Using Barcoded HIV Ensembles (B-HIVE) for Single Provirus Transcriptomics

Article · April 2018

DOI: 10.1002/cpmb.56

3 authors, including:

Heng-Chang Chen
Institut de Génétique Humaine
13 PUBLICATIONS **61** CITATIONS
SEE PROFILE

Eduard Valera Zorita
Centre for Genomic Regulation
9 PUBLICATIONS **101** CITATIONS
SEE PROFILE

Using Barcoded HIV Ensembles (B-HIVE) for Single Provirus Transcriptomics

Heng-Chang Chen,[1,2,3] Eduard Zorita,[1,2,3] and Guillaume J. Filion[1,2]

[1]Genome Architecture, Gene Regulation, Stem Cells and Cancer Programme, Center for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain
[2]University Pompeu Fabra, Barcelona, Spain
[3]These authors contributed equally to this work

The latent HIV reservoir is the main barrier to curing AIDS, because infected cells escape the immune system and antiretroviral therapies. Developing new treatment strategies requires technologies to trace latent proviruses. Here, we describe a genome-wide technique called Barcoded HIV Ensembles (B-HIVE) to measure HIV expression at the single provirus level. The principle of B-HIVE is to tag the genome of HIV with DNA barcodes to trace viral transcripts produced by single proviruses in an infected cell population. This in turn reveals which proviruses are active and which are latent or expressed at low level. B-HIVE is a high-throughput method to identify and quantify thousands of individual viral transcripts per round of infection. It can be applied in different conditions, characterizing the response of single proviruses to different treatments. Overall, B-HIVE gives unprecedented insight into the expression of single proviruses in populations of HIV-infected cells. © 2018 by John Wiley & Sons, Inc.

Keywords: B-HIVE • HIV • latency • molecular barcodes • single-provirus expression

How to cite this article:
Chen, H., Zorita, E., & Filion, G. J. (2018). Using barcoded HIV ensembles (B-HIVE) for single provirus transcriptomics. *Current Protocols in Molecular Biology*, *122*, e56. doi: 10.1002/cpmb.56

INTRODUCTION

More than thirty years after its discovery, HIV remains a significant threat to public health. A small fraction of the viruses can enter a silent state called *latency* where they are innocuous, but also out of reach from the immune system and from antiretroviral therapies. The viruses can remain latent several years (the estimated half-life is 44 months, Finzi et al., 1999) and later awaken to start a new infection if the antiretroviral therapy is interrupted. Latency is thus the major barrier to curing HIV infection. By 2015, the number of people living with HIV was estimated at 36.7 million worldwide (see Internet Resources), indicating an urgent need for a new strategy against the HIV epidemics.

"Shock and kill" is one of the most attractive strategies used to purge patients from HIV infected cells. The principle is to use latency reversing agents (LRAs) such as histone deacetylase inhibitors to deliberately reactivate proviral transcription in cells with a latent infection (Archin et al., 2012). This way, the immune system is expected to detect and kill the infected cells. This strategy has entered exploratory clinical trials, but the early results show limited evidence of efficacy (Cillo et al., 2014). The success of this

Current Protocols in Molecular Biology e56, April 2018
Published online April 2018 in Wiley Online Library (wileyonlinelibrary.com).
doi: 10.1002/cpmb.56

strategy depends on the drugs used to reactivate HIV; however, at present we do not fully understand how the transcriptional state of HIV is affected due to the lack of a proper technique to measure the transcription of individual viruses.

The integration of HIV DNA is one of the critical steps for the persistence of the infection; as such, it has been intensively studied over the past decade (Ciuffi et al., 2005; Debyser, Christ, De Rijck, & Gijsbers, 2015; Schröder et al., 2002; Wang, Ciuffi, Leipzig, Berry, & Bushman, 2007). Integration sites can be revealed by various PCR-based methods such as inverse PCR (Ochman, Ajioka, Garza, & Hartl, 1989; Triglia, Peterson, & Kemp, 1988), ligation-mediated PCR (Mueller & Wold, 1989) and linear amplification-mediated PCR (Schmidt et al., 2007). Although these methods provide rich information on the HIV integration sites; it is currently impossible to have a genome-wide expression landscape of each integrated provirus. Thus, we developed a high-throughput technology called Barcoded HIV Ensembles (B-HIVE, Chen, Martinez, Zorita, Meyerhans, & Filion, 2017) to provide a genome-wide expression map of integrated proviruses at the single-virus level. This map can be combined with other genome-wide techniques such as ATAC-seq (Buenrostro, Giresi, Zaba, Chang, & Greenleaf, 2013; Buenrostro, Wu, Chang, & Greenleaf, 2015), ChIP-seq, 4C and Hi-C (Rao et al., 2014) to study the impact of the chromatin context surrounding the integration site on the transcriptional activity of HIV.

B-HIVE can also be used to measure the efficiency of different LRAs on provirus reactivation (Chen et al., 2017) because it distinguishes the transcriptional state of individual proviruses. For instance, we have showed that vorinostat (also known as SAHA, Bullen, Laird, Durand, Siliciano, & Siliciano, 2014; Spina et al., 2013) and phytohemagglutinin have different effects on the same provirus, and tend to reactivate different subsets of proviruses (Chen et al., 2017).

Here we describe optimized protocols that we use to routinely measure the individual expression of over 1,000 proviruses per infection in Jurkat T cells. The strategy of B-HIVE is to insert DNA barcodes of 20 random nucleotides in a recombinant HIV-based vector (Basic Protocol 1), generate barcoded viruses used for infection (Basic Protocol 2), and use the barcodes to identify the transcripts produced by individual proviruses (Basic Protocol 3). Guidelines for *in-silico* analysis of the results are also provided (Basic Protocol 4). With some minor modifications, B-HIVE can also be carried out in different types of cellular models infected with lentiviral and retroviral vectors tagged with a unique barcode.

BASIC PROTOCOL 1

GENERATION OF THE BARCODED pHCC1 LIBRARY

In this protocol we detail the steps to insert a barcode in the recombinant HIV-based vector pHCC1 by barcoding PCR (see Internet Resources; steps 1 through 4), validate the barcoded pHCC (steps 17 and 18) and prepare the barcoded library (steps 19 through 29). The complete map of pHCC1 is provided (*APPENDIX 1*).

Materials

100 pg recombinant HIV-based vector pHCC1 (the plasmid is available upon request to the authors)
2 U/µl Thermo Scientific Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, cat. no. F530S) containing:
5× Phusion GC Buffer
10 µM Primer GAT431 (5′-[Pho]NNNNNNNNNNNNNNNNNNNNTATAGTGAGTCGTATTAAAAGCGAAAGGGAAACCAGA-3′)
10 µM Primer GAT432 (5′-[Pho]AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTCAAGTCCCTGTTCGGGCGCC-3′)
10 mM dNTPs (Thermo Fisher Scientific, R0181)

100% dimethyl sulfoxide (DMSO; NEB, cat. no. B0515A)
Distilled water
20,000 U/ml *Dpn*I (NEB, cat. no. R0176S)
3,000 U/ml T4 DNA polymerase (NEB, cat. no. M0203S)
0.5 M EDTA (Sigma, cat. no. E9884)
1% (w/v) agarose gel
QIAquick PCR Purification Kit (QIAGEN, cat. no. 28104)
QIAquick Gel Extraction Kit (QIAGEN, cat. no. 28704)
30 U/µl T4 DNA ligase (Thermo Fisher Scientific, cat. no. EL0013)
5% PEG 4000 (included in the T4 DNA ligase kit)
Absolute ethanol (PanReac AppliChem, cat. no. 131086.1214)
75% (v/v) ethanol (PanReac AppliChem, cat. no. 131086.1214)
3 M sodium acetate
20 mg/ml Glycogen (Thermo Fisher Scientific, cat. no. R0561)
Dry ice
10 µM Primer GAT515 (5′-GTTGTAGCTGTCCCAGTATT-3′)
10 µM Primer GAT551 (5′-CTGGCTAACTAGGGAACCCACTGCT-3′)
ElectroMAX™ DH10B™ Cells (Invitrogen, cat. no. 18290015)
1 × S.O.C. medium
LB plate supplemented with 50 µg/ml ampicillin
LB broth
50 µg/ml ampicillin
PureLink™ HiPure Plasmid Maxiprep Kit (Invitrogen, cat. no. K210006)

0.2-ml PCR tubes
PCR C100 Touch™ Thermal Cycler (Bio-Rad)
Nanodrop ND-1000 Spectrophotometer (Thermo Fisher Scientific)
Centrifuge
0.1-cm electrode cuvette (BIO-RAD, cat. no. 165-2089)
Gene Pulser/MicroPulser™ Electroporation Cuvettes, 0.1 cm gap (Bio-Red, cat. no. 1652083)
MicroPulser™ Electroporator (Bio-Red, cat. no. 1652100)
1.5-ml microcentrifuge tubes

Barcoding PCR

The process of barcoding PCR is illustrated in Figure 25C.2.1A-C. See Critical Parameters-Design of the library for further details.

1. Amplify 100 pg pHCC1 in a 0.2-ml PCR tube with the following ingredients in a 50 µl final volume:

   10 µl of 5× Phusion GC Buffer
   5 µl of 10 µM Primer GAT431
   5 µl of 10 µM Primer GAT432
   1 µl of 10 mM dNTPs
   0.5 µl Phusion DNA Polymerase
   1.5 µl of 100% dimethyl sulfoxide (DMSO)
   x µl 100 pg pHCC1
   27 – *x* µl distilled water.

   Prepare two PCR reactions to have minimum 1 µg PCR products. DMSO is used to increase the PCR efficiency.

2. Run the PCR under the following conditions:

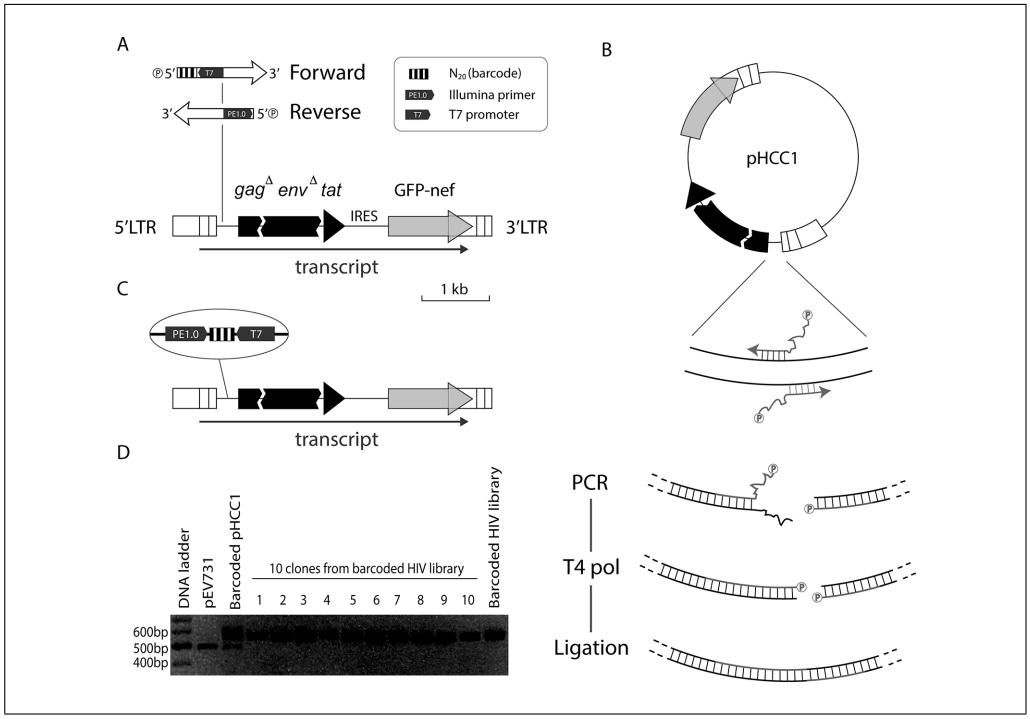

Figure 25C.2.1 Preparation and validation of the barcoded HIV library. (**A**) Structure of the primers with their position on the construct. The forward primer (GAT431) contains a stretch of 20 random nucleotides at the 5′ end that will constitute the barcode, followed by a T7 promoter. The reverse primer (GAT432) contains the sequence of the Illumina sequencing primer PE1.0 at the 5′ end. Both primers are phosphorylated to allow the ligation of the PCR products. (**B**) Schematic of the reaction. The divergent primers amplify the whole backbone. Because the random nucleotides are different between primers, half of the products contain a non-double-stranded end. The T4 polymerase removes the 3′ overhangs and generates blunt products by using the barcode as a template (see Basic Protocol 1, steps 5 through 7). Finally, the blunt products are self-ligated. (**C**) Structure of the HIV construct after the barcoding PCR. The net result is the addition of a 20-nucleotide barcode that is unique for every ligation product, flanked by the Illumina primer PE1.0 and the T7 promoter. (**D**) 10 clones randomly selected in a barcoded library were confirmed by PCR. Only the clones and barcoded HIV library containing an inserted product showed a 566 bp PCR product. Reproduced with permission from Chen et al., 2017.

1 cycle:	1 min	98°C (initial denaturation)
25 cycles:	20 sec	98°C (denaturation)
	1 min	58°C (annealing)
	10 min	72°C (extension)
1 cycle:	10 min	72°C (extension).

3. Add 1 µl *Dpn*I to the PCR products.

   DpnI digests methylated $GA_{(me)}$/TC site; thus, it can be used to remove the plasmid template remaining in PCR products. This step is recommended to increase ligation efficiency in the next two steps.

4. Incubate the reaction for 30 min at 37°C.

Blunt-end the barcoded linear products

5. Adjust the temperature of the reaction to 12°C. Add 1 µl T4 DNA polymerase. Incubate the reaction for 20 min at 12°C .

   Precooling the PCR thermal cycler at 12°C before adding the T4 DNA polymerase will limit the ability of it to degrade the PCR products.

6. Add 2 µl of 0.5 M EDTA into the reaction.

7. Incubate for 20 min at 65°C.

8a. Load 2 µl blunt-ended products on a 1% (w/v) agarose gel. If a single band appears, purify the products with the QIAquick PCR purification kit. Elute the products in a 30 µl final volume.

8b. If the products show multiple bands, gel-purify the band with the expected size of 7.4 kb with the QIAquick Gel extraction kit. Elute the products in a 20 µl final volume.

9. Measure the concentration of the purified products (1 µl) using a Nanodrop spectrophotometer.

The concentration of the products should be between 100 and 200 ng/µl.

Ligate the barcoded linear products

10. Self-ligate the products (19 µl) in 1 × T4 DNA ligase buffer with 2 µl T4 DNA ligase and 25 µl PEG 4000 in a 0.5 ml final volume for 1 hr at 22°C.

We recommend using concentrated T4 DNA ligase for self-ligation. PEG 4000 greatly increases the efficiency of blunt-end DNA ligation.

11. Heat-inactivate the reaction for 10 min at 65°C.

12. Precipitate the ligated products with 1 ml of 100% ethanol, 50 µl 3 M sodium acetate, and 2 µl glycogen in dry ice over 1 hr or at −80°C over 4 hr.

The reaction can be stored up to 3 months at −80°C, if necessary.

13. Centrifuge the reaction for 1 hr at maximum speed, 4°C.

14. Wash the precipitated pellet with 75% ethanol one time followed by centrifugation for 5 min at maximum speed, 4°C.

15. Dry the pellet for 10 min at room temperature.

16. Resuspend the pellet in 11 µl distilled water.

Validate the barcoded pHCC1

Steps 17 and 18 are used to validate whether a barcode was successfully introduced in pHCC1.

17. Amplify barcoded pHCC1 with the primer GAT515 and GAT551 in a 0.2-ml PCR tube with the following ingredients in a 50 µl final volume:

10 µl of 5× Phusion GC Buffer
5 µl of 10 µM Primer GAT515
5 µl of 10 µM Primer GAT551
1 µl of 10 mM dNTPs
0.5 µl Phusion DNA Polymerase
2 µl barcoded pHCC1
26.5 µl distilled water.

Use the plasmid pHCC1 (no barcode) as a negative control for PCR amplification.

18. Load 5 µl the PCR product on a 2% (w/v) agarose gel. The expected size of the product is 566 bp (Fig. 25C.2.1D).

Prepare the barcoded pHCC1 library

19. Add 1 µl barcoded pHCC1 into 20 µl ElectroMAX DH10B competent cells. Maintain the cells on ice.

20. Transfer the mix prepared in step 19 into a pre-cold 0.1-cm electrode cuvette.

    Make sure that the mix is homogeneous and spreads uniformly at the bottom of a cuvette.

21. Place the cuvette inside a MicroPulser Electroporator; select the default setting for *E. coli* electroporation in the electroporator and press the button.

    The default setting used for E. coli electroporation in a 0.1-cm electrode cuvette is the program Ec1 with the following parameters: voltage (kV), 1.8; capacity (μF), 25; resistance (ohm), 200. Check and record the pulse parameters. The time constant should be between 5 and 6 milliseconds after electroporating the cells.

22. Immediately add 0.5 ml of 1× S.O.C. medium into the cuvette and carefully transfer the cells by pipetting into a new 1.5-ml microcentrifuge tube containing 0.5 ml of 1× S.O.C. medium.

23. Place the cells in a shaker at 250 rpm for 1 hr at room temperature.

24. Transfer 1% of the cells (10 μl) and perform three-fold one to ten serial dilutions. Each dilution is in 100 μl final volume.

25. Plate each dilution on the ampicillin-containing LB plate.

26. Incubate the plate overnight at 30°C.

27. Transfer the remaining cells (990 μl) into 200 ml LB broth with ampicillin in the final concentration of 50 μg and culture the cells overnight at 30°C.

28. Calculate the number of colonies the next day.

    It is important to note that each barcoded library should contain at least one million independent clones. Typically, two to five electroporations can be sufficient to obtain a final complexity over one million independent clones. It is recommended to validate some selected colonies by PCR (Fig. 25C.2.1D) with the process mentioned in steps 17 and 18 followed by Sanger sequencing.

29. Extract barcoded pHCC1 library DNA with the PureLink™ HiPure Plasmid Maxiprep Kit.

BASIC PROTOCOL 2

GENERATION OF THE BARCODED VIRIONS AND ISOLATION OF JURKAT T CELLS WITH LATENT BARCODED VIRUSES

The protocols used for the preparation of infectious virions and viral infection are flexible. Here we describe the steps carried out in (Chen et al., 2017) including (1) preparation of single-infection-cycle barcoded HIV by using the barcoded pHCC1 library (steps 1 through 18), (2) infection of Jurkat T cells with barcoded HIV (steps 19 through 23), and (3) isolation of the latently infected Jurkat T cells (steps 24 through 27). Biosafety precautions for working on HIV are followed by the guidelines described in *UNIT 26.5* (Worby & Dixon, 2004).

Materials

HEK 293T cells
Dulbecco's Modified Eagle Medium (DMEM; Gibco, cat. no. 41965-039)
Fetal bovine serum (FBS; Gibco, cat. no. 10270-106)
100× penicillin/streptomycin (Gibco, cat. no. 10378016)
10 μg barcoded pHCC1 library DNA (obtained from Basic Protocol 1)
6.5 μg pCMV-dR8.91 (the plasmid is available upon request to the authors)
3.5 μg pVSV-G (the plasmid is available upon request to the authors)
2 M $CaCl_2$ solution (see recipe)
Nuclease-free water (Ambion, cat. no. AM9938)

2× HBS buffer (see recipe)
100× phosphate buffer (see recipe)
0.05% Trypsin-EDTA (Gibco, cat. no. 25300054). optional
Jurkat T cells
RPMI 1640 Medium, HEPES, no glutamine (Gibco, cat. no. 42401-018)
GlutaMAX™ Supplement (Gibco, cat. no. 35050-061)
Polybrene (Sigma, cat. no. H9268-5G)
1 µg/ml 4′,6′-diamidino-2-phenylindole (DAPI; Thermo Fisher Scientific, cat. no. RE22094430)

10 cm-diameter plates (BD Falcon™, cat. no. 353003)
15- and 50-ml Falcon tubes (Corning®)
2 ml serological pipet (FALCON, cat. no. 357558)
Pipet-aid (Stuart, cat. no. PC2000)
37°C, 5% CO_2 incubator
0.45-µm filter (Merck Millipore, cat. no. SLHA025NB)
Beckman ultracentrifuge tube (Beckman Coulter, cat. no. 358126)
Ultracentrifuge L-100 K (Beckman Coulter)
SW 32 Ti Rotor (Beckman Coulter)
1.5-ml microcentrifuge tubes
6-well plate (COSTAR, cat. no. 3506)
FACSCalibur cell sorter (Becton Dickinson)
FACS Aria-SORP (Becton Dickinson)
Cell culture flasks (surface area 25 cm^2, Corning®)

CAUTION: Follow biosafety protocol and use 10% (v/v) bleach to inactivate HIV in the plates, tubes, plastic pipets, etc.

Plate HEK293T cells

1. Seed 2 × 10^6 HEK 293T cells in 10 ml DMEM medium supplemented with 1 × penicillin/streptomycin and 1 × fetal bovine serum (FBS) in 10 cm-diameter plates the day before transfection. Harvest the cells overnight.

Transfect HEK293T cells

2. Replace culture medium with fresh DMEM culture medium 1 hr before transfection.
3. Prepare transfection reagent A by mixing the following ingredients in a 15-ml Falcon tube.

   10 µg barcoded pHCC1 library DNA
   6.5 µg pCMV-dR8.91
   3.5 µg pVSV-G
   60 µl 2 M $CaCl_2$ solution
   Nuclease-free water to a 500 µl final volume.

   The transfection reagent A is freshly prepared.

4. Prepare transfection reagent B by mixing the following ingredients in a 15-ml Falcon tube.

   490 µl of 2× HBS buffer
   10 µl of 100× phosphate buffer.

   The transfection reagent B is freshly prepared.

5. Add dropwise the contents from transfection reagent A into the 15-ml Falcon tube containing transfection reagent B while bubbling reagent B.

Bubble reagent B by blowing air constantly through a 2-ml serological pipet with a pipet-aid.

6. Incubate the mix for 30 min at room temperature.
7. Add dropwise the mix (1 ml) to HEK 293T cells in a 10 cm-diameter plate. Mix gently by rotating the plate.
8. Place the cells for 16 hr in a 37°C, 5% CO_2 incubator.
9. Replace culture medium of HEK 293T cells with 10 ml 37°C prewarmed fresh DMEM culture medium the next day.

Gently manipulate the cells after transfection.

Harvest and concentrate viral stocks

10. Collect first harvest of supernatant in a 50-ml falcon tube day 2 post-transfection. Add 10 ml 37°C prewarmed fresh DMEM culture medium to the HEK 293T cells in a 10 cm-diameter plate.

Store first harvest of supernatant at 4°C.

11. Collect the second harvest of the supernatant day 3 post-transfection. Pool the supernatant from the first and second harvest in one 50-ml Falcon tube.

Optional step: Trypsinize HEK 293T cells and quickly estimate the transfection efficiency based on the percentage of GFP(+) cells determined through FACS analysis. A successful transfection usually brings over 80% GFP(+) cells.

12. Centrifuge the supernatant for 5 min at 300 × *g*, 4°C.
13. Clear the supernatant of cell debris by filtering through a 0.45-μm filter.
14. Transfer the supernatant into the Beckman ultracentrifuge tube.
15. Ultracentrifuge the tube for 90 min at 28,500 × *g*, 4°C.
16. Carefully aspirate the supernatant. Resuspend the viral pellet with 100 μl of 37°C prewarmed fresh DMEM culture medium.
17. Place the tube for 30 min at room temperature.
18. Transfer the barcoded viral supernatant into a 1.5-ml microcentrifuge tube and store up to 6 months at −80°C.

Titrate the barcoded viral stock before infecting the target cells.

Infect Jurkat T cells with barcoded HIV

19. Seed 1 × 10^6 Jurkat T cells in 3 ml RPMI 1640 medium supplied with 1 × penicillin/streptomycin, 1 × GlutaMAX, 1 × FBS, and 8 μg/ml polybrene per well in a 6-well plate before infection.
20. Add appropriate volumes of barcoded HIV virions at a multiplicity of infection of 0.5 per well.
21. Centrifuge the 6-well plate for 90 min at 1000 × *g*, 32°C.
22. Place the cells for 24 hr in a 37°C, 5% CO_2 incubator.
23. Replace the culture medium of HIV-infected Jurkat T cells with 3 ml 37°C prewarmed fresh RPMI culture medium the next day.

The efficiency of infection can be monitored based on the expression of a GFP reporter in barcoded viruses with a FACSCalibur cell sorter on day 2 post-infection.

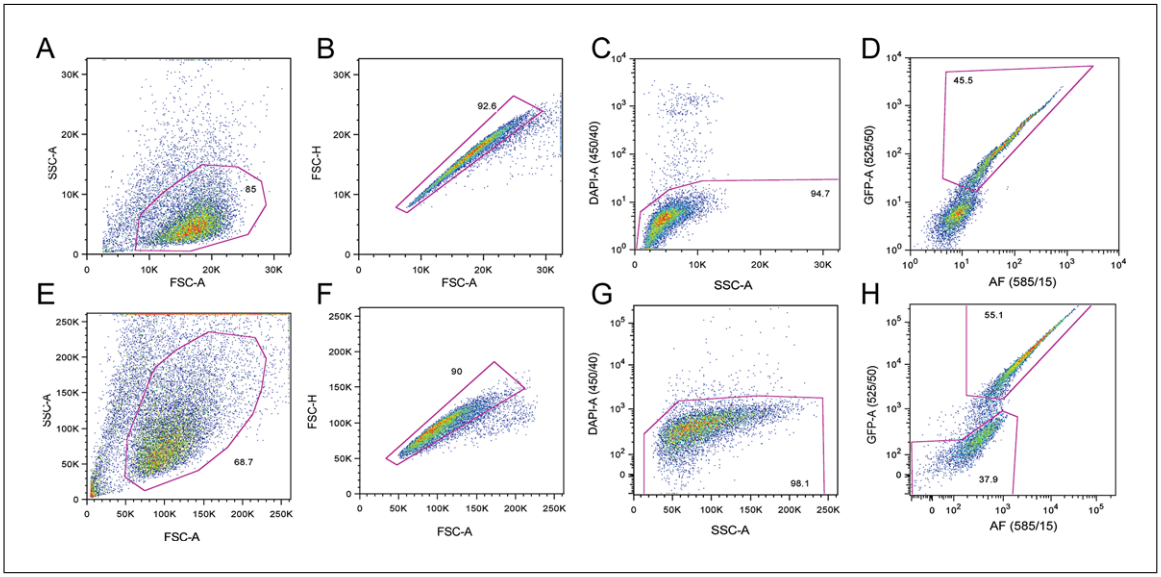

Figure 25C.2.2 FACS concentration and isolation of latently infected Jurkat T cells with barcoded HIV. Representative FACS profiles for sorting GFP(+) cells 4 days post-infection (see Basic Protocol 2, step 24) (**A** to **D**), and for separating GFP(-) from GFP(+) cells in a founder cell pool (see Basic Protocol 2, step 27) (**E** to **H**). (**A** and **E**) Jurkat T cells were selected by FSC-A versus SSC-A (pink line-gated region). (**B** and **F**) Singlets were selected by FSC-A versus FSC-H (pink line-gated region). (**C** and **G**) Live cells (pink line-gated region) were selected by DAPI negative on a SSC-A versus DAPI-A (wavelength: 450/40). (**D** and **H**) GFP(+) cells (day 4 post-infection) and latently infected GFP(-) cells (day 21 post-infection) were acquired by GFP fluorescence (wavelength: 525/50) versus autofluorescence (AF, wavelength: 585/15). See the color version of this figure online. Reproduced with permission from Chen et al., 2017.

FACS sorting to enrich HIV infected cells

24. FACS-sort GFP(+) cells with a FACS Aria-SORP on day 4 post-infection (Fig. 25C.2.2A-D).

    1 µg/ml DAPI is added in each sample to separate the live from the dead cells while sorting.

25. Distribute GFP(+)-sorted infected cells to several small pools with 20,000 infected founder cells per pool.

    The purpose of this approach is threefold: (1) It ensures that every cell in the founder population is infected with at least one active virus, (2) the bottleneck in the founder pool decreases the chances that two viruses have the same barcode, and (3) The long terminal repeat (LTR)-driven GFP expression allows identification of latency events on the basis of loss of fluorescence (see step 27).

26. Expand each founder population for 17 days in a 37°C, 5% CO_2 incubator.

    It is recommended to immediately wash sorted GFP(+) cells after FACS-sorting before separating the cells into different small founder cell pools. Due to the low cell density, we recommend to first grow 20,000 founder cells in culture per well in a 6-well plate without replacing culture medium. After 4 to 5 days of culture in a 6-well plate, semi-dense cells can be transferred to a cell culture 25-cm^2 flask and continue to be expanded.

Separate latently infected cells from the heterogeneous founder cell pool

27. Isolate latently infected GFP(−) cells from GFP(+) cells with a FACS Aria-SORP (Fig. 25C.2.2E-H).

    1 µg/ml DAPI is used to separate the live from dead cells while sorting. FACS-sorted GFP(−) cells are the cells with latent infections. Insert-specific expression of integrated proviruses is mapped in this cell population (see Basic Protocol 3).

PREPARE B-HIVE SAMPLES FOR ILLUMINA HIGH-THROUGHPUT SEQUENCING

This protocol is the main focus of the technology; it aims to map proviruses and measure their individual expression. This protocol comprises four parts: (1) preparation of the sequencing library for mapping provirus integration sites (steps 1 through 27 and Fig. 25C.2.3A), (2) amplification of RNA (steps 28 through 38 and Fig. 25C.2.3B) and DNA barcodes (steps 39 through 43 and Fig. 25C.2.3B), (3) quality control of sequencing libraries (steps 44 through 46), and (4) performing high-throughput sequencing (steps 47 through 49).

Materials

Barcoded HIV-infected Jurkat T cells (see Basic Protocol 2)
AllPrep DNA/RNA Mini Kit (QIAGEN, cat. no. 80204) containing:
    TE buffer
10 U/µl *Bpl*I (Thermo Fisher Scientific, cat. no. ER1311) containing:
  Tango buffer and 1× SAM
5,000 U/ml HpyCH4III (NEB, cat. no. R0618S)
10 mM dNTPs (Thermo Fisher Scientific, cat. no. R0181)
3,000 U/ml T4 DNA polymerase (NEB, cat. no. M0203S)
5,000 U/ml DNA polymerase I, Klenow fragment (NEB, cat. no. M0210S)
10,000 U/ml T4 polynucleotide kinase (NEB, cat. no. M0201S)
0.5 M EDTA (Sigma, cat. no. E9884)
30 U/µl T4 DNA ligase (Thermo Fisher Scientific, cat. no. EL0013)
Absolute ethanol (PanReac AppliChem, cat. no. 131086.1214)
3 M sodium acetate
20 mg/ml Glycogen (Thermo Fisher Scientific, cat. no. R0561)
Dry ice
75% (v/v) ethanol (PanReac AppliChem, cat. no. 131086.1214)
Distilled water
10 U/µl Plasmid-Safe™ ATP-Dependent DNase (Epicentre, E3101K) containing:
    25 mM ATP
    10× Plasmid-Safe reaction buffer
    Plasmid-Safe DNase
MinElute PCR Purification Kit (QIAGEN, cat. no. 28004)
20,000 U/ml *Sac*I (NEB, cat. no. R0156S)
2 U/µl Thermo Scientific Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, cat. no. F530S)
1 µM Primer GAT316 (5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′)
1 µM Primer GAT645 (5′-GCCCTGGTGTGTAGTTCTGCCA-3′)
1 µM Primer GAT024 (5′-AATGATACGGCGACCACCGAGATCTACACTCTTT CCCTACACGACGCTCTTCCGATCT-3′)
1 µM GAT-int Primer [5′-CAAGCAGAAGACGGCATACGAGAT-(6 nt index)-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGAGTGAATTAG CCCTTC-3′]
2% (w/v) agarose gel
Oligotex mRNA Mini Kit (QIAGEN, cat. no. 70022)
ThermoScript RT-PCR System (Invitrogen, cat. no. 11145-024) containing:
    cDNA synthesis buffer
    DTT
    RNaseOUT
    ThermoScript
    DEPC-treated water

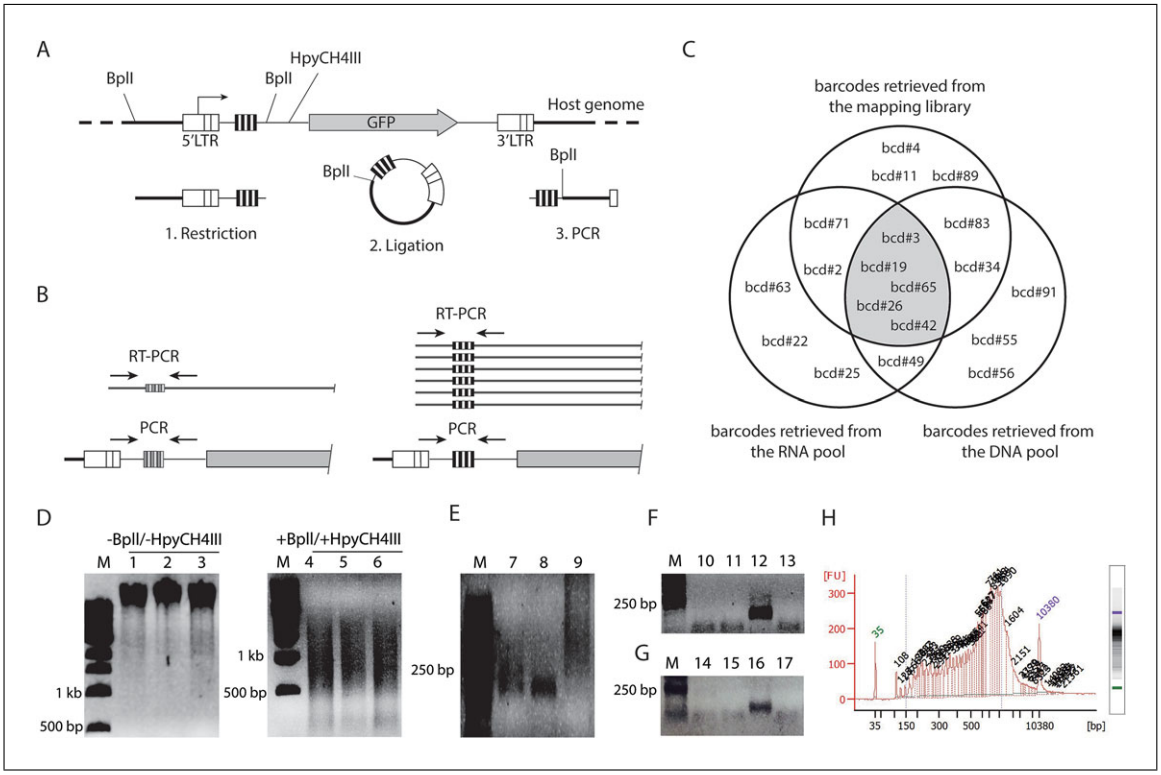

Figure 25C.2.3 Principle of the B-HIVE technology and expected results. (**A**) Provirus mapping (see Basic Protocol 3, steps 1 through 27). Mapping is carried out by inverse PCR. The genome of the cell population is digested by the restriction enzymes BpII and HpyCH4III, which create hybrid human/HIV fragments containing the barcodes. The digestion products are ligated in conditions favoring self-ligation (see Basic Protocol 3, step 6). A PCR with primers in the LTR is then carried out to amplify the ligation product containing the human DNA flanking the insertion and the barcode (see Basic Protocol 3, steps 20 through 24). (**B**) Provirus quantification (see Basic Protocol 3, steps 28 through 43). The provirus on the left has a lower expression than the provirus on the right. The corresponding barcode on the left is less abundant in the RNA than the barcode on the right. Barcode abundance is measured by RT-PCR on the RNA pool versus PCR on the DNA pool. (**C**) Schematic representation of barcodes used to map insert-specific expression of integrated proviruses. Only barcodes lying in the grey region (bcd#3, #19, #26, #42 and #65) in the Venn diagram are considered to be informative. (**D**) Expected results from genomic DNA digested by BpII and HpyCH4III (see Basic Protocol 3, steps 1 through 3). The size of digested products is expected to lie between 500 bp and 1 kb. Lanes 1, 2 and 3: before BpII and HpyCH4III digestion; lanes 4, 5 and 6: after BpII and HpyCH4III digestion. Lanes 1 and 4: uninfected cells; lanes 2 and 5: genomic DNA from cells infected by non-barcoded viruses; lanes 3 and 6: genomic DNA from cells infected with barcoded viruses. (**E**) Expected results from the preparation of the mapping library. 2.0% (w/v) agarose gel displaying a PCR smear obtained from the second round of inverse PCR corresponding to different integrations in cells infected with barcoded viruses (lane 9) (see Basic Protocol 3, step 25). Uninfected cells (lane 7) and cells infected by non-barcoded HIV (lane 8) did not yield any PCR product. (**F**) Expected results from RNA barcode amplification (see Basic Protocol 3, step 37). Only cells infected with barcoded viruses displayed the specific PCR product (lane 12). Uninfected cells (lane 10), cells infected by non-barcoded HIV (lane 11) and a no reverse transcriptase control (lane 13) did not yield any PCR product. (**G**) Expected results from DNA barcode amplification (see Basic Protocol 3, step 41). Only cells infected with barcoded viruses displayed the specific PCR product (lane 16). Uninfected cells (lane 14), cells infected by non-barcoded HIV (lane 15) and a water control (lane 17) did not yield any PCR product. (**H**) The example Bioanalyzer profile of a mapping library. A smear lies between 200 and 700 bp with peaks concentrated from 500 to 700 bp. The *y* axis represents fluorescence units (FU); the × axis represents the size of PCR products. Panels A, B, D and E reproduced with permission Chen et al., 2017.

20 µM RT primer GAT526 (5′-CCCTTTCGCTTTTAATACGACTCACTATA-3′)
1 µM GAT-bcd-amp Primer [5′-CAAGCAGAAGACGGCATACGAGAT-(4 nt index)-TTTTAATACGACTCACTATA-3′]
Qubit™ dsDNA HS Assay Kit (Invitrogen, cat. no. Q32851)
Agilent High Sensitivity DNA Kit (Agilent, cat. no. 5067-4626)
Agilent DNA 1000 Kit (Agilent, cat. no. 5067-1504)

KAPA library quantification kit (Kapa Biosystems, cat. no. KK4835)
NextSeq 500/550 High Output v2 kit (Illumina, cat. no. FC-404-2005)
HiSeq SBS Kit V4 (Illumina, cat. no. FC-401-4002)

37°C incubator
PCR C100 Touch™ Thermal Cycler (Bio-Rad)
2.0-ml tubes (Eppendorf, cat. no. 0030 120.094)
0.2-ml PCR tubes (Eppendorf, cat. no. 0030 124.332)
Refrigerated centrifuge (Eppendorf, cat. no. 5415R)
Qubit™ Assay Tubes (Invitrogen, cat. no. Q32856)
Qubit Fluorometric Quantitation (Invitrogen, cat. no. Q33216)
Bioanalyzer High-Sensitivity DNA Analysis kit (Agilent)
NextSeq sequencer (Illumina)
HiSeq 2000 sequencer (Illumina)

Digest genomic DNA

1. Extract genomic DNA and total RNA from infected Jurkat T cells with an AllPrep DNA/RNA Mini Kit. Elute the DNA from the supplied filter column with 35 µl TE buffer (supplied in the kit).

   Genomic DNA and RNA can be extracted with different protocols. We recommend to concurrently extract genomic DNA and RNA from the same cell population in order to maximize the overlap between DNA and RNA barcodes.

2. Digest 3 µg genomic DNA with 2 µl *Bpl*I and 2 µl HpyCH4III in Tango buffer complemented with 1× SAM in a 50 µl final volume for 3 hr at 37°C.

   Start with at least 2 µg genomic DNA to prepare a mapping library. Load 1.5 µl double-digested product on a 1% (w/v) agarose gel to validate the efficiency of digestion. A good average fragment length of digested genomic DNA is expected to lie between 500 bp to 1 kb (Fig. 25C.2.3D). Long genomic DNA decreases the ligation efficiency in step 6.

3. Heat-inactivate the reaction for 20 min at 65°C using a thermal cycler.

Blunt-end and self-ligate enzyme-digested genomic DNA

4. Precool the thermal cycler to 12°C. Blunt-end double-digested products by dilution of the mix in 1 × T4 DNA ligase buffer with 3.3 µl dNTPs, 4.2 µl T4 DNA polymerase, 0.8 µl DNA polymerase I, Klenow fragment, and 4.2 µl T4 polynucleotide kinase in a 100 µl final volume for 20 min at 12°C.

   Precooling the PCR thermal cycler at 12°C before adding the T4 DNA polymerase is important because heat may degrade the PCR products.

5. Add 2 µl of 0.5 M EDTA and then heat-inactivate the reaction for 20 min at 65°C.

6. Immediately self-ligate blunt-ended products in 1× T4 DNA ligase buffer with 2 µl T4 DNA ligase in a 1 ml final volume overnight at 16°C.

   Ligation is carried out in a large volume to favor intramolecular ligation.

7. Separate each self-ligated product equally into two 2.0-ml tubes following the overnight incubation.

8. Precipitate each tube of products with 1 ml of 100% ethanol, 50 µl of 3 M sodium acetate. and 2 µl glycogen in dry ice over 1 hr or at −80°C over 4 hr.

9. Centrifuge the mix for 1 hr at maximum speed,4°C, in a refrigerated microcentrifuge.

10. Wash the precipitated pellet with 0.5 ml of 75% ethanol one time followed by centrifugation for 5 min at maximum speed, 4°C.

11. Dry the pellet for 10 min at room temperature.

Destroy non-circularized genomic DNA

Steps 12 through 16 are optional. However, we strongly recommend including these steps in order to reduce the amount of nonspecific products, because unligated products may cause PCR artifacts and introduce mapping errors.

12. Resuspend the pellet in 84 µl distilled water.

13. Add 4 µl of 25 mM ATP (supplied in the kit), 10 µl of 10× Plasmid-Safe reaction buffer, and 2 µl Plasmid-Safe DNase with 84 µl products in a 100 µl final volume.

14. Place the reaction for 2 hr at 37°C.

15. Heat-inactivate the reaction for 30 min at 70°C.

16. Purify circularized DNA with a MinElute PCR Purification Kit following the manufacturer's instructions and elute the product in a 20 µl final volume.

Linearize the circularized products

17. Digest circularized DNA with 2 µl *Sac*I in a 30 µl final volume for 3 hr at 37°C.

18. Heat-inactivate the reaction for 20 min at 65°C.

    This step increases the efficiency of the inverse PCRs.

Amplify libraries for mapping provirus integration sites

19. Prepare the reaction in a 0.2-ml PCR tube for the first round of nested PCR. Each PCR reaction mix contains the following ingredients in a 50 µl final volume:

    10 µl of 5× Phusion GC Buffer
    5 µl of 1 µM Primer GAT316
    5 µl of 1 µM Primer GAT645
    1 µl of 10 mM dNTPs
    0.5 µl Phusion DNA Polymerase
    8 µl linearized DNA
    20.5 µl distilled water.

    We recommend preparing two PCR reactions per sample to obtain enough material for Illumina high-throughput sequencing. Always include a no-template control to confirm that the final PCR products are not amplified from contaminating DNA. Primers and PCR conditions are optimized for detecting HIV integration sites in Jurkat T cells infected with barcoded HIV-based pHCC1 vector. Modifications are required when different types of viral vectors and cellular models are used.

20. Run the first round of PCR under the following conditions:

1 cycle:	1 min	98°C (initial denaturation)
10 cycles:	20 sec	98°C (denaturation)
	1 min	65°C (annealing)
	5 min	72°C (extension)
1 cycle:	5 min	72°C (final extension).

21. Transfer 10 µl first-round PCR products for the second round of nested PCR.

22. Prepare a PCR reaction mix in a 0.2-ml PCR tube with the following ingredients in a 50 µl final volume:

10 µl of 5× Phusion GC Buffer
5 µl of 1 µM Primer GAT024
5 µl of 1 µM GAT-int Primer
1 µl of 10 mM dNTPs
0.5 µl Phusion DNA Polymerase
10 µl first-round PCR product
18.5 µl distilled water.

23. Run the second round of PCR under the following conditions: 98°C for 1 min, 98°C for 20 sec, 55°C for 1 min and 72°C for 5 min (2 cycles); 98°C for 20 sec, 65°C for 1 min and 72°C for 5 min (15 cycles); and 72°C for 5 min.

24. Load 5 µl second-round PCR products on a 2% (w/v) agarose gel. The expected size of the smear is between 200 and 700 bp (Fig. 25C.2.3E).

25. Pool together the products from the same sample, if necessary.

26. Elute the products with the MinElute PCR purification kit in a 20 µl final volume.

Continue from step 44 to check the quality of the libraries before Illumina high-throughput sequencing. In the event that primer dimers appear on the gel, PCR products can be purified with Agencourt AMPure XP beads.

Purify mRNA

RNA barcode amplification is carried out by reverse transcription polymerase chain reaction (RT-PCR) to quantify individual viral transcripts. Individual RNA barcode expression measured by B-HIVE can be calibrated by template-switch T7 PCR (see Support Protocol 1).

27. Purify mRNA from total RNA extracted in step 1 with the Oligotex mRNA Mini Kit following the manufacturer's instructions. Elute purified mRNA in a 44 µl final volume.

RT-PCR on mRNA barcodes

28. Prepare a reverse transcription (RT) reaction mix in a 0.2-ml PCR tube with the following ingredients in a 12 µl final volume:

10 µl purified mRNA
1 µl of 20 µM RT primer GAT526
1 µl of 10 mM dNTPs.

We recommend preparing two independent RT reactions for each sample to validate the expression of the same barcode.

29. Incubate the reaction for 1 min at 95°C and immediately put the reaction on ice.

This step is used to denature mRNA. We recommend performing reverse transcription in a PCR thermal cycler.

30. Prepare the RT master mix with the following reagents:

4 µl of 5× cDNA synthesis buffer
1 µl of 0.1 M DTT
1 µl of 40 U/µl RNaseOUT
1 µl of 15 U/µl ThermoScript
1 µl DEPC-treated water.

All reagents are included in the ThermoScript RT-PCR System. The master mix shown above is used for one RT reaction.

31. Add 8 µl master mix to the denatured mRNA.
32. Incubate the reaction mix for 1 hr at 65°C.
33. Heat-inactivate the reaction mix for 5 min at 85°C.
34. Amplify cDNA in a 0.2-ml PCR tube with the following ingredients in a 50 µl final volume:

    10 µl of 5× Phusion GC Buffer
    5 µl of 10 µM Primer GAT024
    5 µl of 10 µM GAT-bcd-amp Primer
    1 µl of 10 mM dNTPs
    0.5 µl Phusion DNA Polymerase
    5 µl cDNA
    23.5 µl distilled water.

35. Run the PCR under the following conditions:

1 cycle:	1 min	98°C (initial denaturation)
27 cycles:	20 sec	98°C (denaturation)
	30 sec	60°C (annealing)
	1 min	72°C (extension)
1 cycle:	5 min	72°C (final extension).

36. Load 5 µl RT-PCR products on a 2% agarose gel. The expected size of the product is 126 bp (Fig. 25C.2.3F).
37. Elute the products with the MinElute PCR purification kit in a 20 µl final volume.

    Continue from step 44 to check the quality of the libraries before Illumina high-throughput sequencing.

Amplify DNA barcodes

Viral transcripts need to be normalized by the amount of DNA barcodes amplified by PCR in the same cell population because founder cells may not divide at the same rate. The combined information reveals the expression landscape of individual proviruses inserted at distinct genomic locations.

38. Amplify DNA barcodes with 200 ng genomic DNA extracted in step 1 with the following ingredients in a 50 µl final volume:

    10 µl of 5× Phusion GC Buffer
    5 µl of 10 µM Primer GAT024
    5 µl of 10 µM GAT-bcd-amp Primer
    1 µl of 10 mM dNTPs
    0.5 µl Phusion DNA Polymerase
    200 ng genomic DNA
    28.5 µl distilled water.

    We recommend measuring dsDNA concentration with the Qubit™ dsDNA HS Assay Kit following manufacturer's instructions. Prepare at least five PCR reactions for each sample.

39. Run the PCR under the following conditions:

1 cycle:	1 min	98°C (initial denaturation)
29 cycles:	20 sec	98°C (denaturation)
	30 sec	58°C (annealing)
	1 min	72°C (extension)
1 cycle:	5 min	72°C (final extension).

40. Load 5 µl PCR products on a 2% (w/v) agarose gel. The expected size of the product is 126 bp (Fig. 25C.2.3G).

41. Pool together the PCR products from the same sample, if necessary.

42. Elute the products with the MinElute PCR purification kit in a 20 µl final volume.

Continue from step 44 to check the quality of the libraries before Illumina high-throughput sequencing.

Quality control of sequencing libraries

43. Measure the concentration of each library with the Qubit™ dsDNA HS Assay Kit.

44. Visualize the quality of each library on a Bioanalyzer.

Run the library for mapping with Agilent High Sensitivity DNA Reagents; run the library of RNA and DNA barcodes amplification with Agilent DNA 1000 Reagents. Calculate the average fragment length (bp) of each library amplicon based on a Bioanalyzer profile.

45. Calculate the size-adjusted concentration of each library based on the concentration measure by qPCR with the KAPA library quantification kit and the average fragment length obtained in step 45.

Perform Illumina high-throughput sequencing

46. Pool libraries together if they need to be sequenced in the same lane of the flow cell.

If demultiplexing must be performed after sequencing, make sure that each library contains a different sequencing index.

47. Adjust pooled libraries to a 4 nM final concentration.

48a. Sequence libraries for mapping as 76-bp paired-end reads on a NextSeq sequencer with NextSeq 500/550 High Output v2 reagents.

48b. Sequence libraries of RNA and DNA barcodes amplification as 50-bp single reads on a HiSeq2000 sequencer with HiSeq SBS Kit V4 reagents.

BASIC PROTOCOL 4

DATA ANALYSES *IN SILICO*

The simplified pipeline provided in this protocol aims to facilitate the analysis process. With this, users can assay the sequencing data of provirus mapping and genomic DNA and RNA barcode amplification. Basic design guidelines of a custom pipeline are described in Support Protocol 3.

Necessary Resources

Hardware:

A 64-bit computer with at least 16 GB of RAM running Linux (Ubuntu 14.04 is preferred).

Software:

Docker CE 17.05
Nextflow 0.24
Python 2.7
Java JRE 1.7

Software included in the Docker container (no installation required):

Python 2.7
R version 3.3.1
bwa 0.7.5a
Starcode 1.1
Seeq 1.1.2
Pipeline source files

Files:

Paired-end read files with integration site information (see Basic Protocol 3, step 49a) in gzip-compressed *FASTQ* format
Single-end read files with DNA and RNA barcodes (see Basic Protocol 3, step 49b) in gzip-compressed *FASTQ* format
Human genome reference in *FASTA* format

Download Programs and files

1. Install Docker Community Edition (see *https://www.docker.com*).
2. Install Java Runtime Environment (see *https://www.java.com*).
3. Download pipeline source files. In a Linux terminal type or copy-paste the following commands:

```
wget https://github.com/gui11aume/
BHIVE_for_single_provirus_transcriptomics/archive/
  master.zip
```

NOTE: The URL above is too long to fit on a single line of this page. Write it as a single word in the terminal, i.e., without separation between `gui11aume/` *and* `BHIVE`.

```
unzip master.zip
rm master.zip
cd BHIVE_for_single_provirus_transcriptomics-master
```

4. Download and install nextflow (see *https://www.nextflow.io* for updated instructions):

```
curl -s https://get.nextflow.io | bash
```

5. Download the human genome reference (skip to use own reference):

```
wget http://hgdownload.cse.ucsc.edu/
goldenpath/hg19/bigZips/chromFa.tar.gz
```

NOTE: As above, write the URL as a single word in the terminal.

```
tar -Oxf chromFa.tar.gz > ref-genome.fasta
rm chromFa.tar.gz
```

6. Build bwa index file (type or copy-paste the following in a single line):

```
docker run -v $(pwd):/data ezorita/bioinformatics
  bwa index
/data/ref-genome.fasta
```

Edit mapping parameters file

7. Open the file `map.cfg` with a text editor.
8. Introduce the information of the experiment replicates (Basic Protocol 3, steps 1 to 27). Instructions are provided in `map.cfg`.

Map provirus integration sites

9. Run the mapping pipeline (type or copy-paste the following in a single line):

```
./nextflow run map.nf --index ref-genome.fasta
  -with-docker
ezorita/bioinformatics -resume
```

Integration site file

10. Review the output file `hiv_integrations.txt`.

    The output file has a tab-separated column format and contains one insertion per row. The reference for the columns of the HIV integration file is

    - brcd: provirus barcode
    - chr: integration chromosome
    - locus: integration nucleotide
    - strand: integration strand
    - reads: count of provirus molecules sequenced
    - mapq: confidence score (from 0 to 100)
    - rep: replicate number

Demultiplexing DNA and mRNA barcodes

Demultiplexing is only necessary if multiple samples were pooled together in Basic Protocol 3, step 47.

11. Open the file `demux.cfg` with a text editor.

12. Introduce information of indices used in barcode PCR (Basic Protocol 3, steps 39-43) and barcode RT-PCR (Basic Protocol 3, steps 28-38). Instructions are provided in `demux.cfg`.

13. Run demultiplexing script:

```
python demux.py demux.cfg
```

Edit expression parameters file

14. Open the file `expr.cfg` with a text editor.

15. Introduce information of DNA and mRNA barcode sequencing files (see Basic Protocol 3, steps 28-43). Instructions are provided in `expr.cfg`.

Compute provirus expression

16. Run the expression pipeline (type or copy-paste the following in a single line):

```
nextflow run expr.nf --integs hiv_integrations.txt
  -with-dockerezorita/bioinformatics -resume
```

Provirus expression file

17. Review the output file `hiv_expression.txt`.

    The output file has a tab-separated column format and contains one provirus per row. The reference for the columns of the HIV expression file is

    - brcd: provirus barcode
    - rep: biological replicate number
    - chr: integration chromosome
    - locus: integration nucleotide
    - strand: integration strand

- reads: count of provirus molecules sequenced
- mapq: confidence score (from 0 to 100)
- dna: count of DNA barcode molecules sequenced
- rna: count of mRNA barcode molecules sequenced
- exprscore: balanced expression score

Replicate correlation plots

18. A pdf file is generated for each biological replicate in the `figures/` folder.

TEMPLATE-SWITCH T7 PCR AND QUANTITATIVE PCR TO CONFIRM PROVIRUS EXPRESSION

SUPPORT PROTOCOL 1

This method can be used to validate RNA barcode expression obtained from B-HIVE after knowing the sequence of the barcodes mapped on the genome. The process of template-switch T7 PCR is illustrated in Figure 25C.2.4A.

Materials

2 µg genomic DNA from barcoded HIV-infected Jurkat T cells (see Basic Protocol 3)
5,000 U/ml HpyCH4III (NEB, cat. no. R0618S) containing:
    1× CutSmart Buffer
HiScribe™ T7 Quick High Yield RNA Synthesis Kit (NEB, cat. no. E2050S)
10 mM $CaCl_2$ (Sigma-Aldrich, cat. no. C4901-100G)
2 U/µl DNaseI (RNase-free) (NEB, cat. no. M0303S)
0.5 M EDTA (Sigma, cat. no. E9884)
RNeasy MinElute Cleanup Kit (Qiagen, cat. no. 74204) containing:
    RNase-free water
20 µM RT primer GAT551 (5′-CTGGCTAACTAGGGAACCCACTGCT-3′)

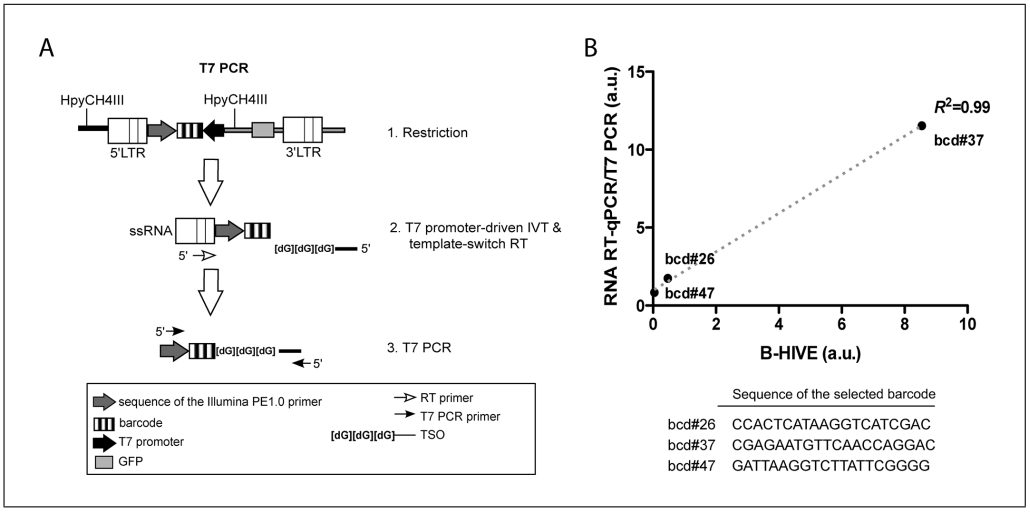

Figure 25C.2.4 Principle of template-switch T7 PCR. (**A**) Schematic representation of the key steps in template-switch T7 PCR. (1) The 5-cutter restriction enzyme HpyCH4III was chosen to fragment genomic DNA between 500 bp to 1 kb (see Support Protocol 1, steps 1 and 2). (2) *In vitro* ssRNA was transcribed from the T7 promoter inserted next to the barcode (see Support Protocol 1, steps 3 through 8). Template switch reverse transcription was further performed with the RT primer (GAT551) annealed on the HIV 5′LTR together with the template switching oligonucleotide (TSO) (GAT997) (see Support Protocol 1, steps 9 through 16). (3) A primer annealing to Illumina PE1.0 and the TSO were used for DNA amplification (see Support Protocol 1, steps 17 through 19). (**B**) The scatter plot shows individual HIV expression measured by B-HIVE against expression measured by template-switch T7 PCR (a.u: arbitrary units). Three individual barcodes were chosen to plot linear regression. Reproduced with permission from Chen et al., 2017.

100 µM template-switch oligonucleotide (TSO) GAT997 (5′-CAAGCAGAAGACGGCATACGAGATGGG-3′)
10 µM Primer GAT551 (5′-CTGGCTAACTAGGGAACCCACTGCT-3′)
10 µM Primer GAT519 (5′-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3′)
SuperScript™ II Reverse Transcriptase Kit (Invitrogen, cat. no. 18064022) containing:
5× First-Strand Buffer
0.1 M DTT
10 mM dNTPs (Thermo Fisher Scientific, cat. no. R0181)
40 U/µl RNasin Plus RNase Inhibitor (Promega, cat. no. N2611)
10 µM Primer GAT024 [sequence provided in Chen et al. (2017)]
10 µM Primer GAT998 (5′-CAAGCAGAAGACGGCATACGAGAT-3′)
Power SYBR Green PCR Master Mix (Applied Biosystems, cat. no. 4367659)
10 µM Single-barcode forward primer (see step 17)
10 µM Single-barcode reverse primer (see step 17)

PCR C100 Touch™ Thermal Cycler (Bio-Rad)
0.2-ml PCR tubes (Eppendorf, cat. no. 0030 124.332)
MicroAmp Optical 384-well Reaction Plate with barcode (Applied Biosystems, cat. no. 4309849)
ViiA 7 Real-Time PCR System (Applied Biosystems)

Digest genomic DNA

1. Digest 2 µg genomic DNA (see Basic Protocol 3, step 1) with 2 µl HpyCH4III in 1× CutSmart buffer in a 20 µl final volume for 3 hr at 37°C.
2. Heat-inactivate the reaction for 20 min at 65°C.

T7-promoter-driven in vitro transcription

3. *In vitro* transcribe 18 µl digested genomic DNA in a 0.2-ml PCR tube with 2 µl T7 RNA polymerase mix and 10 µl NTP buffer in a 30 µl final volume.

   Including a positive control (supplied in the kit) is recommended while performing in vitro transcription.

4. Incubate the reaction for 16 hr at 37°C.
5. Add 2 µl of 10 mM $CaCl_2$ and 1 µl DNaseI in the same reaction the next day.

   $CaCl_2$ is required for DNaseI to work in the in vitro transcription reaction.

6. Incubate the reaction for 30 min at 37°C.
7. Add 2 µl of 0.5 M EDTA and heat-inactivate the reaction for 10 min at 75°C.
8. Purify single-stranded RNA (ssRNA) with an RNeasy MinElute Cleanup Kit. Elute sRNA from the supplied filter column with 14 µl RNase-free water (supplied in the kit).

   The ssRNA can be validated by RT-PCR with 1.5 µl purified ssRNA and the RT primer GAT551; the protocol for RT-PCR is found at Basic Protocol 3, steps 29 through 36. The primers GAT551 and GAT519 are used for PCR amplification. The expected size of the PCR product is 196 bp.

Template-switch amplification

The template-switch protocol is modified from a previous report described by Salimullah, Sakai, Plessy, and Carninci (2011).

9. Prepare a template-switch reaction mix in a 0.2-ml PCR tube with the following ingredients in a 12 µl final volume:

   2 µl ssRNA
   1 µl of 20 µM RT primer GAT551
   8 µl of 100 µM TSO GAT997
   1 µl of 10 mM dNTPs.

10. Incubate the reaction for 5 min at 65°C and immediately place the reaction at 4°C.

11. Prepare the template-switch master mix with the following reagents:

    4 µl of 5× First-Strand Buffer
    2 µl of 0.1 M DTT
    1 µl of 40 U/µl RNasin Plus RNase Inhibitor.

    RNasin Plus RNase Inhibitor is not provided in the SuperScript™ II Reverse Transcriptase kit.

12. Add 7 µl master mix to denatured ssRNA.

13. Incubate the reaction mix for 2 min at 42°C.

14. Immediately add 1 µl SuperScript II reverse transcriptase in the same reaction.

15. Incubate the reaction for 90 min at 42°C.

16. Heat-inactivate for 10 min at 70°C.

    The cDNA products can be validated by PCR amplification with primers GAT024 and GAT998. Run the PCR under the following conditions: 98°C for 1 min; 98°C for 20 sec, 65°C for 1 min and 72°C for 1 min (30 cycles); and 72°C for 5 min. The expected size of the PCR product is 105 bp.

Quantify the expression of individual barcodes by quantitative PCR (qPCR)

17. Transfer 2 µl 10× diluted cDNA in step 16 with 5 µl of 2× Power SYBR Green PCR Master Mix, 0.3 µl of 10 µM single-barcode forward and reverse primers in a 10 µl final volume per well in a 384-well plate.

    The sequence of the single-barcode forward and reverse primers contains a partial sequence of a selected barcode.

    For example:

    Here is the sequence of the selected barcode: CCACTCATAAGGTCATCGAC [bcd#26 from Chen et al. (2017)]

    *The single-barcode forward primer (GAT920) is: 5′-CGCTCTTCCGATCT**CCACTCA**-3′*

    *The single-barcode reverse primer (GAT921) is: 5′-TACGACTCACTATA**GTCGATG**-3′*

    *The sequence **written in bold** hybridizes to the barcode directly.*

Select at least three individual barcodes to perform a linear regression in step 22 (see below).

18. Transfer 1 µl 10× diluted cDNA from Basic Protocol 3, step 34 with 5 µl of 2× Power SYBR Green PCR Master Mix, and 0.3 µl same single-barcode forward and reverse primers used in step 17 in a 10 µl final volume per well in a 384-well plate.

19. Run the qPCR under the following cycling conditions: 95°C for 10 min; and 95°C for 15 sec, 60°C for 1.5 min (40 cycles).

20. Normalize barcode expression with the value obtained from mRNA qPCR (step 18) divided by the value obtained from template-switch T7 qPCR (step 17).

21. Plot the value acquired in step 20 and its corresponding value obtained from B-HIVE in a scatterplot.

    The corresponding normalized RNA barcode counts from B-HIVE is acquired from Basic Protocol 4 step 17.

22. Perform the linear regression to validate barcode expression measured by B-HIVE (Figure 25C.2.4B).

    The measurements by B-HIVE and by T7 qPCR should lie on a straight line. A high value of the coefficient of determination (R^2) indicates that the B-HIVE measurements are as accurate as the template-switch T7 qPCR.

SUPPORT PROTOCOL 2

GENERATION OF THE RECOMBINANT HIV-BASED VECTOR pHCC1

In this protocol we detail the steps to generate the recombinant HIV-based vector pHCC1. The advantage of pHCC1 is due to its relative small size, which increases the efficiency of barcoding PCR described in Basic Protocol 1.

Materials

1 ng plasmid pT2-HSP-amaxaGFP (the plasmid is available upon request to the authors)
10 µM Primer GAT417 (5′-GATCGAATTCGGTACCCAGCTTTTGTTCCCGT-3′)
10 µM Primer GAT418 (5′-GATCGAATTCCAATTCGCCCTATAGTGAGTCG-3′)
2 U/µl Thermo Scientific Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, cat. no. F530S)
10 mM dNTPs (Thermo Fisher Scientific, cat. no. R0181)
100 pg pEV731 HIV-based vector (the plasmid is available upon request to the authors)
10 µM Primer GAT419 (5′-GATCGAATTCACTGCCAATTACCTAGTGGTT-3′)
10 µM Primer GAT420 (5′-GATCGAATTCCATGAACTACTACTGCTAGAGA-3′)
3% dimethyl sulfoxide (DMSO; NEB, cat. no. B0515A)
20,000 U/ml *Dpn*I (NEB, cat. no. R0176S)
10 U/µl *Eco*RI (Thermo Fisher Scientific, cat. no. ER0275)
5,000 U/ml Antarctic phosphatase (NEB, cat. no. M0289S)
MinElute PCR Purification Kit (QIAGEN, cat. no. 28004)
1% (w/v) agarose gel
QIAquick Gel Extraction Kit (QIAGEN, cat. no. 28704)
30 U/µl T4 DNA ligase (Thermo Fisher Scientific, cat. no. EL0013)
One Shot™ Stbl3™ Chemically Competent *E. coli* (Invitrogen, cat. no. C737303)
1 × SOC medium
LB plate supplemented with 50 µg/ml ampicillin
LB broth
50 mg/ml ampicillin
QIAprep Spin Miniprep Kit (QIAGEN, cat. no. 27104)

0.2-ml PCR tubes
PCR C100 Touch™ Thermal Cycler (Bio-Rad)

Generate the backbone and the insert of pHCC1

1. Amplify 1 ng pT2-HSP-amaxaGFP (referred to as backbone hereinafter) in a 0.2-ml PCR tube with the following ingredients in a 50 µl final volume:

   10 µl of 5× Phusion GC Buffer
   5 µl of 10 µM Primer GAT417
   5 µl of 10 µM Primer GAT418
   1 µl of 10 mM dNTPs
   0.5 µl Phusion DNA Polymerase
   1 ng pT2-HSP-amaxaGFP
   28.5 µl distilled water.

   Do not use more than 1 ng plasmid for PCR amplification in each reaction; excessive plasmid DNA decreases the efficiency of PCR amplification. Prepare two PCR reactions to have minimum 1 µg PCR products for following steps.

2. Run the PCR under the following conditions:

1 cycle:	1min	98°C (initial denaturation)
29 cycles:	20 sec	98°C (denaturation)
	1 min	58°C (annealing)
	1 min	72°C (extension)
1 cycle:	5 min	72°C (final extension).

3. Amplify 100 pg HIV-based vector pEV731 (referred to as insert hereinafter) in a 0.2-ml PCR tube with the following ingredients in a 50 µl final volume:

   10 µl of 5× Phusion GC Buffer
   5 µl of 10 µM Primer GAT419
   5 µl of 10 µM Primer GAT420
   1 µl of 10 mM dNTPs
   0.5 µl Phusion DNA Polymerase
   1.5 µl of 3% DMSO
   100 pg pEV731
   27 µl distilled water.

   Prepare two PCR reactions to have minimum 1 µg PCR products for following steps. DMSO is used to increase the PCR efficiency.

4. Run the PCR under the following conditions:

1 cycle:	1 min	98°C (initial denaturation)
29 cycles:	20 sec	98°C (denaturation)
	1 min	58°C (annealing)
	5 min	72°C (extension)
1 cycle:	5 min	72°C (final extension).

5. Add 1 µl *Dpn*I in the reaction containing the backbone.

   The same auxiliary information mentioned in Basic Protocol 1, step 3 applies here.

6. Incubate the reactions for 1 hr at 37°C.

EcoRI digests the backbone and the insert

7. Digest 1 µg backbone with 1 µl *Eco*RI in 1 × *Eco*RI buffer in a 30 µl final volume for 2 hr at 37°C.

8. Add 1 µl Antarctic phosphatase to the reaction from step 7.

9. Incubate the reaction for 1 hr at 37°C.
10. Heat-inactivate the reaction for 20 min at 65°C.
11. Purify *Eco*RI-digested backbone with a MinElute PCR Purification Kit following the manufacturer's instructions and elute the product in a 20 µl final volume.

Continue from step 16.

12. Digest 1 µg insert with 1 µl *Eco*RI in 1 × *Eco*RI buffer in a 30 µl final volume for 3 hr at 37°C.
13. Heat-inactivate the reaction for 20 min at 65°C.
14. Load all *Eco*RI-digested insert on a 1% (w/v) agarose gel.
15. Gel-purify the upper band (~4.6 bp) with the QIAquick Gel Extraction Kit and elute the product in a 20 µl final volume.
16. Measure the concentration of the backbone and the insert.

Ligate the backbone and the insert

17. Ligate 200 ng backbone and insert in 1 × T4 DNA ligase buffer with 2 µl T4 DNA ligase in a 20 µl final volume at 16°C for 18 hr.
18. Heat-inactivate the reaction for 10 min at 65°C.

Stbl3 transformation

19. Add 2.5 µl ligated products into 20 µl Stbl3.
20. Incubate the cells on ice for 30 min, followed by 42°C for 45 sec and on ice for 2 min.
21. Immediately add 90 µl 37°C prewarmed 1 × SOC into the cells.
22. Incubate the bacterial broth for 1 hr at room temperature.
23. Plate all bacterial broth on the ampicillin-containing LB plate.
24. Incubate the plate overnight at 37°C.
25. Pick up several colonies the next day.
26. Inoculate the colonies in 4 ml LB broth containing 50 µg/ml ampicillin to grow overnight.
27. Extract plasmid DNA from bacterial broth with the QIAprep Spin Miniprep Kit the next day.

Validate the pHCC1 construct

28. Digest 1 µg plasmid DNA with 1 µl *Eco*RI in 1 × *Eco*RI buffer in a 10 µl final volume at 37°C for 1 hr.
29. Load 5 µl digested products on a 1% (w/v) agarose gel to validate the construct of the ligated plasmid.

The size of two bands are 4.6 kb (insert) and 2.8 kb (backbone). The correct construct is referred to as pHCC1 hereinafter. pHCC1 will be used as the template to prepare the barcoded HIV library by barcoding PCR (see Basic Protocol 1).

BASIC DESIGN GUIDELINES OF A CUSTOM PIPELINE TO ANALYZE SEQUENCING DATA

In the protocol we show the logic of the analysis pipeline designed for the B-HIVE technology. Briefly, the first step is to determine the integration site of individual proviruses. According to the library design, valid sequenced reads must contain the barcode on the forward read and the sequence of the integration site at the reverse read. By mapping the sequenced genomic DNA sequences, we infer the provirus integration site at nucleotide resolution. Performing a robust association between barcode and insertion site is crucial to all subsequent analyses on an infected population. For materials, see Basic Protocol 3.

Guidelines to validate the read structure from paired-end sequencing (see Basic Protocol 3, step 49a)

1. Perform an inexact search of the T7 promoter sequence at the forward read.
2. Remove the matched sequence and keep the beginning of the read to obtain the barcode.
3. Discard reads without match.

   T7 sequence: TATAGTGAGTCGTATTA; recommended 0 to 2 mismatches for inexact search.

4. Perform an inexact search of the provirus LTR at the reverse read.
5. Remove the matched sequence and keep the rest of the read to obtain the genomic DNA sequence of the integration site.
6. Discard reads without match.

   Reverse complement of the LTR sequence: AGCCCTTCCA; recommended 0 to 2 mismatches.

7. Store pairs of genomic DNA and barcode from reads matching on both ends.

Guidelines to cluster barcodes

8. Cluster the extracted barcodes using Starcode (Zorita, Cuscó, & Filion, 2015); use a matching distance between 1 and 3.
9. Replace the barcodes on the list by their canonicals.

Guidelines to map genomic DNA sequences

10. Map the genomic DNA sequences on a Human Genome reference.
11. Preserve the correspondence between barcode and genomic DNA.

    It is recommended to use the barcode in place of the read name.

12. Discard reads with low mapping quality.
13. Update the list by replacing the genomic DNA sequence with the corresponding integration locus.

Guidelines to filter out PCR artefacts

14. Compute the list of associated barcodes for each distinct integration site.
15. Compute the list of associated integration sites for each distinct barcode.
16. Finally, for each integration site/barcode association, accept only if:

    i. the interaction has at least k observations.
    ii. the rate of interaction of the barcode with the integration site is greater than r_1.

iii. the rate of interaction of the integration site with the barcode is greater than r_2.

Recommended values: $k \geq 5$, $r_1 \geq 0.9$, $r_2 \geq 0.9$.

DNA and RNA barcode extraction and clustering

The expression of each individual provirus is inferred with the ratio between RNA and DNA barcode read counts after single-read sequencing (see Basic Protocol 3, step 49b).

17. Perform an inexact search of the T7 promoter sequence at the DNA and RNA reads, respectively.

18. Remove the matched sequence and keep the beginning of the read to obtain the barcode.

19. Discard reads without match.

    T7 sequence: TATAGTGAGTCGTATTA; recommended 0 to 2 mismatches for inexact search.

20. Cluster DNA and RNA barcodes separately using Starcode; use a matching distance between 1 and 3.

Distribution of provirus expression landscape

21. Merge the RNA and DNA set and keep only the barcodes present in both sets.

22. Compute the expression score of the *i*-th provirus as:

$$s_i = \frac{R_i \sum_j D_j}{D_i \sum_j R_j}$$

where D_j and R_j are the counts of DNA and RNA of the *j*-th barcode, respectively.

This step is used to reduce batch effects.

Guidelines to reveal single provirus transcriptomics

23. Merge the previously merged list with the integration site list.

    Only these barcodes present in both lists shall be kept.

24. Compute the score as in 23 to obtain the expression score of each integration site.

REAGENTS AND SOLUTIONS

Use deionized, distilled water in all recipes (unless otherwise indicated).

Phosphate buffer, 100×

3.5 ml of 1 M $NaH_2PO_4.H_2O$ (final conc. 70 mM)

$CaCl_2$ solution, 2 M

11.1g $CaCl_2$
Nuclease-free water to a 50 ml final volume
Filter-sterilize through a 0.22-µm filter
Store up to 6 months at 4°C

HBS buffer, 2×

3.5 ml of 1 M Na_2HPO_4,H_2O (final conc. 70 mM)
Nuclease-free water to a 50 ml final volume
Adjust the pH value to pH 6.8

Filter sterilize through a 0.22-μm filter
Divide into 10-ml aliquots
Store aliquots up to 6 months at −20°C

Wrong pH value affects transfection efficiency

COMMENTARY

Background Information

Molecular barcodes have a long history in yeast genetics (Goffeau, 2000), but their first use for systematic tracking of gene expression in mammals came with "Thousands of Reporters Integrated in Parallel" (TRIP, Akhtar et al., 2013), a method to study position effects in transgene insertion. The focus of TRIP is to understand how the genome affects the expression of genes delivered through a neutral vector, whereas B-HIVE aims to track individual viral parasites in their host. Building on similar principles, several steps were specifically redesigned for B-HIVE, including 1. the barcoding PCR strategy, 2. the design of the library, 3. the dual DNA and RNA extraction and 4. the data analyses *in silico*.

Critical Parameters

The protocols are based on experimental results set up in Jurkat T cells infected with pHCC1-based vectors (derived from the plasmid pEV731). Several parameters need to be adjusted when working with different biological material.

Design of the library

Two key points for the library design are the location of the barcode and the additional elements included together with the barcode. We recommend inserting the barcode close to the HIV long terminal repeat at either the 5′ or the 3′ end. This will make it easier to find restriction enzyme sites that are not present between the barcode and the end of the provirus, and it will also make the restriction fragments shorter, which will increase the efficiency of the self-ligation (see Basic Protocol 3, step 6). Also, the exact location of the barcode should not disrupt any feature of interest in the virus, such as an important open reading frame.

Along the same lines, it is important that the restriction enzyme site is close to the barcode (ideally 8-16 nucleotides). This is typically achieved by inserting the site together with the barcode. The DNA between the barcode and the restriction enzyme site is present in all the proviruses, so it triggers template switching (Meyerhans, Vartanian, & Wain-Hobson, 1990) between the amplicons during the PCR. This artefact shuffles the barcodes between insertions sites so it must be avoided. The most efficient solution is to keep this fragment short to reduce the probability of template switching. Leaving a few nucleotides has the advantage of keeping a "watermark," i.e., a known sequence after the barcode to facilitate identification and quality control. The chosen restriction enzymes should be frequent in the genome, but absent from the fragment between the barcode and the end of the provirus. We recommend performing single enzyme digestion for restriction enzyme sites of 4 base pairs, and double enzymes digestion for restriction enzyme sites of 5 nucleotides or more.

Including the sequence of the Illumina sequencing primer P5 in the construct is optional. The advantage is that the PCR and RT-PCR products can be sequenced without further steps of library preparation. Moreover, it also allows quantifying barcode abundance by T7 PCR (see Support Protocol 1).

Complexity of barcoded HIV libraries

The key to using barcodes as universal identifiers for tracing HIV after infection is to have a very large initial collection of barcodes. In our published work (Chen et al., 2017), the complexity of the barcoded library was estimated at 1.3 million clones. Out of 20,000 FACS-sorted GFP(+) founder cells, approximately half maintain a long-term infection; thus, barcodes were drawn on average 0.0076 times from a pool of 1.3 million (10,000/1,300,000 = 0.0076). The probability that two viruses carry the same barcode is approximately equal to Prob (X > 1), where X has a Poisson distribution with mean 0.0076. This probability is about 2.87e-5, so duplications are negligible in the condition of infection described in this protocol.

In general, duplications should be kept at a frequency below 0.1%. This means that the number of barcodes in the pool should never exceed 5% of the size of the library (the number of independent clones). We consider this number a borderline case and we simply recommend producing libraries with at least one million independent clones and not creating pools of more than 50,000 barcoded viruses. Also note that imbalance in the representation of the barcodes in the library increases

the risk that two viruses carry the same barcode. Amplifying existing libraries can thus have severe consequences on the diversity of the barcodes, so we recommend preparing new libraries when the stock is exhausted.

Establishment and expansion of founder cell pools

In addition to reducing barcode diversity, establishing founder cell pools ensures that every cell in the founder population is infected with at least one active virus. Moreover, we recommend expanding founder cells at least two weeks to a size of several million cells; so that each infected cell can divide several times to obtain abundant copies of the same barcoded virus (the doubling time of Jurkat T cells is around 48 hr).

Inverse PCR

The amplification method carried out in this protocol is to perform two rounds of inverse PCR with two different pairs of primers in order to increase the yield. We use relatively high melting temperature and limited numbers of PCR cycles to avoid non-specific PCR amplification. In principle, the total number of PCR cycles should not exceed 30. We also use no more than 0.1 μM final concentration of each primer to reduce mispriming and primer dimers.

Library quantification

We recommend quantifying libraries by qPCR with the KAPA library quantification kit. Because the primers provided in this kit anneal to the Illumina sequencing primer sequences, the qPCR output reflects the amount of material that can be sequenced.

Bioinformatic pipelines

The BHIVE bioinformatics pipeline which is available on the accompanying virtual machine (see Internet Resources).

Troubleshooting

Troubleshooting problems and possible solutions are found in the Table 25C.2.1.

Understanding Results

Visualization of libraries of mapping, RNA barcode and DNA barcode amplification

A successful mapping library presents a smear with the expected size between 200 to 700 bp on a 2% (w/v) agarose gel (Fig. 25C.2.3E). High density peaks lie between 500 to 700 bp on a bioanalyzer profile (Fig. 25C.2.3H). A specific PCR product with the expected size of 126 bp is detected on a 2% (w/v) agarose gel from the library of RNA barcode amplification (Fig. 25C.2.3F) and DNA barcode amplification (Fig. 25C.2.3G), respectively. A specific peak is also present on a bioanalyzer profile (data not shown).

Time Considerations

The time for Basic Protocol 1 partly depends on the experimenter. A barcoded library can usually be accomplished within a week. The transfection carried out in Basic Protocol 2 (steps 2 through 18) can be accomplished within in a week once cells are ready-to-use. Two harvests of viral supernatant can be stored up to 2 days at 4°C before concentrating viral stocks (see Basic Protocol 2, steps 12 through 18).

The whole procedure of viral infection and isolation of latently infected cells requires about 1 month (steps 19 through 27). On day 4 post-infection, it takes 1 day to FACS-sort GFP(+) cells (step 24) and immediately distribute the cells to several small pools of 20,000 infected founder cells (step 25). The cell pool is then expanded in culture for at least two weeks so that each founder cell produces enough clones with identical proviruses (step 26). It takes 1 day to isolate latently infected GFP(−) cells from this cell pool by FACS-sorting (step 27). Cells are further cultured for a week followed by genomic DNA and RNA extraction.

Assuming that genomic DNA and mRNA are ready to use in Basic Protocol 3, it takes 3 days to prepare a library for mapping: on the first day the restriction enzyme digestion of genomic DNA requires 3 hr 20 min (steps 1 through 3), blunt-ending digested genomic DNA requires 40 min (steps 4 and 5), and self-ligation is carried out overnight (step 6). On the second day the precipitation of self-ligated products requires ∼2 hr (steps 7 through 12), the Plasmid-Safe DNase digestion requires 2.5 hr (steps 13 through 17) and the *Sac*I digestion requires 3 hr 20 min (steps 18 and 19). On the third day the first (steps 20 and 21) and second (steps 22 through 24) rounds of inverse PCR require ∼1.5 and ∼2.5 hr, respectively. It takes 1 day to amplify RNA barcodes: the mRNA purification requires 1 hr (step 28), the RT requires 1.5 hr (steps 29 through 34) and the PCR amplification requires ∼1 hr (steps 35 and 36). The amplification of DNA barcodes requires ∼1 hr (steps 39 and 40). Purifying the sequencing library with the QIAquick PCR purification kit requires ∼45 min (steps 27, 38, and 43). The time for library quality control (steps 44 through 46) can be achieved in a day and the time for Illumina high-throughput sequencing (steps 47 through

Table 25C.2.1 Troubleshooting Guide for B-HIVE

Step	Problem	Possible	Solution
Basic Protocol 1			
21	An arc occur	Too much plasmid DNA given	Dilute plasmid DNA for electroporation
Basic Protocol 3			
25, 37, 41	Non-specific PCR smears/products shown on the gel	PCR pipettes contaminated by the HIV-based vector	Clean PCR pipettes by 10% (v/v) bleach
		PCR reagents contaminated	Replace with new PCR reagents
Problems detected after data analyses			
Basic Protocol 1			
28	High barcode collisions occurred	Low complexity of the barcoded library	Check the complexity of the barcoded library which is comprised of at least one million barcodes
Basic Protocol 2			
25	Low numbers of mapped integration sites	The number of the founder cell pool is not enough	Increase the number of the founder cell pools
25, 26	Low match between barcodes retrieved from mapping samples and RNA barcode amplification	Divisions of the founder cells are not sufficient	Check the number of the cells used to establish a founder cell pool and/or draw the number of the founder cells down to 10,000 cells per pool
Basic Protocol 3			
39	High variation of normalized RNA barcode expression	DNA barcodes amplified by PCR are not enough to cover all barcodes in the presence of a founder cell pool	Perform more reactions for DNA barcode amplification

49) varies depending on the conditions of the run and the waiting time before processing (if applicable).

Acknowledgments

We would like to thank Dr. Andreas Meyerhans and Dr. Javier Martinez for fruitful interactions, and Dr. Albert Jordan for providing research material. We would like to thank the genomics core facility and the 4D unit of the CRG for their technical support. We acknowledge the support of the Spanish Ministry of Economy and Competitiveness, "Centro de Excelencia Severo Ochoa 2013-2017." We acknowledge the support of the CERCA Programme / Generalitat de Catalunya. This work was funded by the Spanish Ministry of Economy and Competitiveness (Plan Nacional BFU2012-37168), and by the European Research Council (Synergy Grant 609989).

Conflict of Interest

The authors declare no conflicts of interest.

Literature Cited

Akhtar, W., de Jong, J., Pindyurin, A. V., Pagie, L., Meuleman, W., de Ridder, J., ... van Steensel, B. (2013). Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell*, *154*, 914–927. doi: 10.1016/j.cell.2013.07.018.

Archin, N. M., Liberty, A. L., Kashuba, A. D., Choudhary, S. K., Kuruc, J. D., Crooks, A. M., ... Margolis, D. M. (2012). Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. *Nature*, *487*, 482–485. doi: 10.1038/nature11286.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive

epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, *10*, 1213–1218. doi: 10.1038/nmeth.2688.

Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current Protocols in Molecular Biology*, *109*, 21.29.1–21.29.9. doi: 10.1002/0471142727.mb2129s109.

Bullen, C. K., Laird, G. M., Durand, C. M., Siliciano, J. D., & Siliciano, R. F. (2014). New ex vivo approaches distinguish effective and ineffective single agents for reversing HIV-1 latency in vivo. *Nature Medicine*, *20*, 425–429. doi: 10.1038/nm.3489.

Chen, H.-C., Martinez, J. P., Zorita, E., Meyerhans, A., & Filion, G. J. (2017). Position effects influence HIV latency reversal. *Nature Structural & Molecular Biology*, *24*, 47–54. doi: 10.1038/nsmb.3328.

Cillo, A. R., Sobolewski, M. D., Bosch, R. J., Fyne, E., Piatak, M. Jr., Coffin, J. M., & Mellors, J. W. (2014). Quantification of HIV-1 latency reversal in resting CD4+ T cells from patients on suppressive antiretroviral therapy. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 7078–7083. doi: 10.1073/pnas.1402873111.

Ciuffi, A., Llano, M., Poeschla, E., Hoffmann, C., Leipzig, J., Shinn, P., ... Bushman, F. (2005). A role for LEDGF/p75 in targeting HIV DNA integration. *Nature Medicine*, *11*, 1287–1289. doi: 10.1038/nm1329.

Debyser, Z., Christ, F., De Rijck, J., & Gijsbers, R. (2015). Host factors for retroviral integration site selection. *Trends in Biochemical Sciences*, *40*, 108–116. doi: 10.1016/j.tibs.2014.12.001.

Finzi, D., Blankson, J., Siliciano, J. D., Margolick, J. B., Chadwick, K., Pierson, T., ... Siliciano, R. F. (1999). Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nature Medicine*, *5*, 512–517. doi: 10.1038/8394.

Goffeau, A. (2000). Four years of post-genomic life with 6,000 yeast genes. *FEBS Letters*, *480*, 37–41. doi: 10.1016/S0014-5793(00)01775-0.

Meyerhans, A., Vartanian, J. P., & Wain-Hobson, S. (1990). DNA recombination during PCR. *Nucleic Acids Research*, *18*, 1687–1691. doi: 10.1093/nar/18.7.1687.

Mueller, P. R., & Wold, B. (1989). In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. *Science*, *246*, 780–786. doi: 10.1126/science.2814500.

Ochman, H., Ajioka, J. W., Garza, D., & Hartl, D. L. (1989). Inverse Polymerase Chain Reaction. In H. A. Erlich (Ed.), *PCR Technology* (pp. 105–111). Palgrave Macmillan UK: Springer.

Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., ... Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, *159*, 1665–1680. doi: 10.1016/j.cell.2014.11.021.

Salimullah, M., Sakai, M., Plessy, C., & Carninci, P. (2011). NanoCAGE: A high-resolution technique to discover and interrogate cell transcriptomes. *Cold Spring Harbor Protocols*, *2011*, pdb.prot5559. doi: 10.1101/pdb.prot5559.

Schmidt, M., Schwarzwaelder, K., Bartholomae, C., Zaoui, K., Ball, C., Pilz, I., ... von Kalle, C. (2007). High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). *Nature Methods*, *4*, 1051–1057. doi: 10.1038/nmeth1103.

Schröder, A. R., Shinn, P., Chen, H., Berry, C., Ecker, J. R., & Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell*, *110*, 521–529. doi: 10.1016/S0092-8674(02)00864-4.

Spina, C. A., Anderson, J., Archin, N. M., Bosque, A., Chan, J., Famiglietti, M., ... Planelles, V. (2013). An in-depth comparison of latent HIV-1 reactivation in multiple cell model systems and resting CD4+ T cells from aviremic patients. *PLoS Pathogens*, *9*, p.e1003834. doi: 10.1371/journal.ppat.1003834.

Triglia, T., Peterson, M. G., & Kemp, D. J. (1988). A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Research*, *16*, 8186. doi: 10.1093/nar/16.16.8186.

Wang, G. P., Ciuffi, A., Leipzig, J., Berry, C. C., & Bushman, F. D. (2007). HIV integration site selection: Analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Research*, *17*, 1186–1194. doi: 10.1101/gr.6286907.

Worby, C. A. and Dixon, J. E. 2004. RNA Interference in Cultured *Drosophila* Cells. *Current Protocols in Molecular Biology*, *65:26.5*:26.5.1–26.5.7.

Zorita, E., Cuscó, P., & Filion, G. J. (2015). Starcode: Sequence clustering based on all-pairs search. *Bioinformatics*, *31*, 1913–1919. doi: 10.1093/bioinformatics/btv053.

Internet Resources

www.unaids.org/en/resources/fact-sheet

Official updated statistic values of HIV epidemic.

https://lab.thegrandlocus.com/protocols/barcoding-pcr

Optimized protocol of barcoding PCR for the general use for every chosen plasmid.

https://github.com/gui11aume/starcode

Tutorial and Starcode scripts.

https://hub.docker.com/r/gui11aume/bhive/

BHIVE bioinformatics pipeline.

pHCC1 plasmid map and annotation key

Blue	5'LTR
Red	*tat* transcriptional activator
Green	GFP reporter
Yellow	3'LTR
Gray	backbone

TGGAAGGGCTAATTCACTCCCAAAGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTG
ATTAGCAGAACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTG
AGCCAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCATGGGATGGAT
GACCCGGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCGAGAGCTGCATCC
GGAGTACTTCAAGAACTGCTGATATCGAGCTTGCTACAAGGGACTTTCCGCTGGGGACTTTCCAGGGAGGCGTGGCCT
GGGCGGGACTGGGGAGTGGCGAGCCCTCAGATCCTGCATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGT
TAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAG
TGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAA
AATCTCTAGCAGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGACGCAGGACTCG
GCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAATTTTGACTAGCGGAGGCT
AGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGCGATGGGAAAAAATTCGGTT
AAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTT
AATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATC
AGAAGAACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCA
AGGAAGCTTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGACCACCGCACAGCAAGCGGCCGCTGATCTTCA
GACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTA
GGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCC
TTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGACGGTACAGGCCAGACAATTATTG
TCTGGTATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGG
GGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATTTGGGG
TTGCTCTGGAAAACTCATTTGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGG
AATCACACGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACTCCTTAATTGAAGAATC
GCAAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAACA
TAACAAATTGGCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGT
ACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCCACCTCCCAACCCCGAGGGGACC
CGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCGATTAGTGAACGGATCT
CGACGGTATCGATGCCACCATGGAGCCAGTAGATCCTAGACTAGAGCCCTGGAAGCATCCAGGAAGTCAGCCTAAAA
CTGCTTGTACCAATTGCTATTGTAAAAAGTGTTGCTTTCATTGCCAAGTTTGTTTCATAACAAAAGCCTTAGGCATCTC
CTATGGCAGGAAGAAGCGGAGACAGCGACGAAGACCTCCTCAAGGCAGTCAGACTCATCAAGTTTCTCTATCAAAGC
AACCCACCTCCCAATCCCGAGGGGACCCGACAGGCCCGAAGGAATCGAAGAAGAAGGTGGAGAGAGAGACAGAGAC
AGATCCATTCGATGGAGGCGACTACAAGGACGACGATGACAAGTAGAAGCTTGGTACCGAGCTCGGATCCGCCCCTC
TCCCTCCCCCCCCCCTAACGTTACTGGCCGAAGCCGCTTGGAATAAGGCCGGTGTGCGTTTGTCTATATGTTATTTTCC
ACCATATTGCCGTCTTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTTCTTGACGAGCATTCCTAGGGGTCTTT
CCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGCAGTTCCTCTGGAAGCTTCTTGAAGACAAA
CAACGTCTGTAGCGACCCTTTGCAGGCAGCGGAACCCCCCACCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTG
TATAAGATACACCTGCAAAGGCGGCACAACCCCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGG
CTCTCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCAGAAGGTACCCCATTGTATGGGATCTGATCTGGGGCCT
CGGTACACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAAACGTCTAGGCCCCCCGAACCACGGGGACGTGGTTTTC
CTTTGAAAAACACGATGATAATATGGCCACAACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCA
TCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTAC
GGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACC
TACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGC
TACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGA
CACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGT
ACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGC
CACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCC
TGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGGCCTCGAGACCTAG
AAAAACATGGAGCAATCACAAGTAGCAATACAGCAGCTACCAATGCTGATTGTGCCTGGCTAGAAGCACAAGAGGAG
GAGGAGGTGGGTTTTCCAGTCACACCTCAGGTACCTTTAAGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCAC
TTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTAC
CACACACAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGG
TGCTACAAGCTAGTACCAGTTGAGCAAGAGAAGGTAGAAGAAGCCAATGAAGGAGAGAACACCCGCTTGTTACACCC
TGTGAGCCTGCATGGGATGGATGACCCGGAGAGAGAAGTATTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATC
ACATGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGACTTTCCGCTGGGG
ACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATGCTGCATATAAGCAGCTGCTTTTT
GCTTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAG

Figure 25C.2.5

CCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCA
GACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTAGTAGTTCATGGAATTCGGTACCCAGCTTTTGTTCCCGTAACTA
TAACGGTCCTAAGGTAGCGAACTTTAGTGAGGGTTAATTTCGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGT
GAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGA
GTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAAT
GAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCT
CGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAAC
GCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCA
TAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
GATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGC
CTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCC
AAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAAC
CCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGC
TACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCC
AGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGC
AAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGG
AACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAA
TGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCT
ATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCT
TACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGC
CAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAG
CTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTC
GTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGC
GGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTG
CATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAAT
AGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAG
TGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAAC
CCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAA
ATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCA
TTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCA
CATTTCCCCGAAAAGTGCCACCTGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCG
TGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTT
CCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTG
ATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTT
TAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTG
CCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCTT
ACAATTTCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGC
TGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACG
ACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGAATTCACTGCCAATTACCTAGTGGTTTCATTT
ACTCTAAACCTGTGATTCCTCTGAATTATTTTCATTTTAAAGAAATTGTATTTGTTAAATATGTACTACAAACTTAGTAG
T

Figure 25C.2.5 *Continued.*