Published in final edited form as:

Methods Mol Biol. 2016 ; 1354: 119–131. doi:10.1007/978-1-4939-3046-3_8.

Analysis of HIV-1 Gag-RNA interactions in cells and virions by CLIP-seq

Sebla B. Kutluay^{1,†,*} and Paul D. Bieniasz^{1,2,*}

¹Laboratory of Retrovirology, Aaron Diamond AIDS Research Center, The Rockefeller University, 455 First Avenue, New York, NY 10016, USA.

²Howard Hughes Medical Institute, Aaron Diamond AIDS Research Center, The Rockefeller University, 455 First Avenue, New York, NY 10016, USA.

Summary

Next-generation sequencing-based methodologies have revolutionized the analysis of protein-nucleic acid complexes, yet these novel approaches have rarely been applied in virology. Because it has an RNA genome, RNA-protein interactions play critical roles in human immunodeficiency virus type 1 (HIV-1) replication. In many cases, the binding sites of proteins on HIV-1 RNA molecules in physiologically relevant settings are not known. Crosslinking-immunoprecipitation sequencing (CLIP-seq) methodologies, which combine immunoprecipitation of covalently crosslinked protein-RNA complexes with high-throughput sequencing, is a powerful technique that can be applied to such questions as it provides a global account of RNA sequences bound by a RNA-binding protein of interest in physiological settings at near-nucleotide resolution. Here, we describe the application of the CLIP-seq methodology to identify the RNA molecules that are bound by the HIV-1 Gag protein in cells and in virions. This protocol can easily be applied to other viral and cellular RNA-binding proteins that influence HIV-1 replication.

Keywords

HIV-1; Gag; RNA packaging; RNA-binding protein; protein-RNA interaction; CLIP-seq; UV crosslinking; next-generation sequencing; cells; virions; bioinformatics

1. Introduction

Viral and host RNA-binding proteins regulate all major stages of HIV-1 replication, including transcription, splicing and export of viral mRNAs, assembly of infectious virions and reverse transcription. HIV-1 Gag is one such viral RNA-binding protein that coordinates all major steps in virion assembly, including the selective packaging of a dimeric, unspliced viral RNA genome (1–3).

HIV-1 genomic RNA packaging has long been thought to be driven by the binding of the nucleocapsid (NC) domain of Gag to a *cis*-acting packaging element, psi (Ψ), located within

*Corresponding authors: Sebla B. Kutluay (skutluay@wusm.wustl.edu), Paul D. Bieniasz (pbienias@adarc.org).

†Current address: Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO 63110 USA.

the 5' leader of the viral genome (4–7). However, several lines of evidence indicate that Gag-Ψ interaction is not sufficient for packaging and that Gag-RNA interactions are more complex. First, knowledge of the viral RNA sequences that are directly bound by Gag has largely been inferred from genetic studies and limited *in vitro* data. To date, no assay has been able to demonstrate a direct and specific interaction between Ψ and Gag protein in a relevant context, i.e. in live cells and in virions. Second, deletion of Ψ does not completely abolish genome encapsidation (8–10), suggesting that other regions on the viral genome may contribute. Third, Gag undergoes several changes in localization (11,12), multimerization state (12) and is proteolytically processed during particle genesis. It is completely unknown how these changes affect RNA-binding properties of Gag. Fourth, Gag is also thought to promiscuously bind to and package cellular RNAs in proportion to their abundance in the cytosol (13,14). Whether these interactions take place in cells and virions, and if so, how they change during particle genesis is largely unknown.

To identify the RNA molecules directly bound by Gag in physiological settings, we have recently applied CLIP-seq methodologies (15,16) to various stages of particle genesis (17). One CLIP-seq approach that is referred to as photoactivatable-ribonucleoside-enhanced-CLIP (PAR-CLIP) relies on the incorporation of photoreactive ribonucleoside analogs, such as 4-thiouridine (4-SU) and 6-thioguanosine (6-SG) into nascent RNAs in live cells. Exposure of cells to UV light at 365 nm wavelength prior to cell lysis induces the covalent crosslinking of the RNA-binding proteins to their target RNA molecules primarily at these 4-SU and 6-SG-modified sites. Following cell lysis and limited RNase digestion, protein-RNA adducts are immunopurified and the RNA molecules, often about 15–50 nucleotides long, are isolated. Following sequential adapter ligations, RNA is reverse-transcribed, the resulting cDNA is PCR amplified and deep sequenced using the Solexa technology. One distinct advantage of the PAR-CLIP methodology is the introduction of T-to-C (for 4-SU) or G-to-A (for 6-SG) mutations during reverse-transcription, which defines the precise sites within target RNA molecules that are crosslinked to the RNA-binding protein of interest.

As 4-SU/6-SG-mediated crosslinking is more efficient than conventional UV-crosslinking, PAR-CLIP yields more abundant protein-RNA complexes, which can be critical for the success (i.e. high signal to background ratios) of a given CLIP-seq experiment. However, one potential disadvantage of the PAR-CLIP approach is the potential for alteration of RNA structure by incorporation of the ribonucleoside analogs. Therefore, it is worthwhile to perform CLIP-seq experiments in which protein-RNA crosslinks are induced by conventional UV-crosslinking for confirmatory purposes (i.e. HITS-CLIP). In addition, it is important to validate results obtained from a given experiment by using a different immunoprecipitating antibody, varying the ribonucleoside analog and the RNase. Overall, we think that this protocol will not only provide a useful tool for analysis of HIV-1 Gag-RNA interactions, but it can also be adapted to other viral and cellular RNA-binding proteins that regulate the replication of viruses.

2. Materials

2.1 UV-crosslinking, lysis and RNase treatment components

1. 4-thiouridine (Sigma-Aldrich Chemical Company, St. Louis, MO, USA): dissolve in water to a final concentration of 100 μ M.
2. Stratalinker 1800/2400 or an equivalent UV-crosslinking chamber equipped with UV365nm bulbs.
3. Phosphate-buffered saline (PBS), without calcium and magnesium.
4. Ultracentrifuge tubes.
5. 20% sucrose solution (w/v): Prepare in 1 $\times$ PBS, filter and store at 4 $^{\circ}$ C.
6. Protease inhibitor cocktail.
7. NP-40 lysis buffer: 50 mM HEPES, pH 7.5, 150 mM KCl, 2 mM EDTA, 0.5% NP-40, supplemented with 0.5 mM DTT and protease inhibitor cocktail (*see* Note 1).
8. RIPA buffer: 50 mM Tris pH7.4, 1% NP-40, 0.25% Na-deoxycholate, 0.1% SDS, 150 mM NaCl, 1mM EDTA, supplemented with 0.5 mM DTT and protease inhibitor cocktail.
9. RNase A.
10. DNase I.

2.2 Immunoprecipitation, alkaline phosphatase treatment, end-labeling components

1. Citrate-phosphate buffer: 4.7 g/L citric acid, 9.2 g/L Na₂HPO₄, pH 5.0.
2. Protein G magnetic beads.
3. Magnetic stand.
4. Calf intestinal alkaline phosphatase.
5. T4 Polynucleotide kinase.
6. ATP, [γ -³²P], 3000Ci/mmol, 10mCi/ml.
7. ATP.
8. Low-retention microcentrifuge tubes.
9. IP wash buffer: 50 mM HEPES-KOH, pH 7.5, 300 mM KCl, 0.05% NP-40, supplemented with 0.5 mM DTT.
10. LiCl buffer: 250 mM LiCl, 10 mM Tris pH 8.0, 1 mM EDTA, 0.5% NP-40, 0.5% Na-deoxycholate, supplemented with 0.5 mM DTT.
11. NaCl buffer: 50 mM Tris pH 7.4, 1 M NaCl, 1 mM EDTA, 0.1% SDS, 0.5% Na-deoxycholate, 1% NP-40, supplemented with 0.5 mM DTT.

¹·Sigma-Aldrich has replaced NP-40 with Igepal-CA 630, which works equally well in our hands.

12. KCl buffer: 50 mM HEPES-KOH, pH 7.5, 500 mM KCl, 0.05% NP-40, supplemented with 0.5 mM DTT.
13. Dephosphorylation buffer: 50 mM Tris-HCl, pH 7.9, 100 mM NaCl, 10 mM MgCl₂, supplemented with 1 mM DTT.
14. Phosphatase wash buffer: 50 mM Tris-HCl, pH 7.5, 20 mM EGTA, 0.5% NP-40, supplemented with 1 mM DTT.
15. PNK buffer: 50 mM Tris-HCl, pH 7.5, 50 mM NaCl, 10 mM MgCl₂, supplemented with 1 mM DTT.
16. Protein sample buffer (4x): 0.5 M Tris, 1.6 mM EDTA, 8% SDS, 40% glycerol, 0.002% bromophenol blue. Adjust pH to 8.5.
17. Thermal mixer.

2.3 RNA purification components

1. 4–12% Bis-Tris protein gels.
2. MOPS SDS Running buffer (20X): 50 mM MOPS, 50 mM Tris Base, 0.1% SDS, 1mM EDTA. Adjust pH to 7.7.
3. Nitrocellulose membrane.
4. Tris-Glycine transfer buffer (10x): 250 mM Tris, 1.92 M glycine. Prepare 1x buffer containing 20% ethanol.
5. Autoradiography cassettes and film.
6. Proteinase K, recombinant, PCR Grade.
7. Proteinase K buffer (2x): 200 mM Tris-HCl, pH 7.5, 100 mM NaCl, 20 mM EDTA, 100 mM NaCl, 2% SDS.
8. Glycoblu co-precipitant (Life Technologies, Carlsbad, CA, USA).
9. 3M Sodium acetate, pH 5.2.
10. Ethanol:isopropanol (1:1).
11. Acid phenol:chloroform: isoamyl alcohol (125:24:1).

2.4 Adapter ligations and library preparation components

1. 80% ethanol.
2. Nuclease free water.
3. RNase inhibitor.
4. Pure BSA.
5. DMSO.
6. 50% PEG8000.

7. T4 RNA Ligase 2, truncated K227Q (New England Biolabs, Ipswich, MA, USA).
8. T4 RNA Ligase 1 (New England Biolabs).
9. 6% and 15% TBE-Urea gels.
10. TBE-Urea sample buffer (2X): 45 mM Tris, 45 mM boric acid, 1 mM EDTA (free acid), 6% Ficoll type 400, 3.5 M Urea, 0.005% bromophenol blue, 0.025% xylene cyanol.
11. TBE running buffer (10X): 890 mM Tris, 890 mM boric acid, 20 mM EDTA, pH 8.3.
12. Sterile centrifuge tube filters with cellulose acetate membrane (pore size 0.22 µm).
13. Low Molecular Weight Marker (range of 10–100 nt).
14. SuperScript ® III First-Strand Synthesis System (Life Technologies).
15. High-fidelity DNA Polymerase.
16. Low Molecular Weight DNA Ladder (range of 50–500 nt).
17. Diffusion buffer: 0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, pH 8.0, 0.1% SDS.
18. DNA gel extraction kit.
19. Adapters and primer pairs:

3' adapter: 5'-adenylated/TCG TAT GCC GTC TTC TGC TTG-3' dideoxyC

5' barcoded adapters:

5'-rGUU CAG AGU UCU ACA GUC CGA CGA UC AGU NNN UC-3'

5'-rGUU CAG AGU UCU ACA GUC CGA CGA UC GAU NNN UC-3'

5'-rGUU CAG AGU UCU ACA GUC CGA CGA UC GUG NNN UC-3'

5'-rGUU CAG AGU UCU ACA GUC CGA CGA UC ACG NNN UC-3'

5'-rGUU CAG AGU UCU ACA GUC CGA CGA UC UAG NNN UC-3'

5'-rGUU CAG AGU UCU ACA GUC CGA CGA UC AUC NNN UC-3'

Positive control RNA oligo: 5'-rAUAGCUACGAUUGCA-3'

RT/Reverse PCR primer: 5'-CAAGCAGAAGACGGCATACTGA-3'

Forward PCR primer:

5'-AATGATACGGCGACACCGACAGGTTTCAGAGTTCTACAGTCCGA-3'

All adapters used in ligation are HPLC purified. PCR reactions can be performed with standard desalted primers.

3. Methods

3.1 UV-crosslinking, lysis and RNase treatment

1. Two days prior to UV-crosslinking, transfect HEK293T cells with proviral plasmid DNAs. For each CLIP-seq experiment, use six 10-cm dishes, each transfected with 10 µg of proviral plasmid DNA (*see Note 2*).
2. One day post-transfection and 14 hours prior to UV crosslinking, change media on plates with 100 µM 4-SU containing media (*see Note 3*).
3. On the day of UV-crosslinking, collect cell culture supernatants containing virions and set aside for processing as detailed in 8–12.
4. Wash cells with 10 mL of ice-cold PBS, aspirate and irradiate the dish containing cells uncovered at an energy setting of 0.15 J/cm² in a Stratalinker UV-crosslinking chamber equipped with UV365 nm bulbs (*see Note 4*).
5. Add 3 mL of PBS to each plate and collect cells using a rubber policeman. Pellet cells by centrifugation at 500 × *g* for 5 min and discard the PBS. Cells can be flash-frozen and stored at this stage.
6. Lyse cells in 2.5 mL of NP40-lysis buffer and keep on ice for 10 min (*see Note 5*).
7. Transfer lysates to microcentrifuge tubes. Clear lysates by centrifugation at 14000 rpm, 4 °C for 10 min and collect the supernatants.
8. In parallel, process cell culture supernatants containing virions. Pellet cellular debris by centrifugation at 500 × *g* for 5 min and filter the supernatant through a 0.2 µM filter.
9. Add 13 mL of 20% sucrose solution in ultracentrifuge tubes and layer 25 mL of cleared cell culture supernatant on top. Pellet the virions by ultracentrifugation at 27000 rpm, 4 °C, for 90 min.
10. Aspirate the supernatant and resuspend the virions in a total of volume of 500 µL PBS.
11. In a 6-well cell-culture dish UV-crosslink the virions twice as above at an energy setting of 0.15 J/cm². Mix between two irradiations.
12. Collect the virions and add 125 µL of 5x NP40 lysis buffer (*see Note 6*).

²As an alternate to transfection, infected HEK293T cells or other cell types (i.e. suspension cells) can be used in CLIP assays.

³6-thioguanosine (6-SG, Sigma-Aldrich) at a final concentration of 100 µM can be used as an alternative ribonucleoside analog, which yields G-to-A mutations in the sequenced reads. However, the efficiency of UV crosslinking with 6-SG and the resulting mutation rates are significantly lower than 4SU-mediated crosslinking.

⁴If CLIP is done on suspension cells, resuspend cells in 10 mL PBS and spread on a 15-cm cell culture dish. Perform UV-crosslinking twice, mixing cells in between.

⁵Following UV-crosslinking, subcellular fractionation can be performed prior to immunoprecipitation. To this end, we found out that a commercially available fractionation kit (Minute plasma membrane isolation kit (Invent Biotechnologies)) works quite well. As has been observed before (22) immunoprecipitation of Gag from the plasma membrane fraction requires harsher detergent conditions (i.e. RIPA buffer). On the other hand, fractionation of cells by membrane flotation on sucrose cushions followed by immunoprecipitation does not yield sufficient Gag-RNA complexes, likely due to the presence of high concentrations of sucrose and increased immunoprecipitation volumes.

13. Add RNase A and DNase I to lysates at a final concentration of 20 U/mL and 60 U/mL, respectively. Incubate samples at 37 °C for 5 min and transfer to ice (*see* Note 7).

3.2 Immunoprecipitation, alkaline phosphatase treatment, end-labeling:

1. For each cell and virion lysate, prepare 60 μ L and 40 μ L of Protein G magnetic beads, respectively. Wash beads twice with 1 mL and resuspend in two bead volumes of citrate-phosphate buffer.
2. Add 5–10 μ g of antibody per 100 μ L of beads. Incubate on a rotating wheel at room temperature for 45 min (*see* Note 8).
3. Wash beads twice with 1 mL and resuspend in one bead volume of citrate-phosphate buffer.
4. Add the proper amount of antibody-conjugated magnetic beads to cell and virion lysates and incubate on a rotating wheel at 4 °C for 1 hr.
5. Collect beads on a magnetic stand in low-retention microcentrifuge tubes. Wash beads twice each with 1 mL of IP wash buffer, LiCl buffer, NaCl buffer, KCl buffer and dephosphorylation buffer. Briefly spin the beads and remove the remaining buffer.
6. Resuspend beads in 1 bead volume of dephosphorylation buffer containing calf intestinal alkaline phosphatase at a final concentration of 0.5 U/ μ L. Incubate for 10 min at 37°C in a thermal mixer programmed to mix at 1400 rpm for 20 seconds every 2 min.
7. Wash beads twice with 1 mL of phosphatase wash buffer. Incubate on a rotating wheel for 5 min between washes.
8. Wash beads twice with 1 mL of PNK buffer. Briefly spin the beads and remove the remaining wash buffer.
9. Resuspend beads in one bead volume of 1 $\times$ PNK buffer containing 0.5 μ Ci/ μ L γ -³²P-ATP and 1 U/ μ L T4 PNK. Incubate at 37°C for 40 min in a thermal mixer programmed to mix at 1400 rpm for 20 seconds every 2 min.
10. Add non-radioactive ATP at a final concentration of 100 μ M and incubate as above at 37°C for an additional 10 min.

⁶.Although immunoprecipitation of Gag from immature virions was very efficient in 1 $\times$ NP40-lysis buffer, immunoprecipitation of NC from mature particles required harsher detergent conditions (i.e. 1 $\times$ RIPA buffer).

⁷.RNase T1 can be used as an alternative to RNase A. The concentration of RNase for each protein and stock of RNase shall be determined separately. We suggest trying several dilutions of RNases in pilot experiments and move forward with the RNase concentration that yields protein-RNA complexes migrating ~5–10 kDa above the expected molecular weight of the protein of interest in SDS-PAGE. The goal is to obtain RNA molecules that are short enough to be sequenced on an Illumina platform but long enough to be unambiguously mapped to the viral and human genomes.

⁸.To facilitate the purification of Gag-RNA adducts, we typically utilize proviral clones carrying 3 consecutive copies of a HA-tag within the stalk region of matrix domain and perform immunoprecipitations using a mouse monoclonal anti-HA antibody (HA.11, Covance). This approach yields abundant and fairly pure Gag-RNA adducts (17).

11. Wash beads once each with 1 mL of PNK Buffer, LiCl buffer, KCl buffer and NaCl buffer as described above. Briefly spin the beads and remove the remaining wash buffer.
12. Resuspend beads in 50 μ L of 1 $\times$ SDS-PAGE loading buffer and elute protein-RNA complexes by incubation at 72 °C for 10 min in a thermal mixer set to constantly mix at 1400 rpm. Collect the eluates.

3.3 Separation of protein-RNA adducts and purification of RNA

1. Run 45 μ L of the eluate on a 4–12% Bis-Tris polyacrylamide gel and transfer to a nitrocellulose membrane. Use the remaining eluate to test for the efficiency of immunoprecipitation by western blotting.
2. Place the membrane in a plastic wrap and expose to autoradiography film until the protein-RNA adducts can be visualized (*see* Note 9).
3. Using a clean scalpel or razor blade, cut a region of the membrane directly above the expected molecular weight of the protein of interest. Cut the membrane further into smaller pieces and place in low-retention microcentrifuge tubes.
4. Add 400 μ L of 1 $\times$ Proteinase K buffer containing 2 mg/mL Proteinase K to each sample. Incubate for 30 min in a thermal mixer set to 55 °C with constant agitation at 1100 rpm.
5. Supplement with an additional 400 μ g Proteinase K and incubate for 15 min as above (*see* Note 10).
6. Lower the temperature to 37 °C and add 1 volume of phenol:chloroform:isoamyl alcohol. Vortex and incubate at 37 °C in a thermal mixer as above for 10 min.
7. Centrifuge samples at 14000 rpm, 3 min, RT.
8. Collect the supernatants and add 100 μ L 3 M sodium-acetate, 1 μ L glycoblue and 1 mL ethanol:isopropanol (1:1). Mix well and place samples at –20 °C overnight.

3.4 Adapter ligations

1. Pellet RNA by centrifugation at 14000 rpm, 4 °C for 30 min.
2. Wash with 500 μ L 80% ethanol. Centrifuge as above for 5 min.
3. Air-dry RNA pellet and resuspend in 8 μ L water.
4. In parallel, bring 1 pmole of the end-labeled positive control RNA up to 8 μ L with water.

⁹If protein-RNA complexes cannot be visualized following 4–5 hours of exposure, we think it is critical to optimize the preceding steps to obtain more abundant crosslinked protein-RNA complexes, which is critical for obtaining meaningful results from a given CLIP-seq experiment.

¹⁰The goal at this step is to maximize the amount of RNA recovered from the nitrocellulose membrane. If sufficient RNA is recovered after the initial round of Proteinase K digestion, this step can be omitted.

5. Add 100 pmoles of 3' adapter (1 μ L of 100 μ M stock), 2 μ L DMSO and 5 μ L PEG8000 (50%). Mix well and denature RNA by incubation at 72 °C for 2 min. Place samples immediately on ice.
6. To each sample add 2 μ L 10 $\times$ T4 RNA ligase 2 buffer without ATP, 20U of RNase inhibitor, 1 μ L 2 mg/ml BSA and 1 μ L of T4 RNA ligase 2, truncated K227Q (200U/ μ L). Mix well.
7. Incubate samples overnight at 16 °C.
8. Add 20 μ L of 2X TBE-Urea loading buffer to each sample and incubate at 72°C for 2 min. Transfer samples to ice.
9. Load samples on a 15% TBE-Urea gel, while leaving at least one empty well between each sample to avoid cross-contamination. Run at 180V, 70 min.
10. Place gel in plastic wrap and expose to autoradiography film.
11. Cut a gel piece corresponding to the ligated RNA products, including the ligated positive control RNA. Crush the gel into smaller pieces (*see* Note 11).
12. Add 3 volumes of 0.4M NaCl supplemented with 200 units/mL of RNase inhibitor.
13. Incubate samples overnight in a thermal mixer set to constant shaking at 1400 rpm, 4°C.
14. Pass gel slurry through a spin column with cellulose acetate filter. Add 1 μ L glycoblue and 2.5V of ethanol:isopropanol (1:1). Place samples on ice for 20 min.
15. Precipitate RNA as above and resuspend in 10 μ L of ultrapure water. Add 20 pmoles (1 μ L of 20 μ M stock) of barcoded 5' adapter and 2 μ L DMSO. Incubate samples at 72 °C for 2 min and immediately transfer to ice.
16. To each sample add ligation mix containing 2 μ L 10 $\times$ T4 RNA ligase 1 buffer, 20U of RNase inhibitor, 1 μ L of 2 mg/mL BSA, 2 μ L ATP (10 mM), 1 μ L T4 RNA Ligase 1 (10U/ μ L).
17. Incubate samples overnight at 16 °C.
18. If barcoded adapters are used, samples can be pooled at this stage. After adding 20 μ L of 2X TBE-Urea loading buffer to each sample, incubate at 72°C for 5 min. Transfer samples on ice and combine them as desired (*see* Note 12).
19. Process samples as described in steps 9 through 15. Resuspend pelleted RNA in 10 μ L of ultrapure water.

¹¹·To crush gel pieces, poke 4 holes on the bottom of a 0.5 mL microcentrifuge tube and place it in a 1.5 mL low-retention microcentrifuge tube. Place the gel piece in the 0.5 mL tube and centrifuge at 14000 rpm, RT for 3 min. Alternatively, a Teflon pestle can be used.

¹²·If RNA abundance varies significantly between samples, it is preferable to pool them at equimolar concentrations to obtain relatively similar number of sequencing reads from each sample.

3.5 Reverse transcription and PCR amplification of CLIP library

1. Use the SuperScript III First Strand Synthesis System as detailed by the manufacturer. To 8 μ L of RNA, add 1 μ L of reverse transcription primer (10 μ M) and 1 μ L of dNTP (10mM). Incubate at 65 °C for 5 min. Transfer samples on ice.
2. Add 4 μ L $MgCl_2$ (25 mM), 2 μ L 10 $\times$ Reverse Transcription Buffer, 2 μ L DTT (0.1M), 1 μ L RNaseOUT (40U/ μ L) and 1 μ L SSIII RTase (200U/ μ L).
3. Reverse transcribe according to manufacturer's instructions.
4. Set up a 100 μ L PCR reaction containing 10 μ L of cDNA, 50 pmoles of forward and reverse primers, 20 μ L 5 $\times$ High Fidelity Buffer, 2 μ L dNTPs (10 mM), 2 μ L high-fidelity polymerase (2U/ μ L). Run PCR for a total of 15 cycles programmed at 98 °C 15 seconds, 55 °C 30 seconds, 72 °C 15 seconds. Take 20 μ L aliquots at the end of PCR cycle 6, 9, 12 and 15.
5. Run samples on a 6% TBE-Urea gel. Stain with EtBr in 1 $\times$ TBE buffer and excise a region corresponding to the CLIP library (*see* Note 13).
6. Weigh the gel and add 1–2 volumes of diffusion buffer.
7. Incubate at 50°C for 30 min in a thermal mixer set to constant agitation at 1400 rpm.
8. Collect the supernatant and extract CLIP DNA library using a gel extraction kit.
9. As CLIP libraries are derived from short RNA sequences, they can be sequenced on an Illumina platform for 50-cycles.

3.6 Bioinformatics analyses

Despite the astounding progress in the development and accessibility of next-generation sequencing-based experimental approaches, data analysis still remains the major limiting step in many laboratories and institutions. Several tools have been developed and are publicly available for data analyses. We perform our data analysis by running these programs from the command line on our local Linux server, for which a basic knowledge of UNIX operating system and other programming languages is necessary. Below we outline the basic analysis of the CLIP-seq data, including sample command line instructions:

1. Raw reads obtained from the sequencing facility can be processed prior to mapping to human or viral genomes using the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/).
2. Discard reads that contain ambiguous nucleotides, did not contain the 3' adapter sequence or are shorter than 15 nucleotides: `fastx_clipper -c -a TCGTATGC -l 23 -i INFILE.fastq -o OUTFILE` (*see* Note 14).

¹³-We usually observe two bands, one corresponding to adapter-adapter ligations that migrate at ~75 nt and the CLIP library that migrate at 90–150 nt. From the smallest number of PCR cycle (typically 9–12) that yields a visible library, carefully cut the region that corresponds to the CLIP library from the gel. If necessary, repeat the PCR using the purified CLIP library as template.

¹⁴-Length filter at this step is 23 nucleotides (-l 23) as this includes the length of the 5' adapter (8 nucleotides).

3. Separate reads based on their 5' barcode sequences: `cat OUTFILE | /usr/local/bin/fastx_barcode_splitter.pl --bcfile barcodes.txt --bol --mismatches 0 --prefix OUTFILE_split`
4. Collapse reads to generate a set of unique sequences: `fastx_collapser -i OUTFILE_split -o OUTFILE_collapsed`
5. Trim the 5' adapter sequence: `fastx_trimmer -f 9 -v -i OUTFILE_collapsed -o OUTFILE_trimmed`
6. After the reads are “cleaned-up” they can be aligned to the appropriate viral and human genomes using several short-read aligners (i.e. Bowtie, BWA etc.). For alignment to the human genome, we typically allow up to 2 mismatches and report locations for reads with the minimum number of observed mismatches for each read (Bowtie criteria: `-m 10 -v 2 --best --strata` for mapping to hg19, and `-m 1 -v 2` for mapping to the viral genomes).
7. Following the alignment, SAMtools (18) and BEDtools (19) can be used to further process the data.
8. Clusters, which represent binding sites derived from overlapping mapped reads, can be generated using the PARalyzer algorithm (20).
9. The generated clusters can be annotated by in-house scripts using publicly available databases (i.e. ENSEMBL, UCSC) as reference.
10. Following annotation, motif searches within clusters can be performed by the cERMIT algorithm (21).

Acknowledgements

This work was supported by NIH grants R01AI501111 and P50GM103297. S.B.K. was supported in part by an AmFAR Mathilde Krim Postdoctoral Fellowship.

References

1. Kuzembayeva M, Dilley K, Sardo L et al. (2014) Life of psi: how full-length HIV-1 RNAs become packaged genomes in the viral particles. *Virology* 454-455, 362–370. [PubMed: 24530126]
2. Lu K, Heng X, Summers MF (2011) Structural determinants and mechanism of HIV-1 genome packaging. *J Mol Biol* 410, 609–633. [PubMed: 21762803]
3. Rein A, Datta SA, Jones CP et al. (2011) Diverse interactions of retroviral Gag proteins with RNAs. *Trends Biochem Sci* 36, 373–380. [PubMed: 21550256]
4. Aldovini A, Young RA (1990) Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J Virol* 64, 1920–1926. [PubMed: 2109098]
5. Clavel F, Orenstein JM (1990) A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. *J Virol* 64, 5230–5234. [PubMed: 2204725]
6. Lever A, Gottlinger H, Haseltine W et al. (1989) Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J Virol* 63, 4085–4087. [PubMed: 2760989]
7. Luban J, Goff SP (1994) Mutational analysis of cis-acting packaging signals in human immunodeficiency virus type 1 RNA. *J Virol* 68, 3784–3793. [PubMed: 8189516]

8. Clever JL, Parslow TG (1997) Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J Virol* 71, 3407–3414. [PubMed: 9094610]
9. Laham-Karam N, Bacharach E (2007) Transduction of human immunodeficiency virus type 1 vectors lacking encapsidation and dimerization signals. *J Virol* 81, 10687–10698. [PubMed: 17652403]
10. McBride MS, Schwartz MD, Panganiban AT (1997) Efficient encapsidation of human immunodeficiency virus type 1 vectors and further characterization of cis elements required for encapsidation. *J Virol* 71, 4544–4554. [PubMed: 9151848]
11. Jouvenet N, Simon SM, Bieniasz PD (2009) Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles. *Proc Natl Acad Sci U S A* 106, 19114–19119. [PubMed: 19861549]
12. Kutluay SB, Bieniasz PD (2010) Analysis of the initiating events in HIV-1 particle assembly and genome packaging. *PLoS Pathog* 6, e1001200. [PubMed: 21124996]
13. Rulli SJ Jr., Hibbert CS, Mirro J et al. (2007) Selective and nonselective packaging of cellular RNAs in retrovirus particles. *J Virol* 81, 6623–6631. [PubMed: 17392359]
14. Muriaux D, Mirro J, Harvin D et al. (2001) RNA is a structural element in retrovirus particles. *Proc Natl Acad Sci U S A* 98, 5246–5251. [PubMed: 11320254]
15. Hafner M, Landthaler M, Burger L et al. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129–141. [PubMed: 20371350]
16. Licatalosi DD, Mele A, Fak JJ et al. (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 456, 464–469. [PubMed: 18978773]
17. Kutluay SB, Zang T, Blanco-Melo D et al. (2014) Global changes in the RNA binding specificity of HIV-1 Gag regulate virion genesis. *Cell* 159, 1096–1109. [PubMed: 25416948]
18. Li H, Handsaker B, Wysoker A et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. [PubMed: 19505943]
19. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. [PubMed: 20110278]
20. Corcoran DL, Georgiev S, Mukherjee N et al. (2011) PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol* 12, R79. [PubMed: 21851591]
21. Georgiev S, Boyle AP, Jayasurya K et al. (2010) Evidence-ranked motif identification. *Genome Biol* 11, R19. [PubMed: 20156354]
22. Ono A, Waheed AA, Joshi A et al. (2005) Association of human immunodeficiency virus type 1 gag with membrane does not require highly basic sequences in the nucleocapsid: use of a novel Gag multimerization assay. *J Virol* 79, 14131–14140. [PubMed: 16254348]