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.

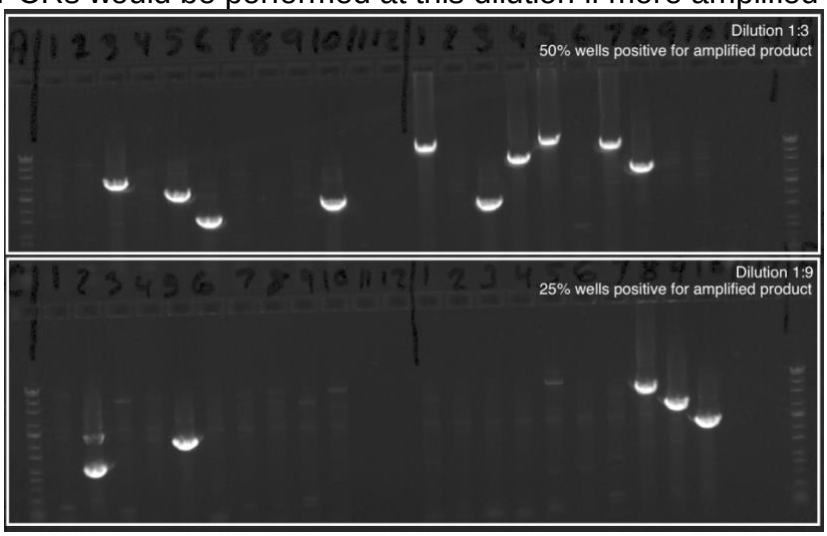
Editor’s and reviewer’s comments are in bold and are addressed point by point below:

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