A new long-read NGS method to sequence HIV-1 integration sites and associated proviruses

00375

Maria Artesi[1#], Vincent Hahaut[1,2#], Basiel Cole[3#], Laurens Lambrechts[3], Ambroise Marçais[4], Olivier Hermine[4], Philip Griebel[5], Natasa Arsic[5], Dominique Bron[2], Vincent Bours[6], Carole Charlier[1], Michel Georges[1], Linos Vandekerckhove[3*], Anne Van den Broeke[1,2*], Keith Durkin[1*]

[1]Unit of Animal Genomics, GIGA Research Institute, ULiège, Liège, Belgium [2]Laboratory of Experimental Hematology, Institut Jules Bordet, ULB, Brussels, Belgium,[3]HIV Cure Research Center, Department of Internal Medicine and Pediatrics, Ghent University Hospital and Ghent University, Ghent, Belgium, Ghent, Belgium, [4]Service d'hématologie, Hôpital Universitaire Necker, Paris, France, [5]Vaccine and Infectious Disease Organization, VIDO-Intervac, University of Saskatchewan, Saskatoon, Canada, [6]Laboratory of Human Genetics, GIGA Research Institute, Liège, Belgium

BACKGROUND

The HIV-1 reservoir represents a major obstacle to HIV cure, making its exploration a priority, however, this task is complicated by its elusiveness, with only ~0.1% of CD4 T cells carrying integrated HIV-1 DNA. Methods based on short-read high throughput sequencing can identify hundreds of HIV-1 insertion sites in a single experiment, but the proviruses within remain a black box. Other approaches can link the sequence of individual proviruses to a specific integration site, but they rely on the whole genome amplification of isolated HIV-1 genomes. As a consequence they are labor intensive and expensive, limiting the number of proviruses one can reasonably interrogate.

METHODS

We have developed Pooled CRISPR Inverse PCR sequencing (PCIP-seq), a method that leverages long reads on the Oxford Nanopore MinION platform to sequence the insertion site and its associated provirus (Figure 1).

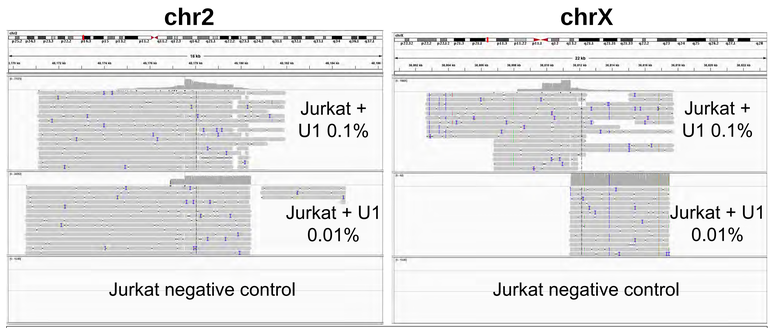


Figure 2 Three PCIP-seq libraries were prepared in parallel using 5 µg of template DNA, all used the same guides and primers. Following sequencing and demultiplexing the Jurkat negative control produced 12,137 reads, Jurkat + U1 0.01% produced 234,421 reads and Jurkat + U1 0.1% 252,913 reads. The resultant reads were mapped to the human genome, the major integration sites observed in U1 on chr2 and chrX are shown. The reads were also mapped the HIV-1 genome. No reads of pure HIV-1 or chimeric HIV-1/host reads were observed in the Jurkat negative control. In Jurkat + U1 0.01% samples 12.6% of the reads were chimeric HIV-1/host, in Jurkat + U1 0.1% this rose to 43.2%.

ADDITIONAL KEY INFORMATION

Contact: kdurkin@uliege.be & anne.vandenbroeke@bordet.be

Work supported by

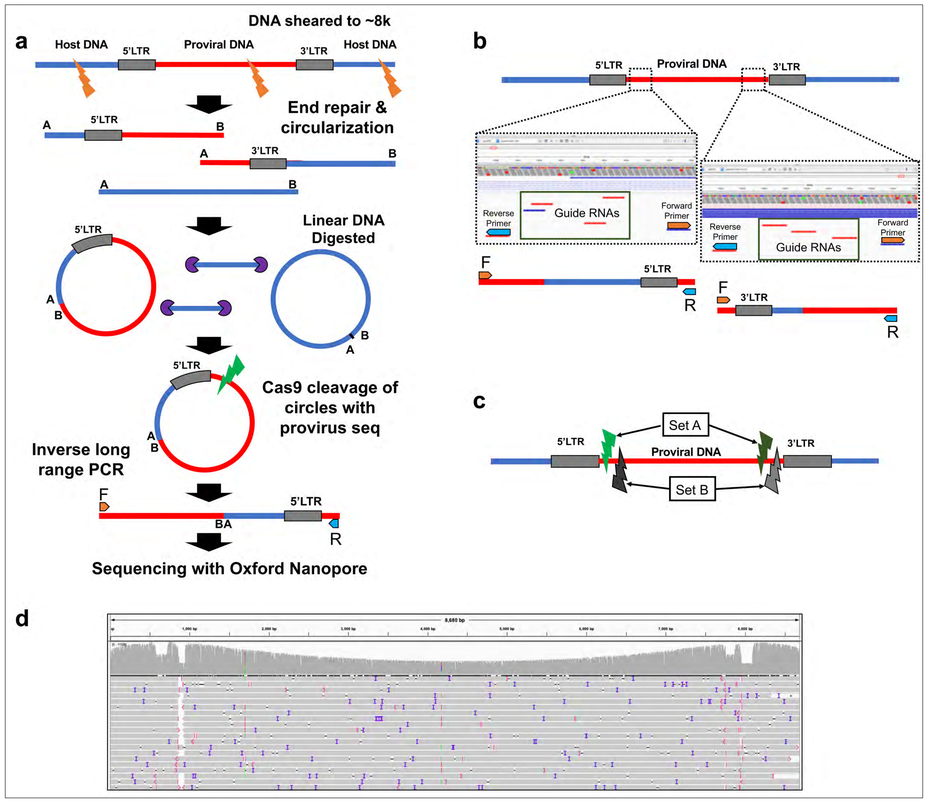


Figure 1. Overview of the PCIP-seq method (a) Simplified outline of method (b) A pool of CRISPR guide-RNAs targets each region, the region is flanked by PCR primers. Guides and primers adjacent to 5' & 3' LTRs are multiplexed. (c) As the region between the PCR primers is not sequenced we created two sets of guides and primers. Following circularization, the sample is split, with CRISPR mediated cleavage and PCR occurring separately for each set. After PCR the products of the two sets of guides and primers are combined for sequencing. (d) Screen shot from the Integrative Genomics Viewer (IGV) showing a small fraction of the resultant reads (grey bars) mapped to the provirus, coverage is shown on top, coverage drops close to the 5' and 3' ends are regions flanked by primers.

RESULTS

We first tested PCIP-seq on 0.1 and 0.01% dilutions of the HIV-1 cell line U1 and demonstrated its utility for examining low proviral loads (Figure 2). We then applied PCIP-seq to CD4 T cells of two HIV-1 patients on long term cART. Using 8 and 12 micrograms of DNA we were able to identify 74 and 159 integration sites respectively, uncovering large numbers of structural variants and SNPs within the integrated proviruses (Figure 3). In patient 02006 we found four clonally expanded full-length proviruses, three appear intact while the fourth has a ~115 bp deletion (Figure 4).

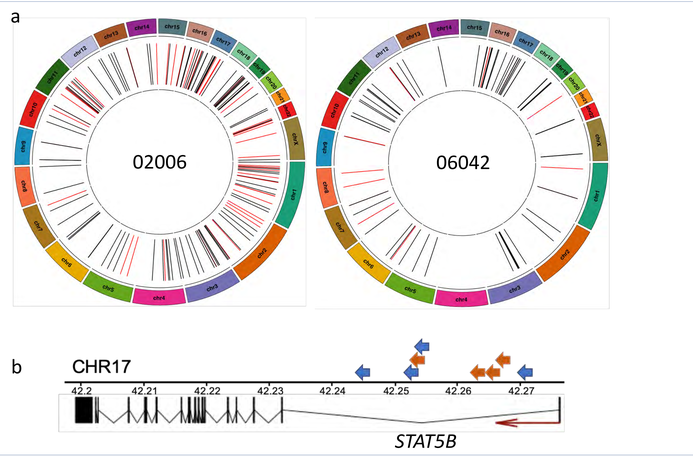


Figure 3. (a) HIV-1 proviral integration sites identified by PCIP-seq in two HIV-1 patients (02006 & 06042). Black lines represent integration sites where the portion of the provirus sequence shows no evidence of a large deletion, red lines indicate sites where a large deletion was observed in the provirus. (b) A hotspot of proviral integration in intron 1 of STAT5B. Arrows represent individual proviruses (02006=blue, 06042=orange) and direction indicates the orientation of the provirus. All proviruses have the same transcriptional orientation as STAT5B.

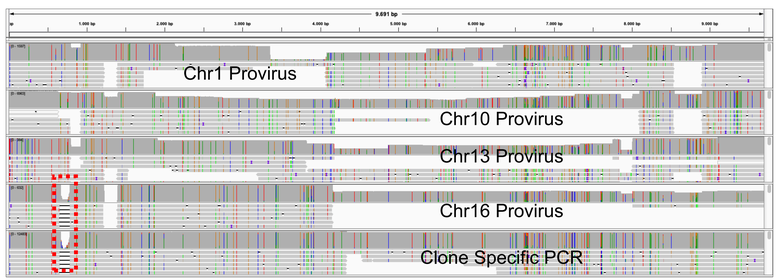


Figure 4. Screen shot from IGV showing the four full proviruses identified in patient 02006. The first 3 appear to be intact but all are integrated in either centromeric repeats or a segmental duplication. The provirus on Chr16 has a ~115 bp deletion disrupting the packaging signal (Ψ).

CONCLUSIONS

Using long reads, we can simultaneously identify the integration site and track clone abundance while also sequencing the HIV-1 provirus inserted at that position. Methods currently used are labor intensive, costly, and only examine a handful of patients. Using PCIP-seq it is feasible to sequence thousands of bases from hundreds of proviruses in a single experiment, opening the landscape of proviral variation and evolution within, and between large numbers of hosts.