consensus building have been colored red, blue, green and orange, respectively. This figure has been modified from Hiener *et al.*[2].

Figure 3: Example agarose gel of PCR amplified HIV-1 proviruses. Wells 1, 3, 6, and 9 contain amplified HIV-1 proviruses containing large internal deletions (size < 9 kb), well 2 contains co-amplification of two HIV-1 proviruses of different lengths (mixture), and well 12 contains positive control. Note the percent of wells containing amplified product is 60%, which is above the percentage required to isolate single templates.

Figure 4: Example output of read mapping. (**A**) Example demonstrating even coverage due to the amplification of a single full-length HIV-1 provirus. Following *de novo* assembly, all reads are mapped to the assembled contig to produce a consensus sequence. The software platform allows the mapped reads to be inspected for sufficient and even coverage. (**B**) Example demonstrating co-amplification of two HIV-1 proviruses of different lengths (mixture). To determine mixtures, reads are mapped to a full-length (~9 kb) reference sequence from the same participant and read mapping inspected. The presence of uneven coverage indicates a mixture. This figure is reproduced with permission from Qiagen[14].

Figure 5: Example visualization of HIV-1 proviral sequences isolated from CD4$^+$ T cell subsets from an individual on long-term ART. Individual HIV-1 proviral sequences are represented by horizontal lines. This figure has been modified from Hiener *et al.*[2].

Table 1: Reagents and volumes for PCR master mixes.

Table 2: Example layout of 96-well plate for quantification of dsDNA.

DISCUSSION:

The FLIPS assay is an efficient and high-throughput method for amplifying and sequencing single, near full-length HIV-1 proviruses. Multiple factors and critical steps in the protocol that influence the number and quality of the sequences obtained have been identified. Firstly, the number of cells and the HIV-1 infection frequency of the cell population influence the number of proviruses amplified. For example, in a previous publication, approximately half as many sequences were obtained from the same number of naïve CD4$^+$ T cells compared to effector memory CD4$^+$ T cells. This is because naïve cells typically have a lower infection frequency than effector memory cells[2]. Secondly, cell lysis is preferable to column-based extraction methods for obtaining genomic DNA as there is no risk of losing DNA in the extraction process. Lastly, as with any PCR-based assay, preventing contamination is critical. Separated clean areas should be designated for preparing master mixes, handling genomic DNA, adding positive controls, DNA purification and

quantification, and library preparation. This is particularly important for single-copy assays such as the one presented here.

Implementation of the FLIPS assay should first include running a positive control such as pNL4-3 plasmids rather than participant samples. This will allow for any troubleshooting prior to the use of HIV-1 positive cells, as the sequences obtained can be compared to available reference sequences for these plasmids. When using HIV-1 positive cells, it is important to consider the HIV-1 subtype (primers designed for FLIPS are specific to subtype B) and the infection frequency of the cell population if little to no proviruses are amplified. Primer sequences can be modified/redesigned to match other subtypes. Additionally, a well containing a positive control should be included in every PCR performed.

FLIPS has overcome the limitations of previous sequencing assays, including SPS. Through amplifying and sequencing near full-length HIV-1 proviruses, FLIPS can determine the potential replication-competency of HIV-1 proviruses. This was not possible using SPS, which sequenced only sub-genomic regions and therefore selected for sequences with intact primer binding sites. Furthermore, FLIPS overcomes the limitations associated with utilizing multiple amplification and sequencing primers, as was employed by previous full-length sequencing assays[1,4]. Through two rounds of PCR targeting the HIV-1 LTR regions combined with NGS, FLIPS decreases the number and complexity of primers required. FLIPS is therefore less susceptible to the consequences of primer mismatches, namely the erroneous identification of defective proviruses and an inability to amplify some proviruses within a population. The FLIPS protocol is also more efficient and allows for a higher throughput of sequencing than previous methods.

Evidently, FLIPS provides advantages over existing methods that determine the genetic composition of HIV-1 proviruses. However, it is important to acknowledge limitations of FLIPS. Firstly, the FLIPS assay has not been developed as a tool for measuring the size of the latent HIV-1 reservoir, as analyses to determine whether FLIPS amplifies every HIV-1 provirus present in a cell population have not been completed. FLIPS is instead useful for making relative comparisons of the composition of the reservoir between different cell populations[2]. Secondly, the replication-competency of intact HIV-1 proviruses cannot be determined with certainty without *in vivo* analyses, such as those performed by Ho *et al.*[4]. Thirdly, FLIPS is not designed to determine the integration site of HIV-1 proviruses.

Minor variations to the FLIPS protocol can increase its application. For example, changes in primer sequences can allow different and multiple HIV-1 subtypes to be amplified and sequenced. Sequencing of plasma HIV-1 virions is possible through the addition of cDNA synthesis prior to nested PCR. Future utilization of single molecule sequencing methods will eliminate the need for *de novo* assembly.

Genetic sequencing of integrated HIV-1 proviruses has increased our understanding of the latent HIV-1 reservoir. FLIPS is an important tool for future studies elucidating the composition and distribution of the latent reservoir. However, the application of FLIPS can extend beyond the reservoir. Future studies may utilize FLIPS to determine particular targets for CRISPR-Cas

technology, or assist in the identification of coding and non-coding regions which make the virus more responsive to latency reversing agents. Viral recombination may be better understood by looking at the junction sites of large internal deletions.

ACKNOWLEDGMENTS:

This work was supported the Delaney AIDS Research Enterprise (DARE) to Find a Cure (1U19AI096109 and 1UM1AI126611-01); an amfAR Research Consortium on HIV Eradication (ARCHE) Collaborative Research Grant from The Foundation for AIDS Research (amfAR 108074-50-RGRL); Australian Centre for HIV and Hepatitis Virology Research (ACH^2); UCSF-GIVI Center for AIDS Research (P30 AI027763); and the Australian National Health and Medical Research Council (AAP1061681). We would like to thank Dr. Joey Lai, Genomics Facility Manager at The Westmead Institute for Medical Research for his training in library preparation and use of his facility, and the Ramaciotti Centre for Genomics (University of New South Wales, Sydney, Australia) for conducting sequencing. We acknowledge with gratitude the participants who donated samples for this study.

DISCLOSURES:

The authors have nothing to disclose.

REFERENCES

1 Bruner, K. M., *et al.* Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nature Medicine.* **22** (9), 1043-1049 (2016).

2 Hiener, B., *et al.* Identification of Genetically Intact HIV-1 Proviruses in Specific CD4(+) T Cells from Effectively Treated Participants. *Cell Reports.* **21** (3), 813-822 (2017).

3 Abram, M. E., Ferris, A. L., Shao, W., Alvord, W. G., Hughes, S. H. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *Journal of Virology.* **84** (19), 9864-9878 (2010).

4 Ho, Y. C., *et al.* Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell.* **155** (3), 540-551 (2013).

5 Harris, R. S., *et al.* DNA deamination mediates innate immunity to retroviral infection. *Cell.* **113** (6), 803-809 (2003).

6 Lecossier, D., Bouchonnet, F., Clavel, F., Hance, A. J. Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science.* **300** (5622), 1112 (2003).

7 Chun, T. W., *et al.* Rebound of plasma viremia following cessation of antiretroviral therapy despite profoundly low levels of HIV reservoir: implications for eradication. *AIDS.* **24** (18), 2803-2808 (2010).

8 Chomont, N., *et al.* HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nature Medicine.* **15** (8), 893-900 (2009).

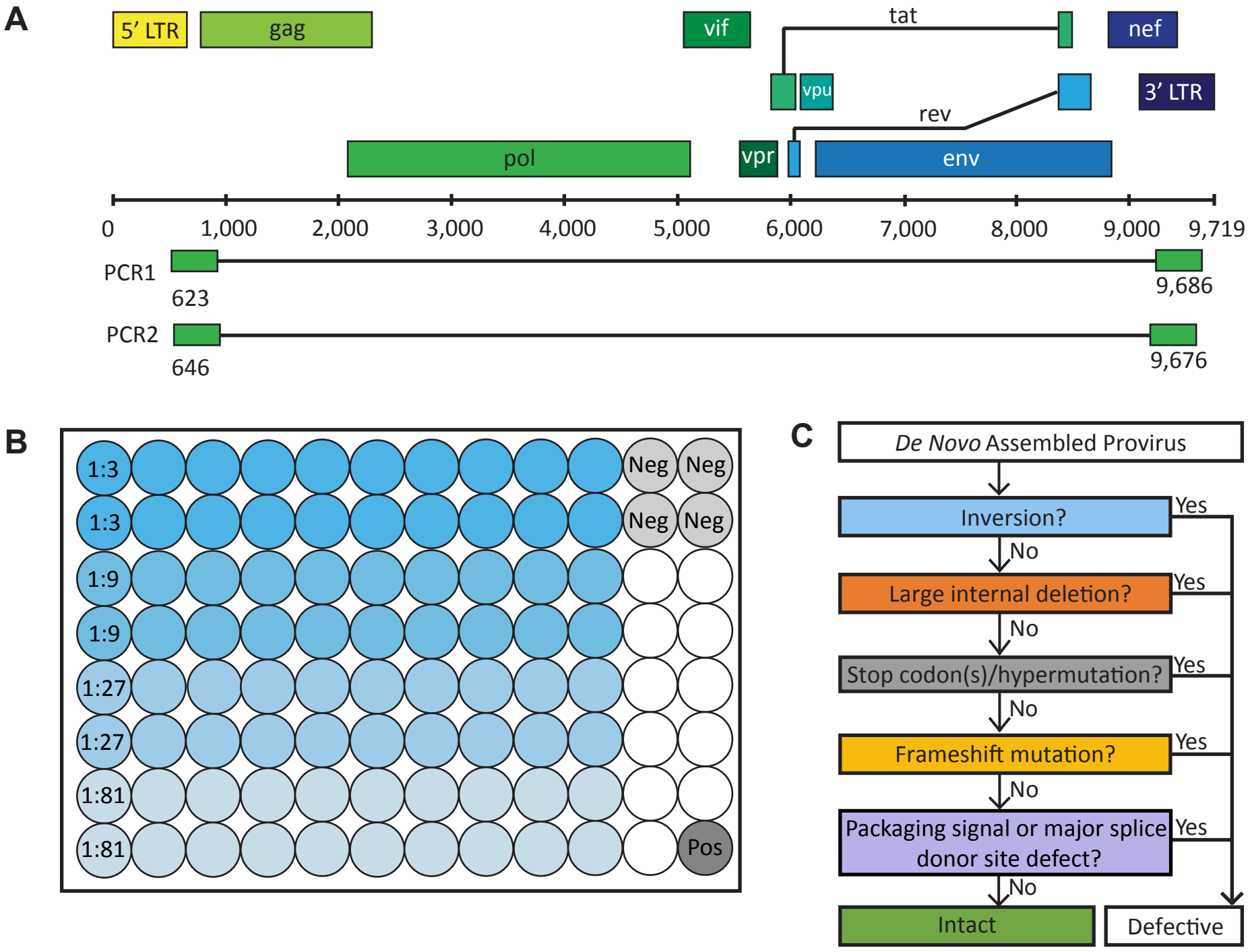
9 Soriano-Sarabia, N., *et al.* Quantitation of replication-competent HIV-1 in populations of resting CD4+ T cells. *Journal of Virology.* **88** (24), 14070-14077 (2014).

10 Josefsson, L., *et al.* The HIV-1 reservoir in eight patients on long-term suppressive antiretroviral therapy is stable with few genetic changes over time. *Proceedings of the National Academy of Sciences USA.* **110** (51), E4987-4996 (2013).

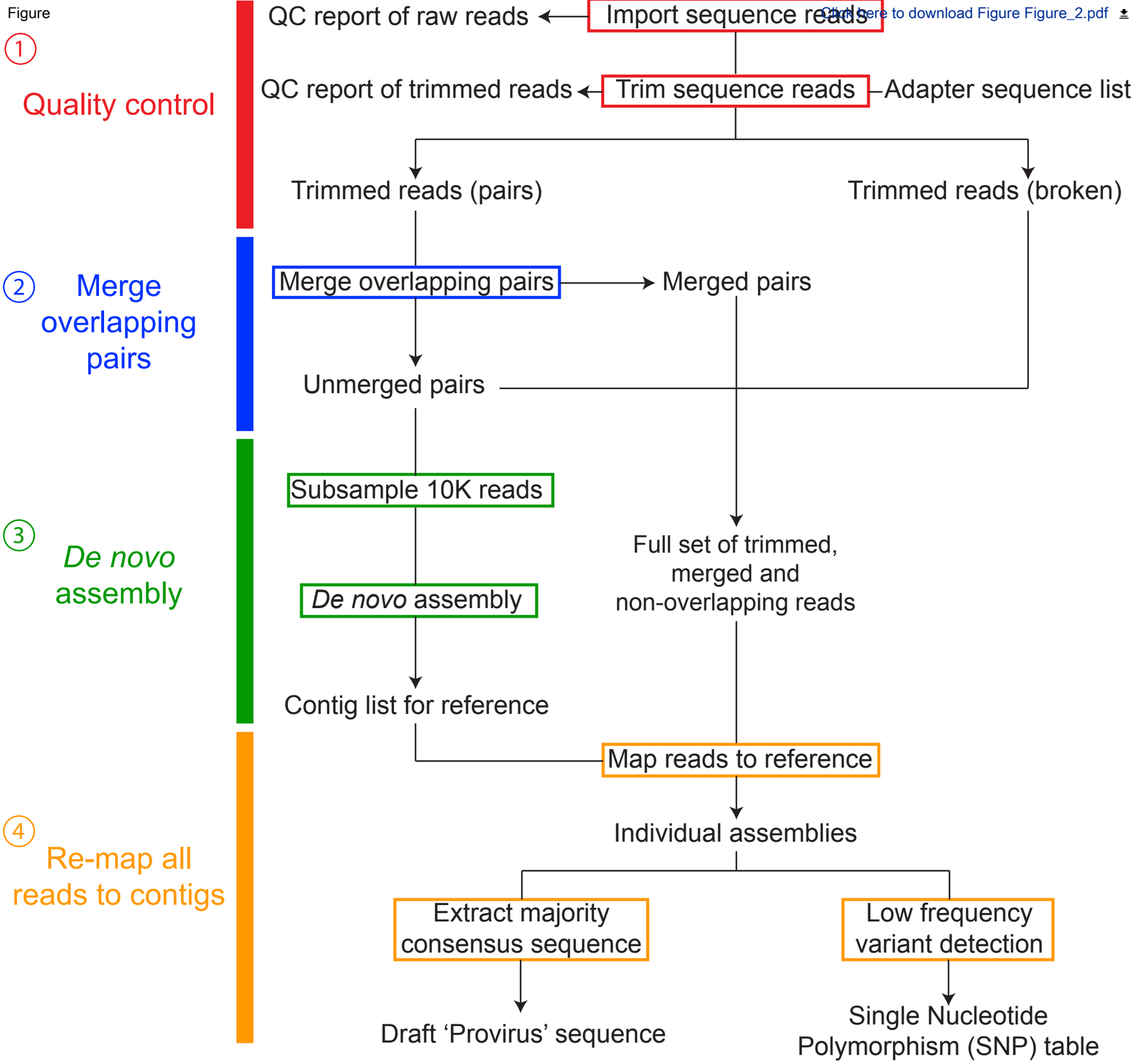
11 Evering, T. H., *et al.* Absence of HIV-1 evolution in the gut-associated lymphoid tissue from patients on combination antiviral therapy initiated during primary infection. *Public Library of Science Pathogens.* **8** (2), e1002506 (2012).
12 von Stockenstrom, S., *et al.* Longitudinal Genetic Characterization Reveals That Cell Proliferation Maintains a Persistent HIV Type 1 DNA Pool During Effective HIV Therapy. *Journal of Infectious Diseases.* **212** (4), 596-607 (2015).
13 Palmer, S., *et al.* Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *Journal of Clinical Microbiology.* **43** (1), 406-413 (2005).
14 Qiagen. *CLC Genomics Version 10*, <https://www.qiagenbioinformatics.com/products/clc-genomics-workbench> (2018).
15 Andrews, S. *FastQC: a quality control tool for high throughput sequence data.*, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).
16 Bolger, A. M., Lohse, M., Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* **30** (15), 2114-2120 (2014).
17 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal.* **17** (1), (2011).
18 Magoc, T., Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics.* **27** (21), 2957-2963 (2011).
19 Langmead, B., Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods.* **9** (4), 357-359 (2012).
20 Bankevich, A., *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology.* **19** (5), 455-477 (2012).
21 Kumar, S., Stecher, G., Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution.* **33** (7), 1870-1874 (2016).
22 Katoh, K., Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution.* **30** (4), 772-780 (2013).
23 Laboratory, L. A. N. *HIV Sequence Database - Gene Cutter tool*, <https://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html> (2018).
24 Foster, J. L., Garcia, J. V. Role of Nef in HIV-1 replication and pathogenesis. *Advances in Pharmacology.* **55** 389-409 (2007).
25 Laboratory, L. A. N. *HIV Sequence Database - Consensus Maker tool*, <HIV Sequence Database - Gene Cutter tool> (2018).
26 Laboratory, L. A. N. *HIV Sequence Database - Hypermut tool*, <https://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html> (2018).
27 Yu, G., Smith, D. K., Zhu, H., Guan, Y., Lam, T. T. Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution.* (2016).

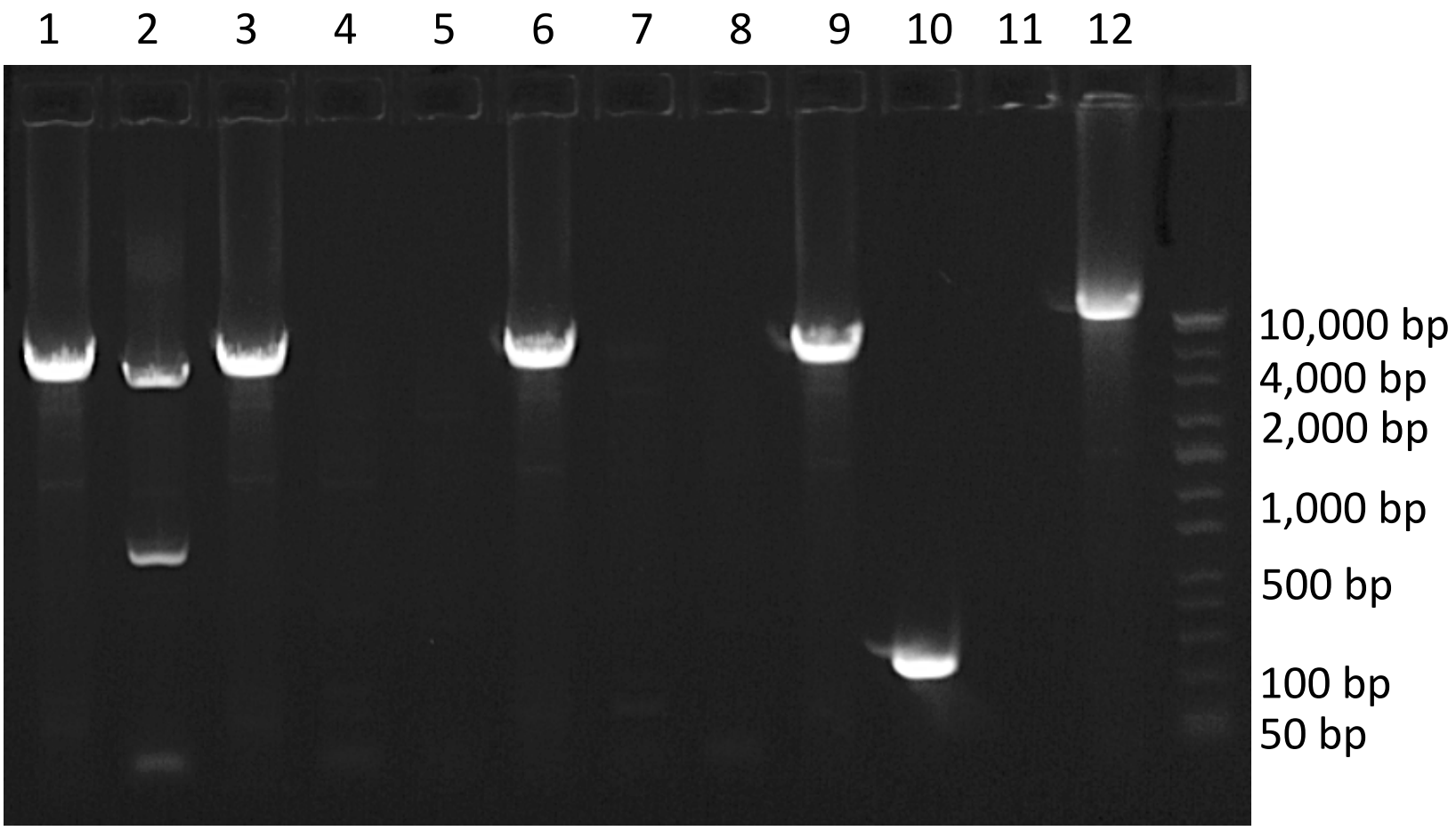
Figure

Click here to download Figure Figure_1.pdf



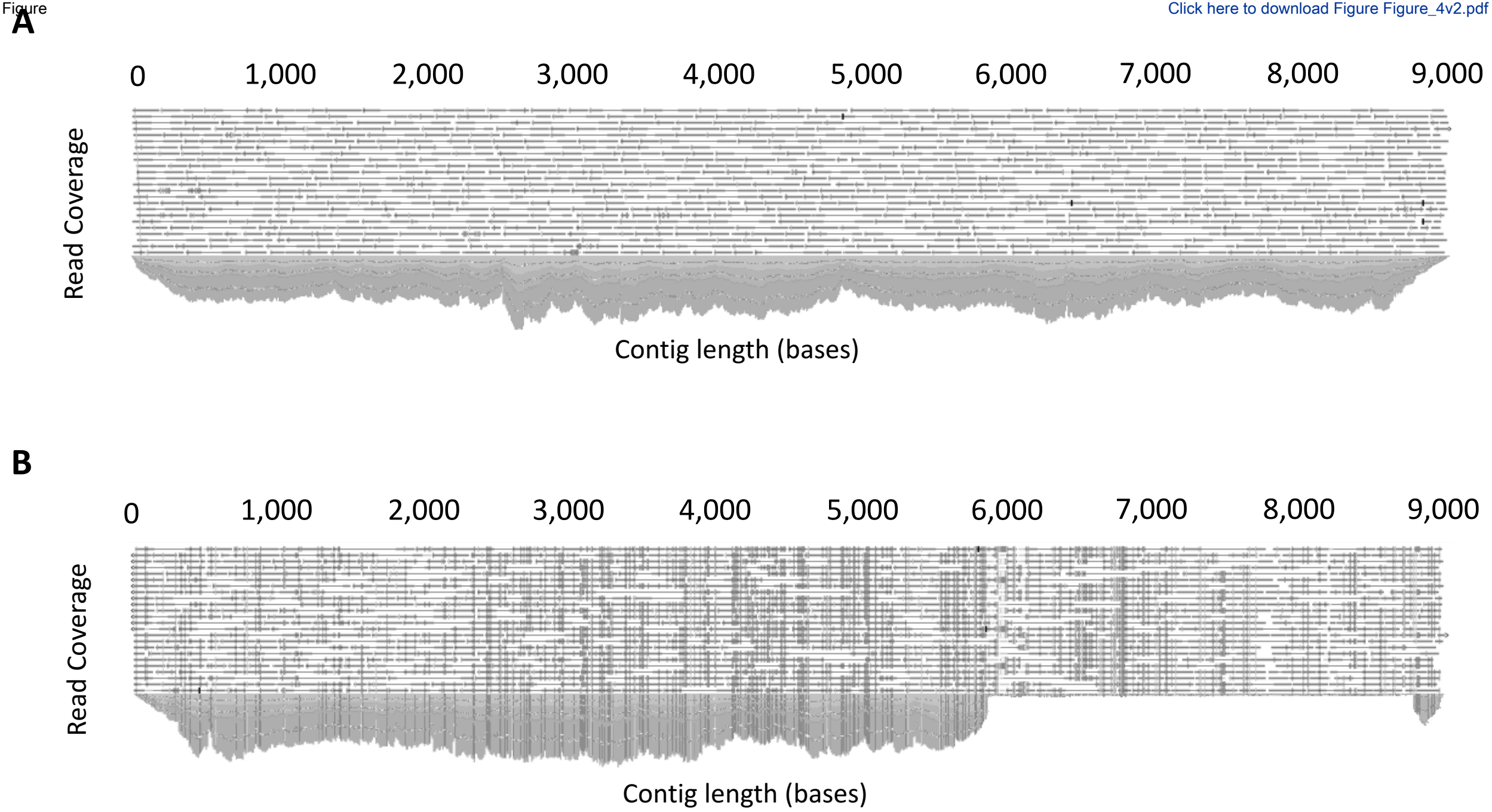
Figure




1
2
3
4
5
6
7
8
9
10
11
12
10,000 bp
4,000 bp
2,000 bp
1,000 bp
500 bp
100 bp
50 bp

Figure

Click here to download Figure Figure_4v2.pdf



Figure

Click here to download Figure Figure 5.pdf

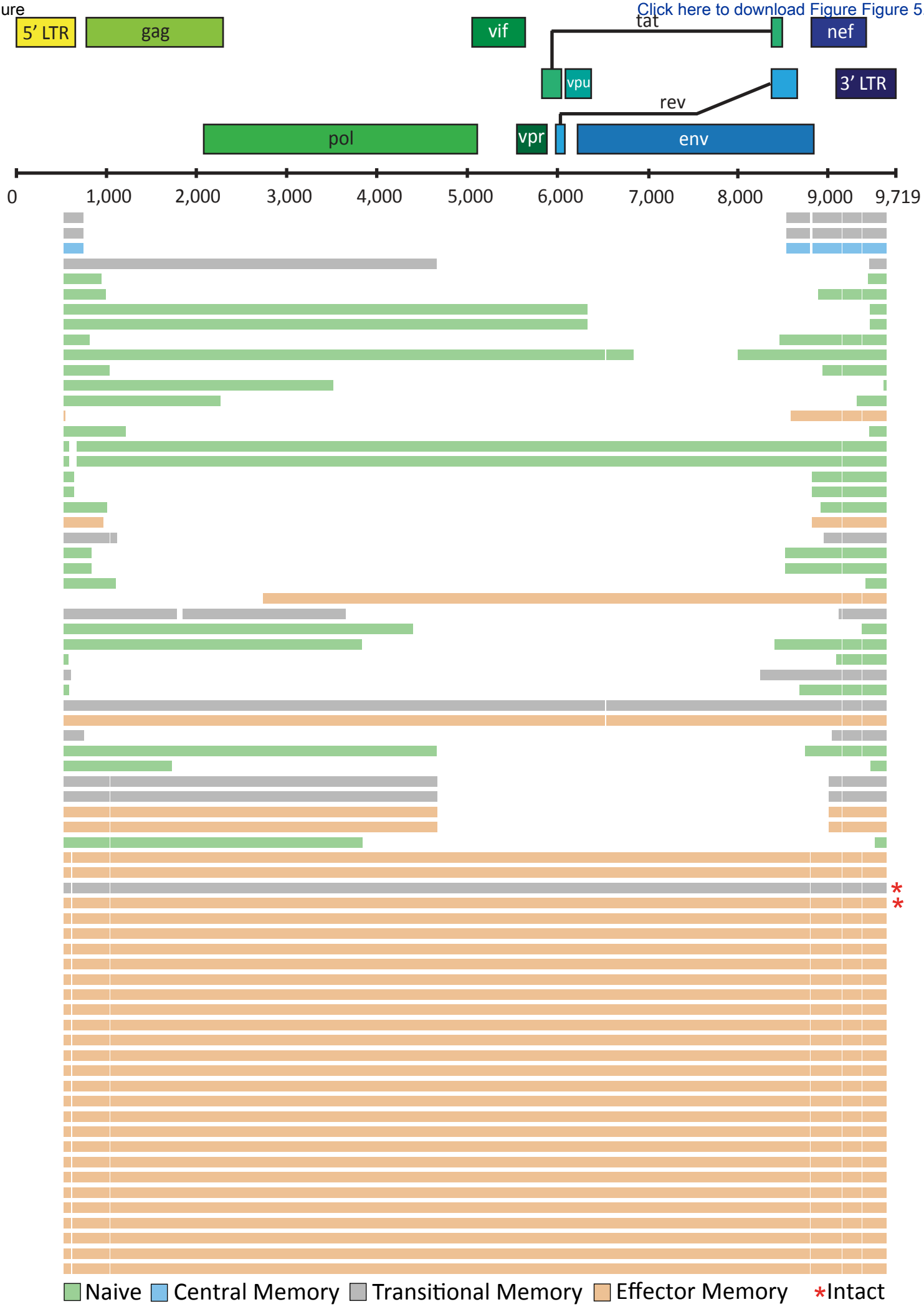


Table 1

Reagent	Final Concentration
Forward Primer	1 μM
Reverse Primer	1 μM
Buffer (10 X)	1 X
$MgSO_4$ (50 mM)	2 mM
dNTP (10 mM)	0.2 mM
DNA polymerase (5 U/μL)	0.025 U/μL
Ultrapure H_2O	

Volume for PCR1 plate (L)	Volume for PCR2 plate (L)
32.3	23.8
32.3	23.8
323	238
129.2	95.2
64.6	47.6
16.2	11.9
2632.5	1939.7

Table 2

	1	2	3	4	5	6
A	Standard (1 ng/μL) 100 μL	Standard (0.1 ng/μL) 100 μL	Standard (0.01 ng/μL) 100 μL	Standard (0.001 ng/μL) 100 μL	Blank 100 μL buffer	
B	Standard (1 ng/μL) 100 μL	Standard (0.1 ng/μL) 100 μL	Standard (0.01 ng/μL) 100 μL	Standard (0.001 ng/μL) 100 μL	Blank 100 μL buffer	
C	Standard (1 ng/μL) 100 μL	Standard (0.1 ng/μL) 100 μL	Standard (0.01 ng/μL) 100 μL	Standard (0.001 ng/μL) 100 μL	Blank 100 μL buffer	
D	Sample 1: 1 μL DNA + 99 μL buffer	Sample 2: 1 μL DNA + 99 μL buffer	Sample 3: 1 μL DNA + 99 μL buffer	Sample 4: 1 μL DNA + 99 μL buffer	Sample 5: 1 μL DNA + 99 μL buffer	Sample 6: 1 μL DNA + 99 μL buffer
E	Sample 1: 1 μL DNA + 99 μL buffer	Sample 2: 1 μL DNA + 99 μL buffer	Sample 3: 1 μL DNA + 99 μL buffer	Sample 4: 1 μL DNA + 99 μL buffer	Sample 5: 1 μL DNA + 99 μL buffer	Sample 6: 1 μL DNA + 99 μL buffer
F	Sample 1: 1 μL DNA + 99 μL buffer	Sample 2: 1 μL DNA + 99 μL buffer	Sample 3: 1 μL DNA + 99 μL buffer	Sample 4: 1 μL DNA + 99 μL buffer	Sample 5: 1 μL DNA + 99 μL buffer	Sample 6: 1 μL DNA + 99 μL buffer
G						
H						

7	8	9	10	11	12
Sample 7: 1 μL DNA + 99 μL buffer	Sample 8: 1 μL DNA + 99 μL buffer	Sample 9: 1 μL DNA + 99 μL buffer	Sample 10: 1 μL DNA + 99 μL buffer	Sample 11: 1 μL DNA + 99 μL buffer	Sample 12: 1 μL DNA + 99 μL buffer
Sample 7: 1 μL DNA + 99 μL buffer	Sample 8: 1 μL DNA + 99 μL buffer	Sample 9: 1 μL DNA + 99 μL buffer	Sample 10: 1 μL DNA + 99 μL buffer	Sample 11: 1 μL DNA + 99 μL buffer	Sample 12: 1 μL DNA + 99 μL buffer
Sample 7: 1 μL DNA + 99 μL buffer	Sample 8: 1 μL DNA + 99 μL buffer	Sample 9: 1 μL DNA + 99 μL buffer	Sample 10: 1 μL DNA + 99 μL buffer	Sample 11: 1 μL DNA + 99 μL buffer	Sample 12: 1 μL DNA + 99 μL buffer

Table of Materials

Click here to download Table of Materials Reagents for FLIPS.xlsx

Material/Equipment	Company	Catalogue Number
UltraPure 1 M Tris-HCl, pH 8.0	Invitrogen	15568025
Nonidet P 40 Substitute solution	Sigma	98379
Tween-20	Sigma	P9416
Proteinase K Solution (20 mg/mL)	Promega	AM2548
96 well thermocycler		
Platinum Taq DNA Polymerase High Fidelity	Invitrogen	11304011
10X High Fidelity Buffer [600 mM Tris-SO_4 (pH 8.9), 180 mM $(NH_4)_2SO_4$]	Invitrogen	11304011
50 mM $MgSO_4$	Invitrogen	11304011
PCR Nucleotide Mix, 10 mM	Promega	C1141
Ultrapure H_2O	Invitrogen	10977023
PCR1 and PCR2 Primers		
PCR Plate 96 Well, Half Skirt, Single Notch Corner, Clear	Axygen	PCR-96M2-HS-C
Microseal 'B' Adhesive Seals	BioRad	MSB1001
HIV-1 NL4-3 Infectious Molecular Clone (pNL4-3)	The following reagent was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: HIV-1 NL4-3 Infectious Molecular Clone (pNL4-3) from Dr. Malcolm Martin (Cat# 114).	
PCR plate spinner or benchtop centrifuge		
E-GEL 48 1% Agarose	Invitrogen	G800801
DirectLoad Wide Range DNA Marker	Sigma	D7058
Mother E-Base device	Invitrogen	EBM03

Gel Doc EZ Gel Documentation System	BioRad	1708270
PlateMax Peelable Heat Sealing	Axygen	HS-200
96 Well 0.8 mL Storage Plate	ThermoFisher Scientific	AB0765
Agencourt AMPure XP (PCR purification kit)	Beckman Coulter	A63880
Ethyl alcohol, Pure. 200 proof, for molecular biology	Sigma	E7023
Magnetic Stand-96	Invitrogen	AM10027
Microplate shaker		
Buffer EB	Qiagen	19086
Quant-iT PicoGreen dsDNA Assay Kit	Invitrogen	P11496
Microplate reader		
Nextera XT DNA Library Preparation Kit	Illumina	FC-131-1096
Nextera XT Index Kit	Illumina	FC-131-2001
Hard-Shell 96-Well PCR Plates	Biorad	HSP9601
Library Quantification Kit - Illumina/Universal	Kapa Biosystems	KK4824
2100 Bioanalyzer	Agilent Technology	
High Sensistivty DNA kit	Agilent Technology	5067-4626
Illumina MiSeq Platform	Illumina	

Comments/Description
Dilute to 5 mM for nested PCR
Any 96 well thermocycler can be used
Desalted. Dilute to (100 M) with H_2O
Required to seal 96 well PCR plates
Diluted to 10^5 copies/mL and used as positive control
Precast 1% agarose gels (two gels used per 96 well plate). Any 1% agarose gel can be substituted. Contains ethidium
Any ladder with a range up to 10 kb can be substituted
Required for running precast 48 well 1% agarose gels

Any visualisation system for ethidium bromide containing agarose gels can be used
Heat sealing film for long term storage
0.8 mL 96 well plate required for magnetic bead based PCR purification
Magnetic bead based PCR purification kit. Other PCR purification kits can be substitued here (e.g. QIAquick PCR Purification Kit (Qiagen Cat#28106)
Optional
Elution buffer
A fluorescence based stain for measuring dsDNA concentration
Required for Nextera XT DNA library preparation kit
Other library quantification kits can be substitued (e.g. JetSeq DNA Library Quantification Lo-Rox Kit (Bioline Cat#BIO-68029)
Automated electrophoresis system . Use in conjunction with a High Sensistivity DNA kit

1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article: Full-Length Individual Proviral Sequencing for the Identification of Intact HIV-1 Genomes

Author(s): Bonnie Hiener, John-Sebastian Eden, Bethany A Horsburgh and Sarah Palmer

Item 1 (check one box): The Author elects to have the Materials be made available (as described at http://www.jove.com/author) via: ☐ Standard Access ☒ Open Access

Item 2 (check one box):

☒ The Author is NOT a United States government employee.

☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. Defined Terms. As used in this Article and Video License Agreement, the following terms shall have the following meanings: "**Agreement**" means this Article and Video License Agreement; "**Article**" means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; "**Author**" means the author who is a signatory to this Agreement; "**Collective Work**" means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; "**CRC License**" means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode; "**Derivative Work**" means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; "**Institution**" means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; "**JoVE**" means MyJove Corporation, a Massachusetts corporation and the publisher of *The Journal of Visualized Experiments;* "**Materials**" means the Article and / or the Video; "**Parties**" means the Author and JoVE; "**Video**" means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. Background. The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. Grant of Rights in Article. In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4** and **7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

4. Retention of Rights in Article. Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. Grant of Rights in Video – Standard Access. This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. Grant of Rights in Video – Open Access. This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this Section 6 is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. Government Employees. If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. Likeness, Privacy, Personality. The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

9. Author Warranties. The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

10. JoVE Discretion. If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have

jove
JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

11. Indemnification. The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in *JoVE* or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

12. Fees. To cover the cost incurred for publication, JoVE must receive payment before production and publication the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

13. Transfer, Governing Law. This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to me one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement required per submission.

CORRESPONDING AUTHOR:

Name:	Bonnie Hiener
Department:	Centre for Virus Research
Institution:	The Westmead Institute for Medical Research
Article Title:	Full-Length Individual Proviral Sequencing for the Identification of Intact HIV-1 Genomes
Signature:	BHiener
Date:	15/02/18

Please submit a signed and dated copy of this license by one of the following three methods:

1) Upload a scanned copy of the document as a pfd on the JoVE submission site;
2) Fax the document to +1.866.381.2236;
3) Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02139

For questions, please email submissions@jove.com or call +1.617.945.9051

20 May 2018

To the Editors of JoVE

The manuscript entitled “Full-Length Individual Proviral Sequencing for the Identification of Intact HIV-1 Genomes” (JoVE58016) was recently reviewed for publication in JoVE. We greatly appreciate the constructive critique from the editor and reviewers as this has improved the quality of our manuscript. We have carefully addressed each point raised and have revised the manuscript accordingly.

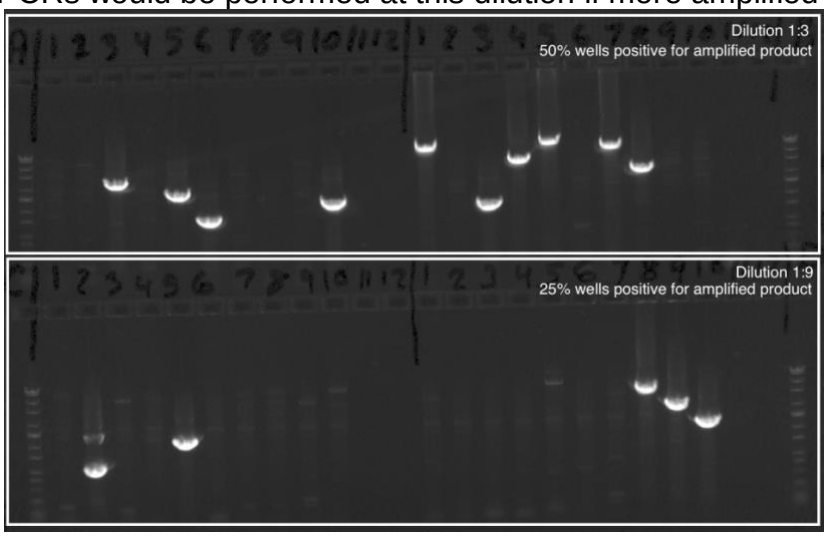
Editor’s and reviewer’s comments are in bold and are addressed point by point below:

Changes recommended by the JoVE Scientific Review Editor:

Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammatical errors.

Protocol Language: Please split up any long steps (e.g. 6.3).
Long steps, including 6.3 have been split into shorter steps

Protocol Detail: Please note that your protocol will be used to generate the script for the video, and must contain everything that you would like shown in the video. Please add more specific details (e.g. button clicks for software actions, numerical values for settings, etc) your protocol steps. There should be enough detail in each step to supplement the actions seen in the video so that viewers can easily replicate the protocol. For steps that will not be filmed, it may be sufficient to cite references where appropriate. Some examples:
1) 2.9: Unclear what we can show here, please clarify.
It is important to emphasize that only wells containing amplified product obtained at endpoint dilution are sent for sequencing. Here, we could show an example gel (below), where the genomic DNA has been diluted 1:3 and 1:9. The wells containing DNA diluted 1:3 are 50% positive for amplified product, whereas the wells containing DNA diluted 1:9 contain 25% amplified product. In this case, only the amplified products obtained at the 1:9 dilution would be sequenced. Additionally, all subsequent PCRs would be performed at this dilution if more amplified proviruses were required.



2) 3.1: Mention spin speed. How is recording done? Using what instrument?
Spin speed and duration are not important. The aim of this step is to pull down any residual liquid from the sides of the wells. As such, the manuscript has been changed to read, “Briefly spin the PCR plate in a PCR plate spinner or centrifuge to pull down any residual contents from the sides of the wells” in all relevant sections.
Additionally, the manuscript has been updated to read, “Write down the original and new well position of each amplified product in a spreadsheet as a record of each amplified product to be sequenced.”

3) 4.1, 4.2, 4.3: Unclear what is done here. In order to film this, all actions must be described explicitly, it is not sufficient to ask the reader to follow manufacturers’ instructions for filmed steps.
This part of the protocol involves following the Illumina Nextera XT Library Preparation protocol very closely. Illumina has extensive instructions on these steps which are addressed in far more detail than we could include here. As such, we believe this should be removed from the filmed steps.

4) 4.3: How is the fragment length determined? Please elaborate.
The fragment lengths are determined by running a small aliquot of the pooled libraries on an Agilent bioanalyzer or similar instrument. This is important for determining the final molarity of the pooled libraries. The manuscript has been updated to read:
“4.3. Quantify the final pooled library using the same dsDNA quantification kit used in step 3.4. Follow the manufacturer’s instructions. Determine the average fragment lengths by running 1 μL of the pooled library on an automated electrophoresis system using an appropriate kit (see table of materials). Use the concentration and average fragment lengths to determine the molarity of the pooled library.”

Protocol Highlight: After you have made all of the recommended changes to your protocol (listed above), please re-evaluate the length of your protocol section. There is a 10-page limit for the protocol text, and a 3- page limit for filmable content. If your protocol is longer than 3 pages, please highlight ~2.5 pages or less of text (which includes headings and spaces) in yellow, to identify which steps should be visualized to tell the most cohesive story of your protocol steps. Remember that the non-highlighted protocol steps will remain in the manuscript and therefore will still be available to the reader.
1) The highlighting must include all relevant details that are required to perform the step. For example, if step 2.5 is highlighted for filming and the details of how to perform the step are given in steps 2.5.1 and 2.5.2, then the sub-steps where the details are provided must be included in the highlighting.
2) The highlighted steps should form a cohesive narrative, that is, there must be a logical flow from one highlighted step to the next.
3) Please highlight complete sentences (not parts of sentences). Include sub-headings and spaces when calculating the final highlighted length.
4) Notes cannot be filmed and should be excluded from highlighting.
5) Please bear in mind that software steps without a graphical user interface/calculations/ command line scripting cannot be filmed.

Discussion: JoVE articles are focused on the methods and the protocol, thus the discussion should be similarly focused. Please ensure that the discussion covers the following in detail and in paragraph form: 1) modifications and

troubleshooting, 2) limitations of the technique, 3) significance with respect to existing methods, 4) future applications and 5) critical steps within the protocol.
The discussion has been written to the above criteria.

References: Please move the in-text http weblinks (Lines 239-244) into the reference list, and use superscripted citations.
All in-text weblinks have been removed and replaced by superscripted citations.

Commercial Language: JoVE is unable to publish manuscripts containing commercial sounding language, including trademark or registered trademark symbols (TM/R) and the mention of company brand names before an instrument or reagent. Examples of commercial sounding language in your manuscript are Taq HiFi Pol, microseal B, Nextera XT, Kapa Illumina Universal, Quant-iT PicoGreen, Bioanalyzer, TapeStation, Illumina MiSeq, clc genomics workbench, etc.
1) Please use MS Word's find function (Ctrl+F), to locate and replace all commercial sounding language in your manuscript with generic names that are not company-specific. All commercial products should be sufficiently referenced in the table of materials/reagents. You may use the generic term followed by "(see table of materials)" to draw the readers' attention to specific commercial names.
2) Please check Table 1 as well.
All commercial language has been removed from both the manuscript and Table 1. However, 'CLC Genomics' has remained and instead referenced.

Table of Materials: Please revise the table of the essential supplies, reagents, and equipment. The table should include the name, company, and catalog number of all relevant materials/software in separate columns in an xls/xlsx file. Please include items such as all instruments used.
Table is complete.

Please define all abbreviations at first use.
Please use standard abbreviations and symbols for SI Units such as µL, mL, L, etc., and abbreviations for non-SI units such as h, min, s for time units. Please use a single space between the numerical value and unit.
These have been checked.

If your figures and tables are original and not published previously or you have already obtained figure permissions, please ignore this comment. If you are re-using figures from a previous publication, you must obtain explicit permission to re-use the figure from the previous publisher (this can be in the form of a letter from an editor or a link to the editorial policies that allows you to re-publish the figure). Please upload the text of the re-print permission (may be copied and pasted from an email/website) as a Word document to the Editorial Manager site in the "Supplemental files (as requested by JoVE)" section. Please also cite the figure appropriately in the figure legend, i.e. "This figure has been modified from [citation]."

Comments from Peer-Reviewers:

Reviewer #1:

Major Concerns:

Lines 361 and 446: it is mandatory to introduce a positive control (diluted plasmid) in the experiment

A well containing a positive control (pWT/BaL plasmid) is included in every PCR. The lines the reviewer is referring to explains the situation where a lab may be implementing the FLIPS assay for the first time. In this situation they may wish to use a positive control to ensure the assay is working, before using cells isolated from HIV-1 positive individuals. As such, the paragraph has been changed to the following:

"Implementation of the FLIPS assay should first include running a positive control such as pWT/BaL or pNL4-3 plasmids rather than participant samples. This will allow for any trouble shooting prior to the use of HIV-1 positive cells as the sequences obtained can be compared to available reference sequences for these plasmids. When using HIV-1 positive cells it is important to consider the HIV-1 subtype (primers designed for FLIPS are specific to subtype B) and the infection frequency of the cell population if little to no proviruses are amplified. Primer sequences can be modified/redesigned to match other subtypes. Additionally, a well containing a positive control should be included in every PCR performed."

Minor Concerns:

1 Lines 137-139 and 178-181: it is not clear if the samples are diluted several times to end point dilution

The first PCR is used to determine the end point dilution. The volume of DNA prepared is only enough for this PCR. Once the endpoint dilution has been determined, the DNA can be freshly diluted to endpoint. This is performed in this way as the amount of DNA is usually limited.

The manuscript has been changed to make this clearer:

"2.2. Serially dilute genomic DNA from 1:3 to 1:81 with Tris-HCl (5 mM, pH 8), **preparing 45 μL for each dilution (enough for 20 wells for each dilution)."**

"2.9. Determine the dilution at which no more than 30% of wells are positive for amplified product. This is the end-point dilution in which the majority (80%) of wells contain amplified product from a single template. This dilution should be **prepared and** used in subsequent PCRs to obtain further proviral amplicons. Record the approximate size of each amplified product."

Figure 3: The quality of the picture is quite bad and we do not see any sample with an amplification of 10.000 bp (there is only one positive control)

There is no amplification at 10 kb as the maximum we sequence is approximately 9 kb. There are 4 amplicons with a length of approximately 9 kb. There is only one positive control as we only include one positive control per PCR plate to minimize contamination to the other wells (as the positive control is a concentrated plasmid). I have tried to improve the quality of the image.

Figure 5 It seems that this Figure is not introduced in the text (should be at line 380?)

Figure 5 reference added.

It is surprising that there is no DMSO reagents

No DMSO reagents were used.

Table 1 : Volumes : the units are not noted
All units are in µL and this is specified in the table.

Reviewer #2:
Manuscript Summary:
The manuscript describes a methodology to obtain near full length sequence of HIV-1 provirus using NGS. Overall the manuscript is cogent and I have only few minor points.

Minor Concerns:
1. For the benefit of readers who are unfamiliar with in-silico tools, this reviewer suggests the authors to elaborate section 5.6 with an example sequences.
2.This reviewer suggests elaborating " such as To ensure each contig represents a single provirus and is not due to the amplification of multiple proviruses within a single well, screen read coverage and variant calling of the final contig" etc.
A reference to representative results and Figure 4 has been added to section 5.6 to allow readers to refer to the section of the manuscript where this issue is further explained.
“Additionally, for de novo assembly, the quality of the assembled contigs can be assessed for sufficient depth (>1000X) and evenness of coverage (Figure 4A). Mixed populations can also be identified at this stage. Mixtures of multiple full-length (~9 kb) proviruses are identified through the presence of multiple SNPs with a frequency of >40%, whereas mixtures of short (containing a large internal deletion) and full-length proviruses can be identified by uneven read coverage following mapping to a full-length reference from the same participant (Figure 4B).”

Reviewer #3:
Characterizing the proviral genome, especially the intact proviral reservoir, has become an important means of assessing the "true" HIV reservoir - i.e., that which can potentially lead to productive HIV infection. In this manuscript, Dr. Hiener and colleagues describe their method for the amplification, sequencing, and analysis of defective and intact proviruses. The manuscript is well-written and provides an important protocol for researchers in the field. For those who want to replicate this protocol, I believe that the biggest hurdle would be in the bioinformatic analysis for the determination of intact or defective proviruses. If the authors would be willing to share this software, it would be a great value to the community.
For those who would like to use CLC genomics, a line has been added to section 5:
“Customized workflow available upon request.”

Other comments:
-Line 58 - would be careful making such a confident statement that 5% of proviruses are replication-competent. The exact proportion of intact proviruses is still up for debate, may vary (e.g., timing of ART), and it's not been proven how many of these apparently intact proviruses really are replication-competent.
This section has been reworded to:

“Two recent studies have found that approximately 5% of HIV-1 proviruses isolated from individuals on long-term ART are genetically intact, and potentially replication competent, and may contribute to the rapid rebound in HIV-1 plasma levels upon cessation of ART.”

-Line 267-268 - Given these are sequences from single-genome amplified products, is it really necessary to have contigs with a minimum of 1000x coverage? Do you have evidence that fewer reads are less reliable? There are studies suggesting that a few hundred reads may be sufficient, even from non-SGA samples [e.g., Henn et al, PLoS Path 2012].
Our minimum coverage is used primarily as a minimum ‘expected’ coverage to limit the number of very short , low coverage contigs, which are artifacts often produced during any assembly. This ensures we generally only return the single provirus from the initial assembly.

-Line 274-275 - If these are from SGA samples, why accept sequences with such high SNPs (up to 40%)?
The SNP cut-off becomes relevant when it is unclear whether a sequence is due to the amplification of >1 full-length (~9 kb) provirus within a single well. This is because the co-amplification of proviruses of different sizes can clearly be seen in the distribution of read coverage across the length of the contig generated. We chose 40% as the SNP cut off as in this situation it is difficult to confidently call the identity of the base at that site.

It is important to emphasize that the aim of FLIPS is to identify which proviruses are genetically intact and hence potentially replication-competent. The majority of proviruses we have sequenced are defective because of an internal deletion. We use a strict process of elimination where after identifying proviruses with large internal deletions we are left with a small population of proviruses that are near full-length. These proviruses are then run through Gene Cutter to identify those which have deleterious stop codons, incomplete codons and frameshift mutations. The majority of these usually have multiple stop codons in multiple genes. If a situation arose where a provirus was deemed defective due to the presence of a single base mutation, then we would need to consider whether this occurs at the same position as a SNP.

Line 306-307 - Could the authors describe the rationale for using the 100bp cut-off for large internal deletions?
Upon reflection we believe a more stringent approach to identifying sequences with internal deletions would first be to have a cut-off of 600 bp that immediately identifies defective proviruses. 600 bp was chosen as this is the approximate length of *nef.* Nef is generally considered a non-essential gene. The remaining sequences can be run through Gene Cutter which will identify those which contain incomplete gene sequences which are <600 bp. Therefore, the manuscript has been changed to the following:
“ 6.3. Deleterious stop codons, frameshift mutations, and deletions: Check all contigs of length >8400 nucleotides for the presence of deleterious stop codons, frameshifts and incomplete gene sequences using the Los Alamos National Laboratory HIV Sequence Database Gene Cutter tool. Gene Cutter divides the contigs into the open reading frames (gag to nef) and translates them to amino acids.

6.3.1. Any contigs containing a stop codon in any gene, excluding nef, are defined as defective. Contigs containing nucleotide deletions or insertions causing frameshift mutations are also identified as defective.

6.3.2. Gene Cutter also identifies incomplete gene sequences and any proviruses with a deletion <600 bp in a gene other than nef can be reclassified as defective due to a large internal deletion."

Lines 320-324 - Can the authors provide exact details on the criteria for the MSD/packaging signal defects. To truly replicate the results and to do this assay well, the interpretation and bioinformatic analysis of the sequences is potentially even more important than the PCR methods themselves.
The exact HXB2 coordinates of the four stem loops of the packaging signal and the MSD site have been added to the manuscript. The manuscript reads as follows:
"6.5. Defects in the MSD and packaging signal: Manually inspect the remaining contigs for defects in the MSD and packaging signal (HXB2 region 670-810). These sites are essential to HIV-1 replication, as proviruses with a deletion in the four stem loops: SL1 (HXB2 691-734), SL2 (HXB2 736-754), SL3 (HXB2 766-779) and SL4 (HXB2 790-810)) of the packaging signal, or a point mutation in the MSD (sequence GT, HXB2 744-745) have previously been identified as replication-incompetent"

Line 344-346 - I expect that some wells may contain bands also due to non-specific amplification of non-HIV sequences. Can the authors report how often this occurs? I would hesitate to label all wells with >1 band as the amplification of >1 provirus. Have the authors performed sequencing from these types of wells? Making this assumption may miss wells that contain usable HIV amplicons and also alter the observed % positive wells. I'm assuming that the authors have found samples where both human and HIV sequences were identified? Were the human sequences disregarded or was that sample discarded completely?
The reviewer is correct in assuming that some wells may contain PCR amplification of non-HIV sequences. In a previous publication where we sequenced 531 sequences from 6 HIV-1 positive individuals, we reported 3-15% of the sequences obtained were due to non-specific amplification[1]. The largest proportion of non-HIV amplifications was seen in cell populations with a low infection frequency, as we generally sequenced as many positive wells as we could obtain.

Additionally, we usually avoid sequencing wells with >1 band due to the potential of obtaining mixtures which would be omitted from analyses. In this case we cannot comment on whether these wells contain an HIV amplicon and a non-specific amplicon. However, when calculating the proportion of intact HIV-1 proviruses in a cell population, we take into account all wells that are positive for any amplification[1].

Lastly, we have sequenced amplicons, which on the gel were a single band, but upon sequencing, have an HIV-1 amplicon and a non-specific amplicon. We only ever consider the contig that the majority of reads map to and that also has >1000X coverage. In these cases, we have kept these sequences as we are confident that the majority of reads map to HIV.

Line 374 - How did the authors handle alignment of HIV sequences of varying sizes. More details would be helpful.
Aligning sequences of varying sizes is a difficult feat for most alignment tools. We have found it most efficient to first manually align each sequence and then use MAFFT to check the alignment. Of the alignment tools we have tried, MAFFT is the most effective computer alignment program to handle this type of alignment. The manuscript has been changed to make these steps clearer:
“5.7.1. Export the final consensus of each proviral sequence into an alignment software such as Molecular Evolutionary Genetics Analysis (MEGA) 7. Align each sequence manually to the HXB2 reference sequence. Trim the 5’ and 3’ ends to positions 666-9650 of HXB2 to remove primer sequences.
5.7.2 Save as a .fasta file. Align using MAFFT version 7 with manual editing where appropriate to obtain the final alignment.”

Line 469-470 - Why can't the FLIPS assay be used as a tool for measuring the size of the HIV reservoir?
If FLIPS amplified every HIV-1 sequence present in a cell population it could be used to measure the size of the HIV-1 reservoir. However, we have not performed these analyses. The manuscript has been changed to read the following:
“Firstly, the FLIPS assay has not been developed as a tool for measuring the size of the latent HIV-1 reservoir as analyses to determine whether FLIPS amplifies every HIV-1 provirus present in a cell population have not been completed.”

Figure 4A and B - A Y-axis is needed to explain what the red bars represent and the green/orange/blue lines underneath with either the "30 - 5314" or "32-6606" labels.
The figure has been modified to include axis labels. Red bars, colours and numbers have been removed as they are not relevant to the overall message this figure is conveying which is the difference in the evenness of read coverage when a single HIV-1 provirus is sequenced, compared to a mixture. The figure legend has also been updated.

1 Hiener, B. et al. Identification of Genetically Intact HIV-1 Proviruses in Specific CD4(+) T Cells from Effectively Treated Participants. Cell Rep. **21** (3), 813-822 (2017).

Kind regards,

BHiener

Bonnie Hiener (Corresponding Author),
Centre for Virus Research
The Westmead Institute for Medical Research
176 Hawkesbury Road
Westmead, Sydney, New South Wales
Australia

Reprint permission for figures 1, 2 and 5
https://www.cell.com/cell-reports/authors?code=cell-site

Permissions

You do not need to ask our permission to use images or information that you have previously published in a Cell Press journal; we ask only that you cite the original publication. For information on how to request permission to use previously published work, please see our Permissions page. If excerpts from other copyrighted works are included in your manuscript, you must obtain written permission from the copyright owners and credit the sources in the article. To obtain permission to use material from Cell Press and Elsevier journals and books, e-mail permissions@elsevier.com. If you have adapted a figure from a published figure, please check with the copyright owners to see if permission is required and include a complete citation/reference for the original article. Obtaining permissions can take up to several weeks. As lack of appropriate permissions can delay publication, we recommend that you request permission at the time of submission.

Print permission for figure 4

Dear Bonnie,
Thank you for contacting QIAGEN Bioinformatics Support. We are happy to hear that you intend to use CLC Genomics Workbench as a source for visualization of your results for your paper.

>I am writing to request permission to use a screenshot of the CLC genomics workbench interface in the JoVE publication...
You can use the the screenshot that you have sent us as part of your article.
To cite a CLC Genomics Workbench, please cite our web site.
For example, CLC Genomics Workbench <the version you are using> (https://www.qiagenbioinformatics.com/)
A detailed explanation how to proceed you can find on our FAQ page.
You can also find a list of papers that already have cited the CLC Genomics Workbench on our web site:
https://www.qiagenbioinformatics.com/?result=citations&s=CLC+Genomics+Workbench
Hopefully you will find a proper example which will serve you as a template.

For any further questions, do not hesitate to contact us.

Best wishes,
Momchil Topalov
Associated Bioinformatics Tech Service Specialist

QIAGEN - Sample to Insight

Phone American: +1-866-464-3684 | Danish: +45-80 82 0167 | German: +49 034133975301
ts-bioinformatics@qiagen.com
www.qiagenbioinformatics.com