

Washington University School of Medicine

Digital Commons@Becker

Open Access Publications

2018

CLIP-related methodologies and their application to retrovirology

Paul D. Bieniasz

The Rockefeller University

Sebla B. Kutluay

Washington University School of Medicine in St. Louis

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs

Please let us know how this document benefits you.

Recommended Citation

Bieniasz, Paul D. and Kutluay, Sebla B., "CLIP-related methodologies and their application to retrovirology." *Retrovirology*. 15, 35. (2018).

https://digitalcommons.wustl.edu/open_access_pubs/6928

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.

REVIEW

Open Access



CLIP-related methodologies and their application to retrovirology

Paul D. Bieniasz¹ and Sebla B. Kutluay^{2*} 

Abstract

Virtually every step of HIV-1 replication and numerous cellular antiviral defense mechanisms are regulated by the binding of a viral or cellular RNA-binding protein (RBP) to distinct sequence or structural elements on HIV-1 RNAs. Until recently, these protein–RNA interactions were studied largely by in vitro binding assays complemented with genetics approaches. However, these methods are highly limited in the identification of the relevant targets of RBPs in physiologically relevant settings. Development of crosslinking-immunoprecipitation sequencing (CLIP) methodology has revolutionized the analysis of protein–nucleic acid complexes. CLIP combines immunoprecipitation of covalently crosslinked protein–RNA complexes with high-throughput sequencing, providing a global account of RNA sequences bound by a RBP of interest in cells (or virions) at near-nucleotide resolution. Numerous variants of the CLIP protocol have recently been developed, some with major improvements over the original. Herein, we briefly review these methodologies and give examples of how CLIP has been successfully applied to retrovirology research.

Background

Following the integration of proviral DNA into the host cell chromosome, genesis of new HIV-1 particles is initiated by the host RNA Polymerase II-mediated synthesis of a single poly-cistronic viral RNA species [1]. This transcript undergoes varying levels of alternative splicing generating over 40 different RNA species, an event orchestrated by the host cellular splicing machinery and cis-acting elements on viral RNAs [1, 2]. Like cellular mRNAs, all viral RNAs contain 5' 7-methylguanosine (m7G) caps and 3' polyA tails [1, 3]. While fully spliced viral RNAs can exit the nucleus via canonical nuclear export pathways, the partially spliced and unspliced viral RNAs depend on the viral Rev and cellular Crm1 proteins for nuclear export [4]. All viral mRNAs are subsequently translated in the cytosol, but the unspliced full-length viral RNAs also serve as the viral genome and are packaged into virions by the viral major structural protein Gag. Following their release from the plasma membrane, particles undergo a maturation step triggered by the viral

protease enzyme. During this process, Gag and Gag-Pol proteins are cleaved into their constituent domains, the CA domain of Gag forms a conical lattice and the viral RNA genome condenses with the cleaved NC domain of Gag and viral enzymes inside this conical core [5, 6]. Thus, virtually every step in HIV-1 replication depends on a complex and changing set of interactions between viral RNAs and the multitude of trans-acting viral and cellular RNA-binding proteins. Historically, the interactions between these proteins and their RNA targets have largely been mapped by genetic studies, complemented by limited in vitro approaches. Comprehensive analysis of these interactions in physiologically relevant settings was effectively impossible prior to the recent development of cutting-edge next-generation sequencing-based methodologies. These methods, collectively referred to as CLIP (crosslinking-immunoprecipitation coupled with next-generation sequencing), allow the global identification of RNA targets of RNA-binding proteins (RBPs) in physiological settings in unprecedented detail. In this review, we provide a detailed outline of the existing CLIP methodologies, discuss their advantages and shortcomings (based partly on our own experience) and give examples of how CLIP has been successfully applied to retrovirology research.

*Correspondence: kutluay@wustl.edu

² Department of Molecular Microbiology, Washington University School of Medicine, Saint Louis, MO 63110, USA

Full list of author information is available at the end of the article



(See figure on next page.)

Fig. 1 Outline of CLIP

Principles of CLIP and variant methodologies

In simple terms, CLIP is a powerful methodology with which one can identify the RNA targets of RNA-binding proteins in physiological settings, ranging from live cells to virus particles and even animal tissues. The inception of the original CLIP protocol [7, 8] and its subsequent coupling to next-generation sequencing [9] has revolutionized the study of protein–RNA interactions. Since then, several other versions of CLIP have been developed. The salient steps of the existing CLIP methodologies are (Fig. 1): (1) protein–RNA complexes are covalently crosslinked in live cells/tissues/virions; (2) Cells/tissues/virions are lysed and treated with limited amounts of RNases leaving small fragments of RNA molecules (~20 to 50 nucleotides) protected by the protein of interest; (3) Protein–RNA complexes are immunoprecipitated, and non-specific RNAs and proteins are removed by stringent washes. Because the protein–RNA complexes are covalently crosslinked, these stringent conditions, in principle, do not affect purification of target protein–RNA adducts. (4) The purified protein–RNA complexes are radioactively labeled and separated by SDS-PAGE. (5) Bound RNA is isolated either directly from SDS-PAGE gels or from nitrocellulose membranes following transfer by Proteinase K treatment. (6) Eluted RNA is ligated to adapters, reverse transcribed, the resulting cDNA is PCR amplified and subjected to sequencing. (7) Sequencing reads are processed and mapped to reference genomes. Depending on the method used, the resulting library contains nucleotide substitutions or deletions at the site of crosslinking, which allows mapping of the site of protein–RNA interactions at near-nucleotide resolution. Subsequent analyses include determination of the significantly enriched binding sites, identification of the binding motifs within them as well as other custom analyses. In the remainder of this section we will review the currently existing CLIP methods and give an overview of the widely used CLIP data analysis tools and pipelines.

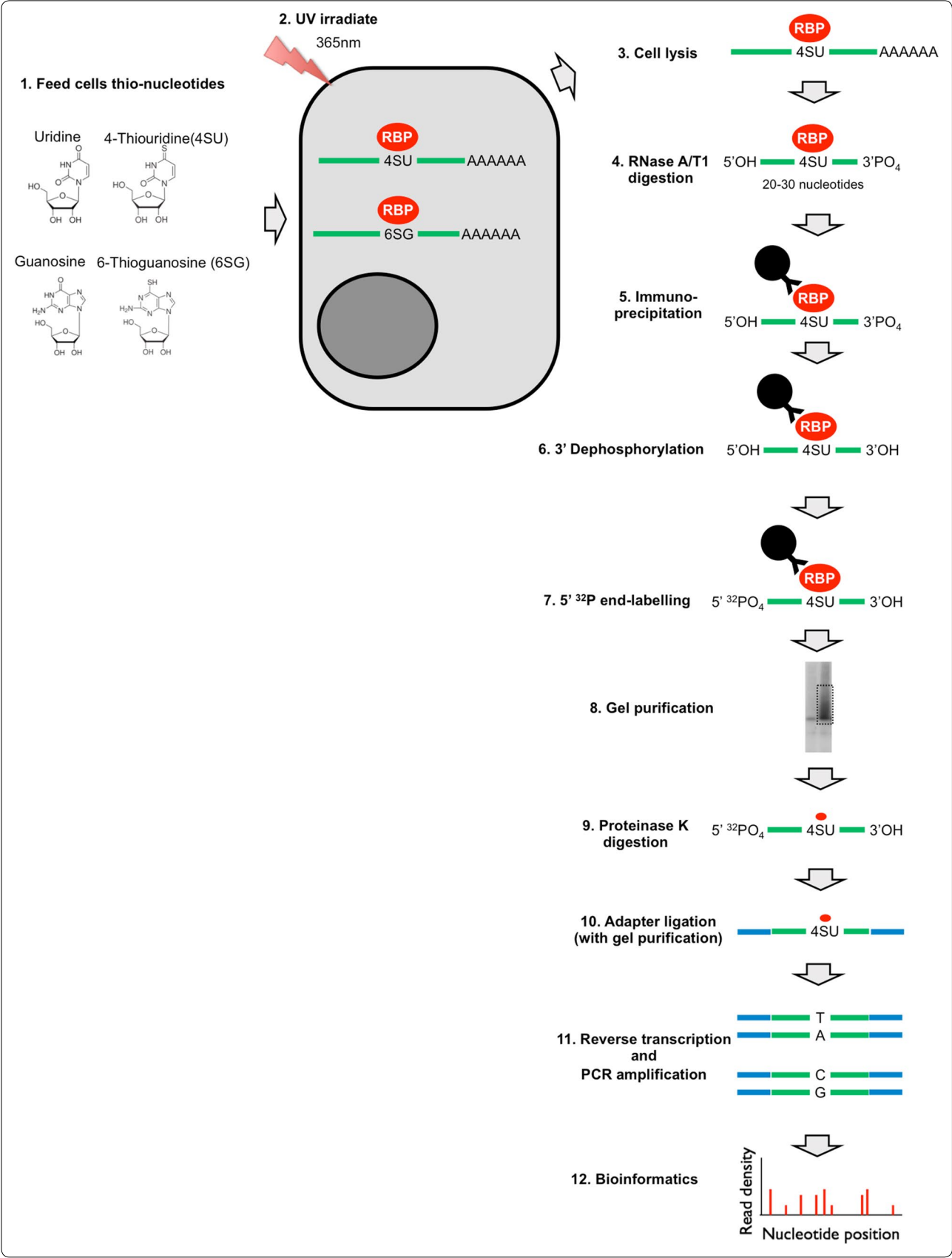
HITS-CLIP

Historically, protein–RNA interactions were studied largely using in vitro binding assays with pure proteins and RNAs. Alternatively, GST-pulldown and immunoprecipitation-based assays were conducted on cell lysates followed by downstream quantitative analysis of RNA by Q-RT-PCR or microarrays. A major drawback of these cell lysate-based approaches was their limited ability to identify direct interactions between a RBP and its target RNA molecules. Their limited power was due at least in

part to the presence of contaminating protein and RNA molecules in the isolated RBP–RNA complexes. Development of the original CLIP protocol [7, 8], in which the protein–RNA complexes were UV-crosslinked in vivo and immunoprecipitated under stringent conditions to remove the contaminating proteins and RNA molecules marked the first advancement over these traditional methods. While the initial CLIP methodology relied on cloning and subsequent sequencing of the RNA targets, the coupling of CLIP to high-throughput sequencing, HITS-CLIP, allowed global transcriptome-wide analysis of RBP–RNA crosslinks [9]. HITS-CLIP relies on UV crosslinking of protein–RNA complexes at UV254 nm. As such, HITS-CLIP can be applied to animal tissues due to its high level of penetration. Following crosslinking and immunoprecipitation of protein–RNA complexes, ligation to the radioactively labeled 5' adapter is performed while the protein–RNA adducts are attached to beads. This allows the removal of unligated 5' adapter by further rounds of bead washing, which substantially reduces the appearance of adapter–adapter ligation products following downstream processing. The isolated protein–RNA adducts are separated by SDS-PAGE and transferred to nitrocellulose membranes. As naked RNA molecules are not retained on the nitrocellulose membranes, protein–RNA complexes are purified further during this step. Transfer to nitrocellulose membranes has been utilized in other CLIP approaches and in our experience confers a major advantage over the originally described PAR-CLIP approach described below. Protein-crosslinked RNA is further purified from nitrocellulose membranes by proteinase K treatment, ligated to the 3' adapters and PCR-amplified prior to sequencing. Detailed bioinformatics analyses of HITS-CLIP datasets revealed that reverse transcriptase (RT) introduces deletions at the site of crosslinking [10], albeit at a fairly low frequency, allowing HITS-CLIP to reach to near nucleotide-resolution identification of binding sites.

PAR-CLIP

A major advantage of PAR-CLIP [11] over HITS-CLIP is the use of ribonucleoside analogs, including 4-thiouridine (4SU) and 6-thioguanosine (6SG), that significantly enhance the efficiency of protein–RNA crosslinking. In PAR-CLIP experiments, cells are typically grown in the presence of ribonucleoside analogs for up to 16 h and UV-crosslinked at a longer wavelength (365 nm). As such, in contrast to HITS-CLIP, application of PAR-CLIP is largely limited to cell culture systems (an exception



being *C. elegans* which can be grown in 4SU containing media and efficiently UV-crosslinked due to its transparency [12]). Although the original PAR-CLIP description utilized an inducible tagged RNA-binding protein [11], we and many other groups have successfully adapted PAR-CLIP to study endogenous proteins, including HIV-1 NC and IN [13, 14], Argonaute [15–18], as well as other proteins involved in RNA biogenesis and metabolism [19–24]. A potential disadvantage of the PAR-CLIP protocol is the cellular toxicity that may be induced by 4SU treatment depending on the cell type, the dose and incubation time [25]. Thus, optimal conditions that allow efficient protein–RNA crosslinking without major toxicity should be determined on a case-by-case basis. Nevertheless, PAR-CLIP allows accurate nucleotide resolution mapping of target RNA sites due to mutations introduced by RT (T-to-C for 4SU and G-to-A for 6SG) precisely at the site of crosslinking during cDNA synthesis. While allowing nucleotide resolution mapping, use of ribonucleoside analogs may inadvertently enrich RNA elements with distinct nucleotide composition or alter RNA structure [26], which may subsequently affect protein binding. Careful validation of PAR-CLIP experiments with different ribonucleoside analogs and RNases should, in principle, address these potential problems.

iCLIP

Identification of the precise crosslinking site in the HITS-CLIP and PAR-CLIP approaches relies respectively on deletions and substitutions introduced by RT during cDNA synthesis. However, read-through at crosslinking sites appears to be a relatively rare event as compared to truncations that occur as a result of RT stalling at these sites [27, 28]. Thus, a major shortcoming of HITS-CLIP and PAR-CLIP approaches is the loss of a large fraction (estimated to be >80%) of the starting material due to the inability to recover truncated reverse transcription products. iCLIP [29] has been designed to address this problem by ligation of a 3' adaptor while protein–RNA complexes are still on beads followed by introduction of a two-part cleavable adaptor into cDNA during reverse transcription. The resulting cDNA is circularized and subsequently linearized with a restriction enzyme, which allows the recovery of a larger fraction of truncated cDNAs. In addition, as circularization is done at high temperatures, structured cDNA molecules are recovered at a much higher efficiency. As a result of this enrichment, iCLIP can yield higher complexity libraries and has been proposed to perform better than previous approaches in identification of the precise site of crosslinking [28–30]. Application of iCLIP on a large scale by the ENCODE consortium indicated that the success rate in generating libraries was low for many RBPs, which was ascribed to

the low efficiency of the circularization step [31]. However, several studies that utilized iCLIP have generated libraries with sufficiently high complexity and sensitivity, and these parameters were not carefully assessed by the ENCODE consortium. The remainder of the iCLIP protocol is similar to HITS-CLIP and PAR-CLIP approaches. BrdU-CLIP [32] and FAST-iCLIP [33] are iterations of the iCLIP protocol, which provide alternative cDNA and RNA purification methods, respectively. For example, by exchanging the 3' ddC blocker from the standard iCLIP 3' adaptor with a 3' biotin moiety and subsequent purification of ligation products on streptavidin beads, FAST-iCLIP is reported to reduce the time required to perform iCLIP by 50%.

eCLIP

The eCLIP protocol [31] proposes to address some of the shortcomings of previous CLIP approaches by including two separate adapter ligation steps (i.e. in the HITS-CLIP and PAR-CLIP protocols). In eCLIP, the immunoprecipitated RNA is first ligated to an indexed 3' RNA adapter while complexes are still on the immunoprecipitation beads, and to a 3' single-stranded (ss) DNA adapter after reverse transcription. As reverse transcription frequently terminates at the RBP–RNA crosslinking site, the ligation of the 3' ssDNA adapter to the terminated cDNA fragments allows higher recovery rates of the starting material and helps in identification of the binding sites as in iCLIP. In addition, as the first 3' RNA adapter already contains the indexes, samples can be combined at an earlier stage than in other protocols saving processing time. While adapter ligations conducted on beads has been inefficient in our hands (see below), the authors suggest that increased T4 RNA ligase concentration and the addition of high concentrations of polyethylene glycol (PEG8000) and DMSO in ligation reactions enable ligation efficiencies of up to >90% [31]. In addition, RNA radiolabeling and autoradiographic visualization steps can be omitted allowing even faster library preparation times. However, these steps in our experience are highly important to purify the target protein–RNA complexes away from other proteins and RNA molecules that have non-specifically immunoprecipitated. Thus, the specificity of eCLIP libraries should be carefully evaluated, as also reviewed by a recent study [34]. Finally, inclusion of a size-matched input control (SMInput) in eCLIP enables efficient background normalization and controls for any inherent biases in library generation. The remainder of the eCLIP protocol shares many of the same steps as other CLIP approaches, in particular iCLIP.

irCLIP

Similar to eCLIP, irCLIP has been developed to overcome some of the shortcomings of previous CLIP methodologies by simplifying the library generation steps, increasing the yield and complexity of the CLIP library, and allowing faster processing times. One of the major differences of this approach is the utilization of a 3' adapter conjugated to an infrared fluorescent dye [35], which provides a more sensitive and faster way of tracking the target RNA molecules compared to radioactive labeling. Similar to FAST-iCLIP, the adapter ligated RNA library is purified by streptavidin beads. CLIP has an inherent bias against identification of protein binding events on structured RNA elements due to stalling of RT at these sites. Although not proven, irCLIP may mitigate this problem by utilizing thermostable enzymes for circularization and reverse transcription steps to take place at 60 °C, which helps to resolve potential RNA secondary structures [35]. Other aspects of the irCLIP protocol, such as on-bead nuclease digestions and Proteinase K digestion in SDS have previously been utilized within the context of PAR-CLIP experiments [11, 14]. As in iCLIP and eCLIP, the irCLIP procedure achieves single-nucleotide resolution by recovery of truncated cDNAs after the reverse transcription stage.

Customizing CLIP

The major shortcomings of all of the above CLIP approaches include technically challenging and labor-intensive protocols, and loss of the starting material at several inefficient steps in the procedure. This problem is further exacerbated if the initial protein–RNA complexes are not abundant due to low levels of expression in cells (virions), low crosslinking or immunoprecipitation efficiencies. These problems can often lead to a final library with insufficient complexity and enrichment of environmental contaminating sequences. When we adapted the CLIP protocol to study HIV-1 Gag–RNA interactions [36], we took advantage of both HITS-CLIP and PAR-CLIP protocols as detailed in [14]. In our experience, 4-SU-mediated crosslinking yielded more abundant Gag-, MA- and NC-RNA complexes, that was critical for generating libraries with sufficient sequence diversity for successful sequencing. While the original PAR-CLIP protocol relied on electroelution of protein–RNA complexes from SDS-PAGE gels, we opted for transfer of protein–RNA complexes to nitrocellulose membranes following SDS-PAGE (as in HITS-CLIP). As naked RNA oligonucleotides are not immobilized on nitrocellulose membranes, this step provides an added level of protein–RNA complex purification. While the HITS-CLIP and many other protocols call for ligation of adapters while the protein–RNA complexes are on beads, the PAR-CLIP library

generation protocol in solution was significantly more efficient in our hands with 3' and 5' adapter ligations routinely working at >90 and 50% efficiency. Although seemingly more cumbersome, sequential ligation of adapters provides more control over monitoring the ligation efficiency and substantially decreases contaminating adapter–adapter ligation products. Additionally, we have utilized barcoded and degenerate sequence containing adapters, which enabled us to combine multiple samples (typically up to eight) and distinguish between independent ligation versus PCR overamplification events, respectively. Finally, due to some of the potential inherent biases of the PAR-CLIP approach discussed above, we typically validate our findings using different ribonucleosides (4SU vs. 6SG) and RNases (RNase A vs. RNase T1).

CLIP data analyses

CLIP data analyses can be summarized in four major steps: (1) pre-processing of sequencing reads. (2) mapping of reads to reference genomes, (3) subjecting mapped reads to cluster finding algorithms to define binding sites, (4) analysis of binding sites for enrichment of certain features including where within a gene body the binding site is located, presence of distinct motifs or nucleotide composition. Recently a few pipelines that can perform the majority of these steps have been developed and include the PARCLIPsuite [37], CLIPZ [38], CIMS [39] and CLIP-seq tools [40]. Below, we will go through some of the publicly available and most frequently used standalone tools that can be utilized for analyses of CLIP data sets. For a more detailed review of these tools and algorithms we refer the readers to detailed recent reviews [41–44]. Implementing many of these analysis pipelines requires some level of coding knowledge and familiarity with shell scripting.

1. Pre-processing of sequencing reads: The resulting CLIP libraries in all of the above protocols will contain some form of 3' and 5' adaptors. In the majority of cases, these adaptors contain barcodes and degenerate sequences (N_{3-10}), which allow multiplexing and differentiating between independent ligation versus PCR overamplification events, respectively. In these circumstances, a typical pipeline will involve removing low quality reads, collapsing of raw reads into unique reads, demultiplexing samples, discarding short reads (typically less than 15 nucleotides) and trimming the adaptors prior to mapping. One of the most commonly used tools is the FASTX_toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), which provides a number of functions to accomplish all of these steps. Other alternatives, with more limited

functions include Cutadapt [45], Trimmomatic [46], PRINSEQ [47] as well as custom scripts.

2. Mapping to reference genomes: The reads that pass the above filtering steps are mapped onto reference genomes or transcriptomes. The most commonly used mapping algorithms used for this task include Bowtie [48], Bowtie2 [49], STAR [50], Novoalign (<http://www.novocraft.com/products/novoalign/>), RMAP [51], TopHat [52], GSNap [53], SOAP [54] and BWA [55], some with unique advantages over others depending on whether mapping is done on a genome versus transcriptome. The choice of algorithm and the parameters for mapping will need to be finely tuned depending on which CLIP methodology is employed and the properties of the RBP of interest. For example, PAR-CLIP reads are expected to contain a number of T-to-C substitutions, and thus mismatches (typically ≤ 2 for reads between 15 and 40 nucleotides) should be allowed during mapping. While all algorithms allow mapping with mismatches, not all can handle deletions, which arise as a result of UV₂₅₄ nm crosslinking in HITS-CLIP and related methods. For example while the original Bowtie algorithm did not allow gaps during alignment, Bowtie2 was developed to enable alignments with indels. Similarly, if mapping is done on transcriptomes, alignment algorithms such as STAR, which allow higher accuracy and speed for mapping spliced transcripts should be preferred. However, mapping to the transcriptome will clearly lead to the exclusion of reads derived from introns, which may constitute the primary binding sites for various splicing regulatory proteins. Thus, a general strategy whereby CLIP reads are mapped first to the transcriptome and the remaining reads are mapped to the genome may work the best for proteins for which there is no information on the types of targeted RNA molecules.
3. Peak calling: The next essential step in CLIP analysis is identification of the true binding sites by what is often referred to as *peak calling*. In simple terms, peak calling is the process by which clusters of reads that map to distinct locations are separated from background reads that may stem from unspecific binding events or contaminants during the CLIP procedure. Peaks are typically defined based on a number of variables such as read depth relative to surrounding regions, presence of expected and absence of unwanted mutations (as in the case of PAR-CLIP-based approaches) and peak shape. While peak calling can be based solely on CLIP data, additional controls such as data derived from replicates and negative controls (i.e. immunoprecipitations done with isotype controls and/or conducted

in lysates lacking the RBP of interest) can further increase specificity of peak calling. Comparison of the CLIP peaks with transcript abundance derived from matching RNA-seq experiments allows the discrimination of whether a binding event is merely a result of transcript abundance or a more specific interaction between the RBP and its target RNA. Several peak calling programs have been developed and include Piranha [56], CLIPper [57], PIPE-CLIP [58], Pyicos [59] that work with all CLIP variants, and PARalyzer [60] and wavClusteR that are specifically developed for PAR-CLIP analysis. For more details on the statistical models underlying these programs, we refer the readers to detailed reviews on this topic [41, 42].

4. Post-processing analyses: Following the identification of peaks, further analyses are typically conducted to identify the specific rules that may determine protein binding. For example, many studies generally assess what classes of RNAs and where within those transcripts binding sites are located and whether there are distinct motifs within the binding sites. While the former analyses are done usually by custom scripts, programs such as MEME [61], HOMER [62] and cERMIT [63] are commonly used for motif discovery. Finally, binding sites derived from CLIP experiments can further be analyzed by programs that are commonly used in gene expression profiling experiments for gene ontology and pathway analyses.

Application of CLIP techniques in retrovirology

Novel insights into selective HIV-1 genome packaging

All major steps of HIV-1 particle assembly are orchestrated by the major structural protein, Gag [6]. Gag undergoes major changes in its subcellular localization, structure and oligomeric state during this process. Immediately following its synthesis, Gag exists as a diffuse pool of monomers and low-order multimers in the cytosol, where it initially binds to the viral RNA genome [64, 65]. Concurrent with binding to the plasma membrane Gag undergoes a major structural change and oligomerizes around the viral genome [65]. Following the release of immature particles from the host cell's plasma membrane, particles undergo maturation—Gag is subjected to several proteolytic cleavages, which liberates NC and other constituent domains. NC remains bound to the viral genome and condenses with it inside the remodeled conical capsid lattice. Thus, a crucial property of Gag is its ability to select two copies of the viral genome for packaging in the cytosol and remain bound to them through various subcellular settings and configurations.

The mechanism by which HIV-1 selectively packages a dimeric unspliced viral genome is based largely on prior

observations with simple retroviruses, as well as genetic studies and limited in vitro data. Selective packaging of the HIV-1 genome is governed in part by binding of the nucleocapsid (NC) domain of Gag to a highly structured *cis*-acting packaging element, psi (Ψ), within the 5' leader of the viral genome, composed of sequences in the unique 5' region (U5) and between the tRNA primer binding site (PBS) and the 5' portion of the Gag open reading frame (ORF). However, disruption of Ψ only modestly decreases HIV-1 RNA encapsidation [66–68], and sequences outside Ψ can increase virion RNA levels and viral vector titers [69–73]. In addition, viral RNA is not necessary for particle assembly and cellular RNAs can be packaged in its absence [74, 75]. Thus, although several lines of evidence have long indicated that sequences other than Ψ can contribute to genome packaging, determining the identities and features of these elements remained a challenge, due largely to lack of proper assays to study this process in cells.

Application of the CLIP methodology to the study of Gag-RNA interactions during different stages of particle assembly in cells revealed previously unanticipated rules of selective genome packaging [14]. First, nucleotide-resolution mapping of Gag binding to the HIV-1 genome in the cytosol revealed selective binding to sequences that coincide nearly precisely with a minimal element that can drive genome packaging. This minimal psi (Ψ) element adopts alternative structures, one of which favors genome packaging [76–78]. Second, in addition to Ψ , cytoplasmic Gag was bound to additional discrete elements on the viral RNA, including Rev Responsive Element (RRE), another highly structured region that mediates the export of HIV-1 RNAs from the nucleus. Although Gag-RRE interactions appeared to be dispensable for genome packaging, a more recent study has implied a role for it in preventing Gag from moving away from the viral RNA genome in the cytosol [79]. Third, mapping of Gag binding sites within the cellular mRNAs revealed a striking contrast between the binding preference of cytosolic versus membrane-bound Gag; while cytosolic Gag preferentially bound to GU-rich motifs, A-rich mRNA sequences were found to be enriched in plasma membrane-bound mRNA molecules. Remarkably, the nucleotide composition of the cellular mRNA targets of Gag at the plasma membrane mirrored the unusual A-rich nucleotide composition of the HIV-1 genome [14]. Finally, upon proteolytic cleavage of Gag in mature virions, the NC binding preference reverted back to GU-rich mRNA sequences and discrete viral RNA elements including Ψ . Together, these findings suggest that upon binding of monomeric Gag to the viral genome through Ψ , multimerization-dependent changes in the RNA binding specificity of Gag may drive the selective packaging of the A-rich viral genome. In line with this

model, a recent study has shown that longer segments of the Gag ORF, but not Ψ alone, can gradually increase the packaging of heterologous RNAs into virions [80]. Thus as part of the selective RNA packaging process, the role of Gag- Ψ interaction may be to nucleate further assembly of Gag oligomers on the viral genome [81].

HIV-1 MA-tRNA interactions

In addition to the NC domain, the matrix (MA) domain of Gag had long been suspected to bind RNA, based largely on in vitro assays [82–87]. The N-terminal basic amino acids of MA that are thought to bind RNA also mediate binding to cellular membranes [83, 88–92]. However, MA-RNA interaction has been thought to be fairly non-specific, and whether it actually occurs in cells could not be addressed until the application of the CLIP methodology. By releasing MA from Gag by Factor Xa protease-mediated cleavage as part of the CLIP procedure, following UV-crosslinking of Gag-expressing cells, MA was bound to a specific set of tRNAs in the cytosol [36]. In fact, MA-tRNA interactions constituted the most frequent binding event between cytosolic Gag and RNA. Notably, MA-tRNA interaction was lost upon binding of Gag to the plasma membrane and RNase treatment of cell lysates expressing Gag led to significantly higher levels of membrane associated Gag [36]. Together, these findings suggested that occlusion of MA basic residues by specific tRNAs may target HIV-1 assembly to the plasma membrane and prevent nonproductive assembly on intracellular membranes. Alternatively, tRNA binding by MA may temporally regulate membrane binding and assembly [93]. Recent in vitro liposome binding assays also revealed that a specific set of RNAs, including Ψ , total yeast tRNA and tRNA^{Pro} can inhibit Gag binding to negatively charged lipid membranes lacking PI(4, 5)P₂ [94]. Interestingly, tRNA^{Lys}, which was one of the most frequently bound to tRNAs by MA in cells [36], did not prevent Gag binding to liposomes [94]. As this study only tested the ability of in vitro transcribed tRNAs in regulating Gag membrane binding, it remains to be seen whether tRNAs containing the complete set of post-transcriptional modifications exhibit differences in MA binding in vitro.

In addition to regulation of Gag membrane binding, MA-tRNA interactions could have other functions. An obvious possibility is regulation of viral and/or host translation. As a result of the unusually A-rich nature of the HIV-1 genome [95–97], Ile, Lys, Glu and Val codons are overrepresented in the Gag and Pol ORFs [98]. Notably, tRNA^{Lys}, tRNA^{Glu} and tRNA^{Val} were found to be amongst the most frequently bound by MA, suggesting the possibility of MA enhancing the translation of Gag and Pol by sequestering these specific set of

tRNAs. Alternatively, it is conceivable that by sequestering tRNAs, MA could inhibit translation of host mRNAs whose products may block viral replication. Indeed, one report has suggested that interaction of MA with host translation elongation factors via a tRNA bridge could inhibit *in vitro* translation [84]. It remains to be determined whether MA-tRNA interactions in a relevant infection setting can influence viral or host translation. Finally, it is possible that if not bound by tRNAs, the basic patch on MA may nonspecifically bind to the viral genome and even prevent the proper interaction of NC with the genome, which may inhibit subsequent steps of infection. In a similar scenario, MA binding to small RNAs might be a mechanism to avoid aggregation by a protein that has two distinct RNA binding domains and an intrinsic tendency to multimerize.

Role of IN-RNA interactions in particle maturation

The morphological changes that occur during HIV-1 particle maturation are often thought to be dependent only on proteolytic cleavage of Gag. The cleaved CA domain of Gag forms the conical lattice within which the viral genome condenses, along with the cleaved NC domain of Gag as well as viral enzymes integrase (IN) and reverse transcriptase (RT), cleavage products of the Pol polyprotein. However, more than two decades ago, mutational studies of the HIV-1 IN indicated that it may also play an active role in proper particle maturation [99–110]. In particular, a set of mutations referred to as Class II IN mutations, were shown to lead to the formation of morphologically aberrant “eccentric” particles, in which the viral ribonucleoproteins complexes (vRNPs) are mislocalized outside the conical CA lattice [101, 103, 111]. Although IN is known to bind DNA through several charged residues scattered throughout the protein (reviewed in [112]) and can bind to RNA *in vitro* with some specificity [113], why and how mutations within IN would specifically lead to mislocalization of vRNPs in virions remained enigmatic.

The recent development of allosteric integrase inhibitors (ALLINIs) reignited research in this area. While ALLINIs were initially developed to target IN binding to the cellular cofactor LEDGF, it was later shown that these compounds primarily act during particle maturation and lead to morphological aberrations in particles similar to those induced by the aforementioned Class II IN mutations [114–119]. Biochemical analysis of IN *in vitro* and in virions revealed that ALLINIs induce aberrant IN multimerization [103, 111, 120–123] through catalytic core domain–C-terminal domain interactions at the dimer–dimer interface [116]. By employing CLIP and complementary *in vitro* approaches, recent studies have shown that low-order multimers of IN binds to distinct

structured elements on the viral genome, including TAR, with high affinity [13]. Notably, while ALLINIs indirectly block these interactions by inducing IN oligomerization, mutations of basic amino acids within the C-terminal domain of IN can abolish IN-RNA binding directly without altering the multimeric state of IN. Inhibition of IN-RNA interactions leads to mislocalization vRNPs and IN outside the conical capsid core [124]. Surprisingly, CLIP experiments reveal that the pattern of NC binding on the vRNA genome seems to be unaffected by IN mutations or ALLINIs, despite the mislocalization of vRNPs in eccentric particles [124]. Together, these aberrations in virion morphology are accompanied by premature degradation of vRNPs and IN, and spatial separation of RT from vRNPs, explaining the early reverse transcription block of these particles in target cells [124]. Thus, CLIP has been key in unveiling the key role of IN-RNA interactions during virion morphogenesis that ensure the correct localization of core components inside the CA lattice during particle maturation.

Incorporation of APOBEC3 proteins into virions

While viral RNAs contain sequence and structural elements that regulate key steps in HIV-1 replication, they can also be recognized by host defense mechanisms. Infiltration of the host APOBEC3 (A3) proteins into virus particles by binding viral RNAs is a prime example of this process. A3 proteins are a family of cytidine deaminases that inhibit the replication of a broad range of viruses and retroelements (reviewed in [125, 126]). A3s inhibit replication in two ways. One mechanism involves the deamination of cytidines to uridines in (–) strand DNA during reverse transcription, resulting in the accumulation dG-to-dA mutations on the coding strand [127–130] and lethal hypermutation. Additionally, A3 proteins have been shown to induce a deamination-independent block, by binding to reverse transcriptase and inhibiting reverse transcription [131–135]. Packaging of A3 proteins into HIV-1 virions is required for their antiviral activity and depends on the NC domain of Gag and its associated RNA [136–141]. A3 proteins appear to be promiscuous RNA binding proteins and it has been difficult to determine whether they selectively target viral or cellular RNAs to infiltrate into particles. For example, there is evidence to indicate that viral genome [142], 7SL RNA, a cellular RNA that is normally part of the signal recognition particle and is enriched in retroviral particles [143], or both cellular and viral RNAs [140, 141] can mediate packaging of A3G into particles. As many of these studies largely relied only on genetic assays, whether A3 proteins exhibit any preference towards a specific set of RNAs, or sequence features within them in a relevant setting remained unknown. Nevertheless, the presence

of a discrete RNA binding domain in A3G implies some level of selectivity in RNA binding, much like other RBPs [144, 145].

Three recent studies employing CLIP have provided insight into the RNA-binding properties of several A3 proteins in infected cells and in virions [146, 147]. The earlier iCLIP-based study indicated that although the viral genome is enriched amongst A3F and A3G-bound RNAs, a diverse set of RNAs could drive the incorporation of A3F and A3G into virions [146]. A subsequent PAR-CLIP-based study confirmed some of these findings in that A3 proteins were shown to bind similar classes of cellular RNAs and HIV-1 RNA was bound preferentially over cellular RNAs in infected cells. However, the PAR-CLIP approach provided a higher resolution assessment of A3-RNA interactions in cells, likely due to the ability to more accurately identify the site of crosslinking. Most importantly, detailed analysis of A3 binding sites revealed that the A3 proteins partly mimic the RNA-binding specificity of NC, in that they target RNA sequences that are G-rich and A-rich [147]. This model provides some explanation of how A3 proteins are incorporated efficiently into virions in the presence of a vast excess of cellular RNA molecules. This model invokes a bias in the binding of A3 proteins to RNA molecules of a given sequence composition, as a way of maintaining broad RNA binding specificity, while removing the need to occupy all mRNA sequences present in an infected cell. One recent study, the first to reveal a crystal structure of an A3 protein in complex with an RNA showed that the A3H protein has a particular propensity to bind to seven-nucleotide duplexes, in a manner that was independent of the nucleotide sequences forming the duplexes [148]. Accompanying CLIP experiments showed that the sites in the HIV-1 genome to which A3H was most frequently bound were invariably predicted to contain 7nt duplexes.

Role of zinc finger antiviral protein (ZAP) in imposing compositional bias on viral genomes

The genomes of vertebrates are marked with a paucity of CG dinucleotides [149], a feature that is well understood to have been caused by the action of CG-specific DNA methyl transferases and methyl-cytosine deamination, over hundreds of millions of years. More mysteriously, inspection of the composition of the genomes of RNA viruses in vertebrates, reveal that they mimic this CG-poor state, even though they are not substrates for DNA methyl transferases [150–152]. Recent work, in the context of HIV-1 has shown that the paucity of CG dinucleotides is essential for viral replication, and that the appearance of too many CG dinucleotides in the viral genome causes cytoplasmic depletion of viral RNA [153]. The apparently destabilizing effect of CG dinucleotides

was cumulative, and found to be induced by CG dinucleotides in both translated portions of an mRNA and also in untranslated exons. Further experiments showed that zinc finger antiviral protein (ZAP) [154] a protein that encodes four CCCH zinc fingers in its N-terminal domain is essential for mediating the deleterious effects of CG dinucleotides. Indeed, HIV-1 mutants containing segments whose CG-content mimicked a random nucleotide sequence could not replicate in unmanipulated cells containing an intact *ZAP* gene, but could replicate with wild-type kinetics in cells rendered ZAP-deficient by CRISPR-Cas9 editing [153].

While previous studies had shown that ZAP had antiviral activity against a number of RNA viruses, several conventional techniques could not identify a common sequence motif or RNA feature that could explain how ZAP was able to specifically target viral RNA sequences [154, 155]. RNA elements that could confer sensitivity to ZAP when inserted into a reporter RNA were large, leading to the proposal that a specific tertiary structure constituted a ZAP recognition site. However, RNA elements that conferred sensitivity to ZAP did so in both orientations [156], effectively refuting these models. CLIP experiments showed unambiguously that ZAP binds directly and selectively to RNA elements that contain CG dinucleotides, but exhibits no preferential binding to RNA elements containing GC or any other dinucleotide [153]. Interestingly, these results suggest that ZAP arose to exploit a compositional difference between host mRNAs and RNAs from viruses have high CG content. However, the dinucleotide composition of HIV-1, appears to have adapted to evade ZAP and it is possible that ZAP has driven the purging of CG dinucleotides from a range of RNA viruses.

Identification of m⁶A marks on HIV-1 RNAs

Like proteins and DNA, RNA can undergo a number of chemical modifications that subsequently affect its metabolism, function and localization. While tRNAs and rRNAs are subjected to the most diverse set of modifications, recent transcriptome-wide studies revealed the presence of numerous mRNA modifications [157–162]. Methylation of adenosine at the N6 position (m⁶A) is the most prevalent of these and has been proposed to regulate several aspects of RNA metabolism, including splicing, nuclear export, localization, stability and translation [163]. m⁶A modification is catalyzed by a nuclear “writer” protein complex, composed of two methyltransferase-like enzymes, METTL3 and METTL4, and their cofactor Wilms tumor 1-associated protein (WTAP). This modification can be reversed by two RNA demethylases, or “erasers”, ALKBH5 (a-ketoglutarate-dependent dioxygenase homolog 5) and FTO (fat mass and obesity

associated). m⁶A-modifications on mRNAs can be bound by three related cytosolic “reader” proteins called YTH-domain containing family 1 (YTHDF1), YTHDF2, and YTHDF3. Exactly how binding of these proteins on modified nucleotides regulate mRNA metabolism is currently unknown. Nonetheless, m⁶A modifications can be found on mRNAs of diverse viruses that replicate in the nucleus, including SV40 [164], adenovirus [165, 166], influenza A virus [167] as well as retroviruses such as avian sarcoma virus [168] and Rous sarcoma virus [169, 170]. Until recently, whether HIV-1 mRNAs contained m⁶A modifications and how this affected virus replication was not known.

Three recent studies have addressed this question by immunoprecipitating methylated HIV-1 RNAs from infected cells using a m⁶A-specific antibody followed by high throughput sequencing of the immunoprecipitated mRNAs [171–173]. Strikingly, there was virtually no overlap in the m⁶A sites identified in these independent studies. This lack of consistency can in part be explained by the different approaches taken. The first published study that has utilized a RIP-seq approach, in which m⁶A-modified RNAs were immunoprecipitated from cell lysates and sequenced, found m⁶A modifications throughout the viral genome [172]. In contrast, a later study, which included a PAR-CLIP-based crosslinking step following immunoprecipitation of m⁶A-modified RNAs, found that the m⁶A modifications were exclusively localized within the viral 3′ UTRs [171]. Importantly, parallel YTHDF PAR-CLIP experiments conducted in this latter study revealed binding sites at or near the modified nucleotides, reinforcing the findings from m⁶A-specific immunoprecipitations [171]. A third study similarly coupled YTHDF HITS-CLIP with m⁶A-seq [173] and identified putative modification sites within 3′ and 5′ UTRs of HIV-1 mRNAs. Notably, none of these sites overlapped with those identified in the former studies. Thus, while CLIP methodologies have been highly instrumental in identification of m⁶A sites on HIV-1 RNAs, cross-validation of reagents (i.e. cell lines, viruses, m⁶A antibodies) and methods (i.e. m⁶A-seq, PAR- vs. HITS-CLIP) will be necessary to reach to a consensus in future studies.

Conclusions

Application of the CLIP methods to questions in retrovirology will undoubtedly continue to increase, given the large number of RBPs that are known and continuing to emerge as key regulators of retroviral replication. Several poorly explored areas in retrovirology will benefit from these approaches. One of the immediate applications of this methodology will be in determining how the alternative splicing of HIV-1 transcripts is regulated by cellular hnRNP and SR splicing-regulatory proteins. Although

the families of hnRNP and SR proteins constitute more than 50 proteins, only a few have been shown to play roles in HIV-1 RNA splicing. In addition, none of the studies performed to date determined where on viral RNAs these proteins bind. Instead, *in vitro* splicing reporters and genetic assays were used, which are prone to artefacts. Another exciting area of research where CLIP and related methodologies may make a major impact is the sensing of viral nucleic acids in infected cells. HIV-1 infection induces high levels of interferon and other cytokines during the acute phase of infection, suggesting that viral nucleic acids are sensed in infected cells. While a few isolated studies indicated that viral reverse transcription products or RNA elements can be sensed in certain settings, it remains to be determined what features of viral nucleic acids are sensed and whether viral RNA or DNA elicits an inflammatory response. While A3 proteins provide a good example of how viral RNAs can be targeted by antiviral host proteins, it is plausible that many other cellular proteins that can recognize and target viral RNAs. CLIP will be a key tool in unveiling novel cellular proteins that participate at the HIV-1-host interface. Finally, although CLIP has so far only been applied to HIV-1 biology, it will certainly find broad applications in retrovirology and virology more generally as the methods and next-generation sequencing becomes more accessible.

Authors' contributions

Both authors read and approved the final manuscript.

Author details

¹ Howard Hughes Medical Institute and Laboratory of Retrovirology, The Rockefeller University, New York, NY 10065, USA. ² Department of Molecular Microbiology, Washington University School of Medicine, Saint Louis, MO 63110, USA.

Acknowledgements

Not applicable.

Competing interests

The authors declare they have no competing interests.

Availability of data and materials

Not applicable.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

Funding was provided by National Institute of General Medical Sciences (Grant Nos. U54 GM-103297, GM-122458) and National Institute of Allergy and Infectious Diseases (Grant No. AI-50111).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 14 March 2018 Accepted: 17 April 2018
Published online: 02 May 2018

References

- Karn J, Stoltzfus CM. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harb Perspect Med*. 2012;2:a006916.
- Stoltzfus CM. Chapter 1. Regulation of HIV-1 alternative RNA splicing and its role in virus replication. *Adv Virus Res*. 2009;74:1–40.
- Leblanc J, Weil J, Beemon K. Posttranscriptional regulation of retroviral gene expression: primary RNA transcripts play three roles as pre-mRNA, mRNA, and genomic RNA. *Wiley Interdiscip Rev RNA*. 2013;4:567–80.
- Cullen BR. Nuclear mRNA export: insights from virology. *Trends Biochem Sci*. 2003;28:419–24.
- Briggs JA, Krausslich HG. The molecular architecture of HIV. *J Mol Biol*. 2011;410:491–500.
- Sundquist WI, Krausslich HG. HIV-1 assembly, budding, and maturation. *Cold Spring Harb Perspect Med*. 2012;2:a006924.
- Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. CLIP identifies Nova-regulated RNA networks in the brain. *Science*. 2003;302:1212–5.
- Ule J, Jensen K, Mele A, Darnell RB. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods*. 2005;37:376–86.
- Licalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, Darnell JC, Darnell RB. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*. 2008;456:464–9.
- Zhang C, Darnell RB. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol*. 2011;29:607–14.
- Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*. 2010;141:129–41.
- Jungkamp AC, Stoekius M, Mecnas D, Grun D, Mastrobuoni G, Kempa S, Rajewsky N. In vivo and transcriptome-wide identification of RNA binding protein target sites. *Mol Cell*. 2011;44:828–40.
- Kessl JJ, Kutluay SB, Townsend D, Rebusburg S, Slaughter A, Larue RC, Shkriabai N, Bakouche N, Fuchs JR, Bieniasz PD, Kvaratskhelia M. HIV-1 integrase binds the viral RNA genome and is essential during virion morphogenesis. *Cell*. 2016;166:1257–68.
- Kutluay SB, Bieniasz PD. Analysis of HIV-1 Gag-RNA interactions in cells and virions by CLIP-seq. *Methods Mol Biol*. 2016;1354:119–31.
- Gottwein E, Corcoran DL, Mukherjee N, Skalsky RL, Hafner M, Nusbaum JD, Shamulailatpam P, Love CL, Dave SS, Tuschl T, Ohler U, Cullen BR. Viral microRNA targetome of KSHV-infected primary effusion lymphoma cell lines. *Cell Host Microbe*. 2011;10:515–26.
- Hafner M, Lianoglou S, Tuschl T, Betel D. Genome-wide identification of miRNA targets by PAR-CLIP. *Methods*. 2012;58:94–105.
- Jaskiewicz L, Bilen B, Hausser J, Zavolan M. Argonaute CLIP—a method to identify in vivo targets of miRNAs. *Methods*. 2012;58:106–12.
- Kishore S, Jaskiewicz L, Burger L, Hausser J, Khorshid M, Zavolan M. A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat Methods*. 2011;8:559–64.
- Chang X, Li B, Rao A. RNA-binding protein hnRNPLL regulates mRNA splicing and stability during B-cell to plasma-cell differentiation. *Proc Natl Acad Sci USA*. 2015;112:E1888–97.
- Degrauwe N, Schlumpf TB, Janiszewska M, Martin P, Caudey A, Proviero P, Riggi N, Suva ML, Paro R, Stamenkovic I. The RNA binding protein IMP2 preserves glioblastoma stem cells by preventing let-7 target gene silencing. *Cell Rep*. 2016;15:1634–47.
- Kim KK, Yang Y, Zhu J, Adelstein RS, Kawamoto S. Rbfox3 controls the biogenesis of a subset of microRNAs. *Nat Struct Mol Biol*. 2014;21:901–10.
- Schonemann L, Kuhn U, Martin G, Schafer P, Gruber AR, Keller W, Zavolan M, Wahle E. Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes Dev*. 2014;28:2381–93.
- Uemura Y, Oshima T, Yamamoto M, Reyes CJ, Costa Cruz PH, Shibuya T, Kawahara Y. Matr3 binds directly to intronic pyrimidine-rich sequences and controls alternative splicing. *Genes Cells*. 2017. <https://doi.org/10.1111/gtc.12512>.
- Lee AS, Kranzusch PJ, Cate JH. eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature*. 2015;522:111–4.
- Burger K, Muhl B, Kellner M, Rohmoser M, Gruber-Eber A, Windhager L, Friedel CC, Dolken L, Eick D. 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. *RNA Biol*. 2013;10:1623–30.
- Testa SM, Disney MD, Turner DH, Kierzek R. Thermodynamics of RNA-RNA duplexes with 2- or 4-thiouridines: implications for antisense design and targeting a group I intron. *Biochemistry*. 1999;38:16655–62.
- Urlaub H, Hartmuth K, Luhrmann R. A two-tracked approach to analyze RNA-protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles. *Methods*. 2002;26:170–81.
- Sugimoto Y, Konig J, Hussain S, Zupan B, Curk T, Frye M, Ule J. Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol*. 2012;13:R67.
- Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol*. 2010;17:909–15.
- Haberman N, Huppertz I, Attig J, Konig J, Wang Z, Hauer C, Hentze MW, Kulozik AE, Le Hir H, Curk T, Sibley CR, Zarnack K, Ule J. Insights into the design and interpretation of iCLIP experiments. *Genome Biol*. 2017;18:7.
- Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundaraman B, Blue SM, Nguyen TB, Surka C, Elkins K, Stanton R, Rigo F, Guttman M, Yeo GW. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods*. 2016;13:508–14.
- Weyn-Vanhenryck SM, Mele A, Yan Q, Sun S, Farny N, Zhang Z, Xue C, Herre M, Silver PA, Zhang MQ, Krainer AR, Darnell RB, Zhang C. HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep*. 2014;6:1139–52.
- Flynn RA, Martin L, Spitalo RC, Do BT, Sagan SM, Zarnegar B, Qu K, Khavari PA, Quake SR, Sarnow P, Chang HY. Dissecting noncoding and pathogen RNA-protein interactomes. *RNA*. 2015;21:135–43.
- Lee FCY, Ule J. Advances in CLIP technologies for studies of protein-RNA interactions. *Mol Cell*. 2018;69:354–69.
- Zarnegar BJ, Flynn RA, Shen Y, Do BT, Chang HY, Khavari PA. irCLIP platform for efficient characterization of protein-RNA interactions. *Nat Methods*. 2016;13:489–92.
- Kutluay SB, Zang T, Blanco-Melo D, Powell C, Jannain D, Errando M, Bieniasz PD. Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. *Cell*. 2014;159:1096–109.
- Garzia A, Meyer C, Morozov P, Sajek M, Tuschl T. Optimization of PAR-CLIP for transcriptome-wide identification of binding sites of RNA-binding proteins. *Methods*. 2017;118–119:24–40.
- Khorshid M, Rodak C, Zavolan M. CLIPZ: a database and analysis environment for experimentally determined binding sites of RNA-binding proteins. *Nucleic Acids Res*. 2011;39:D245–52.
- Moore MJ, Zhang C, Gantman EC, Mele A, Darnell JC, Darnell RB. Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. *Nat Protoc*. 2014;9:263–93.
- Maragkakis M, Alexiou P, Nakaya T, Mourelatos Z. CLIPSeqTools—a novel bioinformatics CLIP-seq analysis suite. *RNA*. 2016;22:1–9.
- Bottini S, Pratella D, Grandjean V, Repetto E, Trabucchi M. Recent computational developments on CLIP-seq data analysis and microRNA targeting implications. *Brief Bioinform*. 2017. <https://doi.org/10.1093/bib/bbx063>.
- Uhl M, Houwaart T, Corrado G, Wright PR, Backofen R. Computational analysis of CLIP-seq data. *Methods*. 2017;118–119:60–72.
- Liu Q, Zhong X, Madison BB, Rustgi AK, Shyr Y. Assessing computational steps for CLIP-seq data analysis. *Biomed Res Int*. 2015;2015:196082.
- Reyes-Herrera PH, Ficarra E. Computational methods for CLIP-seq data processing. *Bioinform Biol Insights*. 2014;8:199–207.
- Chen C, Khaleel SS, Huang H, Wu CH. Software for pre-processing Illumina next-generation sequencing short read sequences. *Source Code Biol Med*. 2014;9:8.

46. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
47. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011;27:863–4.
48. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10:R25.
49. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
50. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
51. Smith AD, Chung WY, Hodges E, Kendall J, Hannon G, Hicks J, Xuan Z, Zhang MQ. Updates to the RMAP short-read mapping software. *Bioinformatics*. 2009;25:2841–2.
52. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25:1105–11.
53. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*. 2010;26:873–81.
54. Li R, Li Y, Kristiansen K, Wang J. SOAP: short oligonucleotide alignment program. *Bioinformatics*. 2008;24:713–4.
55. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
56. Uren PJ, Bahrami-Samani E, Burns SC, Qiao M, Karginov FV, Hodges E, Hannon GJ, Sanford JR, Penava LO, Smith AD. Site identification in high-throughput RNA-protein interaction data. *Bioinformatics*. 2012;28:3013–20.
57. Lovci MT, Ghanem D, Marr H, Arnold J, Gee S, Parra M, Liang TY, Stark TJ, Gehman LT, Hoon S, Massier KB, Pratt GA, Black DL, Gray JW, Conboy JG, Yeo GW. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat Struct Mol Biol*. 2013;20:1434–42.
58. Chen B, Yun J, Kim MS, Mendell JT, Xie Y. PIPE-CLIP: a comprehensive online tool for CLIP-seq data analysis. *Genome Biol*. 2014;15:R18.
59. Althammer S, Gonzalez-Vallinas J, Ballare C, Beato M, Eyraes E. Pycos: a versatile toolkit for the analysis of high-throughput sequencing data. *Bioinformatics*. 2011;27:3333–40.
60. Corcoran DL, Georgiev S, Mukherjee N, Gottwein E, Skalsky RL, Keene JD, Ohler U. PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol*. 2011;12:R79.
61. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. 2009;37:W202–8.
62. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38:576–89.
63. Georgiev S, Boyle AP, Jayasurya K, Ding X, Mukherjee S, Ohler U. Evidence-ranked motif identification. *Genome Biol*. 2010;11:R19.
64. Jouvenet N, Simon SM, Bieniasz PD. Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles. *Proc Natl Acad Sci USA*. 2009;106:19114–9.
65. Kutluay SB, Bieniasz PD. Analysis of the initiating events in HIV-1 particle assembly and genome packaging. *PLoS Pathog*. 2010;6:e1001200.
66. Clever JL, Parslow TG. Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J Virol*. 1997;71:3407–14.
67. Laham-Karam N, Bacharach E. Transduction of human immunodeficiency virus type 1 vectors lacking encapsidation and dimerization signals. *J Virol*. 2007;81:10687–98.
68. McBride MS, Panganiban AT. Position dependence of functional hairpins important for human immunodeficiency virus type 1 RNA encapsidation in vivo. *J Virol*. 1997;71:2050–8.
69. Berkowitz RD, Hammarskjöld ML, Helga-Maria C, Rekosh D, Goff SP. 5' regions of HIV-1 RNAs are not sufficient for encapsidation: implications for the HIV-1 packaging signal. *Virology*. 1995;212:718–23.
70. Chamanian M, Purzycka KJ, Wille PT, Ha JS, McDonald D, Gao Y, Le Grice SF, Arts EJ. A cis-acting element in retroviral genomic RNA links Gag-Pol ribosomal frameshifting to selective viral RNA encapsidation. *Cell Host Microbe*. 2013;13:181–92.
71. Das AT, Klaver B, Klasens BI, van Wamel JL, Berkhout B. A conserved hairpin motif in the R-U5 region of the human immunodeficiency virus type 1 RNA genome is essential for replication. *J Virol*. 1997;71:2346–56.
72. McBride MS, Schwartz MD, Panganiban AT. Efficient encapsidation of human immunodeficiency virus type 1 vectors and further characterization of cis elements required for encapsidation. *J Virol*. 1997;71:4544–54.
73. Richardson JH, Child LA, Lever AM. Packaging of human immunodeficiency virus type 1 RNA requires cis-acting sequences outside the 5' leader region. *J Virol*. 1993;67:3997–4005.
74. Muriaux D, Mirro J, Harvin D, Rein A. RNA is a structural element in retrovirus particles. *Proc Natl Acad Sci USA*. 2001;98:5246–51.
75. Rulli SJ Jr, Hibbert CS, Mirro J, Pederson T, Biswal S, Rein A. Selective and nonselective packaging of cellular RNAs in retrovirus particles. *J Virol*. 2007;81:6623–31.
76. Keane SC, Heng X, Lu K, Kharytonchyk S, Ramakrishnan V, Carter G, Barton S, Hosic A, Florwick A, Santos J, Bolden NC, McCowin S, Case DA, Johnson BA, Salemi M, Telesnitsky A, Summers MF. RNA structure. Structure of the HIV-1 RNA packaging signal. *Science*. 2015;348:917–21.
77. Keane SC, Van V, Frank HM, Sciandra CA, McCowin S, Santos J, Heng X, Summers MF. NMR detection of intermolecular interaction sites in the dimeric 5'-leader of the HIV-1 genome. *Proc Natl Acad Sci USA*. 2016;113:13033–8.
78. Lu K, Heng X, Garyu L, Monti S, Garcia EL, Kharytonchyk S, Dorjsuren B, Kulandaivel G, Jones S, Hiremath A, Divakaruni SS, LaCotti C, Barton S, Tummlillo D, Hosic A, Edme K, Albrecht S, Telesnitsky A, Summers MF. NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science*. 2011;334:242–5.
79. Becker JT, Sherer NM. Subcellular localization of HIV-1 gag-pol mRNAs regulates sites of virion assembly. *J Virol*. 2017;91:e02315–6.
80. Liu Y, Nikolaitchik OA, Rahman SA, Chen J, Pathak VK, Hu WS. HIV-1 sequence necessary and sufficient to package non-viral RNAs into HIV-1 particles. *J Mol Biol*. 2017. <https://doi.org/10.1016/j.jmb.2017.06.018>.
81. Comas-Garcia M, Datta SA, Baker L, Varma R, Gudla PR, Rein A. Dissection of specific binding of HIV-1 Gag to the 'packaging signal' in viral RNA. *Elife*. 2017. <https://doi.org/10.7554/eLife.27055>.
82. Alfadhli A, Still A, Barklis E. Analysis of human immunodeficiency virus type 1 matrix binding to membranes and nucleic acids. *J Virol*. 2009;83:12196–203.
83. Chukkappalli V, Oh SJ, Ono A. Opposing mechanisms involving RNA and lipids regulate HIV-1 Gag membrane binding through the highly basic region of the matrix domain. *Proc Natl Acad Sci USA*. 2010;107:1600–5.
84. Cimarelli A, Luban J. Translation elongation factor 1- α interacts specifically with the human immunodeficiency virus type 1 Gag polyprotein. *J Virol*. 1999;73:5388–401.
85. Levin JG, Mitra M, Mascarenhas A, Musier-Forsyth K. Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol*. 2010;7:754–74.
86. Ramalingam D, Duclair S, Datta SA, Ellington A, Rein A, Prasad VR. RNA aptamers directed to human immunodeficiency virus type 1 Gag polyprotein bind to the matrix and nucleocapsid domains and inhibit virus production. *J Virol*. 2011;85:305–14.
87. Ott DE, Coren LV, Gagliardi TD. Redundant roles for nucleocapsid and matrix RNA-binding sequences in human immunodeficiency virus type 1 assembly. *J Virol*. 2005;79:13839–47.
88. Chukkappalli V, Inlora J, Todd GC, Ono A. Evidence in support of RNA-mediated inhibition of phosphatidylserine-dependent HIV-1 Gag membrane binding in cells. *J Virol*. 2013;87:7155–9.
89. Hill CP, Worthylake D, Bancroft DP, Christensen AM, Sundquist WI. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. *Proc Natl Acad Sci USA*. 1996;93:3099–104.
90. Saad JS, Miller J, Tai J, Kim A, Ghanam RH, Summers MF. Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. *Proc Natl Acad Sci USA*. 2006;103:11364–9.
91. Shkriabai N, Datta SA, Zhao Z, Hess S, Rein A, Kvaratskhelia M. Interactions of HIV-1 Gag with assembly cofactors. *Biochemistry*. 2006;45:4077–83.
92. Zhou W, Parent LJ, Wills JW, Resh MD. Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. *J Virol*. 1994;68:2556–69.

93. Holmes M, Zhang F, Bieniasz PD. Single-cell and single-cycle analysis of HIV-1 Replication. *PLoS Pathog.* 2015;11:e1004961.
94. Todd GC, Duchon A, Inlora J, Olson ED, Musier-Forsyth K, Ono A. Inhibition of HIV-1 Gag-membrane interactions by specific RNAs. *RNA.* 2017;23:395–405.
95. Kypr J, Mrazek J. Unusual codon usage of HIV. *Nature.* 1987;327:20.
96. Grantham P, Perrin P. AIDS virus and HTLV-I differ in codon choices. *Nature.* 1986;319:727–8.
97. Sharp PM. What can AIDS virus codon usage tell us? *Nature.* 1986;324:114.
98. Berkhout B, van Hemert FJ. The unusual nucleotide content of the HIV RNA genome results in a biased amino acid composition of HIV proteins. *Nucleic Acids Res.* 1994;22:1705–11.
99. Bukovsky A, Gottlinger H. Lack of integrase can markedly affect human immunodeficiency virus type 1 particle production in the presence of an active viral protease. *J Virol.* 1996;70:6820–5.
100. Engelman A. In vivo analysis of retroviral integrase structure and function. *Adv Virus Res.* 1999;52:411–26.
101. Engelman A, Englund G, Orenstein JM, Martin MA, Craigie R. Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. *J Virol.* 1995;69:2729–36.
102. Johnson BC, Metifiot M, Ferris A, Pommier Y, Hughes SH. A homology model of HIV-1 integrase and analysis of mutations designed to test the model. *J Mol Biol.* 2013;425:2133–46.
103. Jurado KA, Wang H, Slaughter A, Feng L, Kessl JJ, Koh Y, Wang W, Ballandras-Colas A, Patel PA, Fuchs JR, Kvaratskhelia M, Engelman A. Allosteric integrase inhibitor potency is determined through the inhibition of HIV-1 particle maturation. *Proc Natl Acad Sci USA.* 2013;110:8690–5.
104. Mohammed KD, Topper MB, Muesing MA. Sequential deletion of the integrase (Gag-Pol) carboxyl terminus reveals distinct phenotypic classes of defective HIV-1. *J Virol.* 2011;85:4654–66.
105. Lu R, Limon A, Devroe E, Silver PA, Cherepanov P, Engelman A. Class II integrase mutants with changes in putative nuclear localization signals are primarily blocked at a postnuclear entry step of human immunodeficiency virus type 1 replication. *J Virol.* 2004;78:12735–46.
106. Lu R, Ghory HZ, Engelman A. Genetic analyses of conserved residues in the carboxyl-terminal domain of human immunodeficiency virus type 1 integrase. *J Virol.* 2005;79:10356–68.
107. Limon A, Devroe E, Lu R, Ghory HZ, Silver PA, Engelman A. Nuclear localization of human immunodeficiency virus type 1 preintegration complexes (PICs): V165A and R166A are pleiotropic integrase mutants primarily defective for integration, not PIC nuclear import. *J Virol.* 2002;76:10598–607.
108. Leavitt AD, Robles G, Alesandro N, Varmus HE. Human immunodeficiency virus type 1 integrase mutants retain in vitro integrase activity yet fail to integrate viral DNA efficiently during infection. *J Virol.* 1996;70:721–8.
109. Jenkins TM, Engelman A, Ghirlando R, Craigie R. A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization. *J Biol Chem.* 1996;271:7712–8.
110. Shehu-Xhilaga M, Hill M, Marshall JA, Kappes J, Crowe SM, Mak J. The conformation of the mature dimeric human immunodeficiency virus type 1 RNA genome requires packaging of pol protein. *J Virol.* 2002;76:4331–40.
111. Fontana J, Jurado KA, Cheng N, Ly NL, Fuchs JR, Gorelick RJ, Engelman AN, Steven AC. Distribution and redistribution of HIV-1 nucleocapsid protein in immature, mature, and integrase-inhibited virions: a role for integrase in maturation. *J Virol.* 2015. <https://doi.org/10.1128/JVI.01522-15>.
112. Engelman A, Cherepanov P. Retroviral integrase structure and DNA recombination mechanism. *Microbiol Spectr.* 2014;2:1–22.
113. Allen P, Worland S, Gold L. Isolation of high-affinity RNA ligands to HIV-1 integrase from a random pool. *Virology.* 1995;209:327–36.
114. Christ F, Voet A, Marchand A, Nicolet S, Desimie BA, Marchand D, Bardiot D, Van der Veken NJ, Van Remoortel B, Strelkov SV, De Maeyer M, Chaltin P, Debyser Z. Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. *Nat Chem Biol.* 2010;6:442–8.
115. Fader LD, Malenfant E, Parisien M, Carson R, Bilodeau F, Landry S, Pesant M, Brochu C, Morin S, Chabot C, Halmos T, Bousquet Y, Bailey MD, Kawai SH, Coulombe R, LaPlante S, Jakalian A, Bhardwaj PK, Wernic D, Schroeder P, Amad M, Edwards P, Garneau M, Duan J, Cordingley M, Bethell R, Mason SW, Bos M, Bonneau P, Poupart MA, Faucher AM, Simoneau B, Fenwick C, Yoakim C, Tsantrizos Y. Discovery of BI 224436, a noncatalytic site integrase inhibitor (NCINI) of HIV-1. *ACS Med Chem Lett.* 2014;5:422–7.
116. Gupta K, Brady T, Dyer BM, Malani N, Hwang Y, Male F, Nolte RT, Wang L, Velthuisen E, Jeffrey J, Van Duyn GD, Bushman FD. Allosteric inhibition of human immunodeficiency virus integrase: late block during viral replication and abnormal multimerization involving specific protein domains. *J Biol Chem.* 2014;289:20477–88.
117. Kessl JJ, Jena N, Koh Y, Taskent-Sezgin H, Slaughter A, Feng L, de Silva S, Wu L, Le Grice SF, Engelman A, Fuchs JR, Kvaratskhelia M. Multimode, cooperative mechanism of action of allosteric HIV-1 integrase inhibitors. *J Biol Chem.* 2012;287:16801–11.
118. Le Rouzic E, Bonnard D, Chasset S, Bruneau JM, Chevreuil F, Le Strat F, Nguyen J, Beauvoir R, Amadori C, Brias J, Vomscheid S, Eiler S, Levy N, Delelis O, Deprez E, Saib A, Zamborlini A, Emiliani S, Ruff M, Ledoussal B, Moreau F, Benarous R. Dual inhibition of HIV-1 replication by integrase-LEDGF allosteric inhibitors is predominant at the post-integration stage. *Retrovirology.* 2013;10:144.
119. van Bel N, van der Velden Y, Bonnard D, Le Rouzic E, Das AT, Benarous R, Berkhout B. The allosteric HIV-1 integrase inhibitor BI-D affects virion maturation but does not influence packaging of a functional RNA genome. *PLoS ONE.* 2014;9:e103552.
120. Balakrishnan M, Yant SR, Tsai L, O'Sullivan C, Bam RA, Tsai A, Niedziela-Majka A, Stray KM, Sakowicz R, Cihlar T. Non-catalytic site HIV-1 integrase inhibitors disrupt core maturation and induce a reverse transcription block in target cells. *PLoS ONE.* 2013;8:e74163.
121. Desimie BA, Schrijvers R, Demeulemeester J, Borrenberghs D, Weydert C, Thys W, Vets S, Van Remoortel B, Hofkens J, De Rijck J, Hendrix J, Bannert N, Gijssbers R, Christ F, Debyser Z. LEDGIs inhibit late stage HIV-1 replication by modulating integrase multimerization in the virions. *Retrovirology.* 2013;10:57.
122. Sharma A, Slaughter A, Jena N, Feng L, Kessl JJ, Fadel HJ, Malani N, Male F, Wu L, Poeschla E, Bushman FD, Fuchs JR, Kvaratskhelia M. A new class of multimerization selective inhibitors of HIV-1 integrase. *PLoS Pathog.* 2014;10:e1004171.
123. Gupta K, Turkki V, Sherrill-Mix S, Hwang Y, Eilers G, Taylor L, McDaniel C, Wang P, Temelkoff D, Nolte RT, Velthuisen E, Jeffrey J, Van Duyn GD, Bushman FD. Structural basis for inhibitor-induced aggregation of HIV integrase. *PLoS Biol.* 2016;14:e1002584.
124. Madison MK, Lawson DQ, Elliott J, Ozanturk AN, Koneru PC, Townsend D, Errando M, Kvaratskhelia M, Kutluay SB. Allosteric HIV-1 integrase inhibitors lead to premature degradation of the viral RNA genome and integrase in target cells. *J Virol.* 2017. <https://doi.org/10.1128/JVI.00821-17>.
125. Harris RS, Dudley JP. APOBECs and virus restriction. *Virology.* 2015;479–480:131–45.
126. Stavrou S, Ross SR. APOBEC3 proteins in viral immunity. *J Immunol.* 2015;195:4565–70.
127. Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS, Malim MH. DNA deamination mediates innate immunity to retroviral infection. *Cell.* 2003;113:803–9.
128. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature.* 2003;424:99–103.
129. Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature.* 2003;424:94–8.
130. Lecossier D, Bouchonnet F, Clavel F, Hance AJ. Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science.* 2003;300:1112.
131. Holmes RK, Koning FA, Bishop KN, Malim MH. APOBEC3F can inhibit the accumulation of HIV-1 reverse transcription products in the absence of hypermutation. Comparisons with APOBEC3G. *J Biol Chem.* 2007;282:2587–95.
132. Iwatani Y, Chan DS, Wang F, Maynard KS, Sugiura W, Gronenborn AM, Rouzina I, Williams MC, Musier-Forsyth K, Levin JG. Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. *Nucleic Acids Res.* 2007;35:7096–108.

133. Newman EN, Holmes RK, Craig HM, Klein KC, Lingappa JR, Malim MH, Sheehy AM. Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr Biol*. 2005;15:166–70.
134. Gillick K, Pollpeter D, Phalora P, Kim EY, Wolinsky SM, Malim MH. Suppression of HIV-1 infection by APOBEC3 proteins in primary human CD4(+) T cells is associated with inhibition of processive reverse transcription as well as excessive cytidine deamination. *J Virol*. 2013;87:1508–17.
135. Pollpeter D, Parsons M, Sobala AE, Coxhead S, Lang RD, Bruns AM, Papaioannou S, McDonnell JM, Apolonia L, Chowdhury JA, Horvath CM, Malim MH. Deep sequencing of HIV-1 reverse transcripts reveals the multifaceted antiviral functions of APOBEC3G. *Nat Microbiol*. 2018;3:220–33.
136. Alce TM, Popik W. APOBEC3G is incorporated into virus-like particles by a direct interaction with HIV-1 Gag nucleocapsid protein. *J Biol Chem*. 2004;279:34083–6.
137. Cen S, Guo F, Niu M, Saadatmand J, Deflassieux J, Kleiman L. The interaction between HIV-1 Gag and APOBEC3G. *J Biol Chem*. 2004;279:33177–84.
138. Luo K, Liu B, Xiao Z, Yu Y, Yu X, Gorelick R, Yu XF. Amino-terminal region of the human immunodeficiency virus type 1 nucleocapsid is required for human APOBEC3G packaging. *J Virol*. 2004;78:11841–52.
139. Schafer A, Bogerd HP, Cullen BR. Specific packaging of APOBEC3G into HIV-1 virions is mediated by the nucleocapsid domain of the gag polyprotein precursor. *Virology*. 2004;328:163–8.
140. Svarovskaia ES, Xu H, Mbisa JL, Barr R, Gorelick RJ, Ono A, Freed EO, Hu WS, Pathak VK. Human apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3G (APOBEC3G) is incorporated into HIV-1 virions through interactions with viral and nonviral RNAs. *J Biol Chem*. 2004;279:35822–8.
141. Zennou V, Perez-Caballero D, Gottlinger H, Bieniasz PD. APOBEC3G incorporation into human immunodeficiency virus type 1 particles. *J Virol*. 2004;78:12058–61.
142. Khan MA, Kao S, Miyagi E, Takeuchi H, Goila-Gaur R, Opi S, Gipson CL, Parslow TG, Ly H, Strebel K. Viral RNA is required for the association of APOBEC3G with human immunodeficiency virus type 1 nucleoprotein complexes. *J Virol*. 2005;79:5870–4.
143. Wang T, Tian C, Zhang W, Luo K, Sarkis PT, Yu L, Liu B, Yu Y, Yu XF. 7SL RNA mediates virion packaging of the antiviral cytidine deaminase APOBEC3G. *J Virol*. 2007;81:13112–24.
144. Auweter SD, Oberstrass FC, Allain FH. Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res*. 2006;34:4943–59.
145. Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol*. 2007;8:479–90.
146. Apolonia L, Schulz R, Curk T, Rocha P, Swanson CM, Schaller T, Ule J, Malim MH. Promiscuous RNA binding ensures effective encapsidation of APOBEC3 proteins by HIV-1. *PLoS Pathog*. 2015;11:e1004609.
147. York A, Kutluay SB, Errando M, Bieniasz PD. The RNA binding specificity of human APOBEC3 proteins resembles that of HIV-1 nucleocapsid. *PLoS Pathog*. 2016;12:e1005833.
148. Bohn JA, Thummar K, York A, Raymond A, Brown WC, Bieniasz PD, Hatzioannou T, Smith JL. APOBEC3H structure reveals an unusual mechanism of interaction with duplex RNA. *Nat Commun*. 2017;8:1021.
149. Karlin S, Mrazek J. Compositional differences within and between eukaryotic genomes. *Proc Natl Acad Sci*. 1997;94(19):10227–32.
150. Karlin S, Doerfler W, Cardon LR. Why is CpG suppressed in the genomes of virtually all small eukaryotic viruses but not in those of large eukaryotic viruses? *J Virol*. 1994;68:2889–97.
151. Rima BK, McFerran NV. Dinucleotide and stop codon frequencies in single-stranded RNA viruses. *J Gen Virol*. 1997;78(Pt 11):2859–70.
152. Greenbaum BD, Levine AJ, Bhanot G, Rabadan R. Patterns of evolution and host gene mimicry in influenza and other RNA viruses. *PLoS Pathog*. 2008;4:e1000079.
153. Takata MA, Goncalves-Carneiro D, Zang TM, Soll SJ, York A, Blanco-Melo D, Bieniasz PD. CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature*. 2017;550:124–7.
154. Gao G, Guo X, Goff SP. Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. *Science*. 2002;297:1703–6.
155. Muller S, Moller P, Bick MJ, Wurr S, Becker S, Gunther S, Kummerer BM. Inhibition of filovirus replication by the zinc finger antiviral protein. *J Virol*. 2007;81:2391–400.
156. Guo X, Carroll JW, Macdonald MR, Goff SP, Gao G. The zinc finger antiviral protein directly binds to specific viral mRNAs through the CCCH zinc finger motifs. *J Virol*. 2004;78:12781–7.
157. Carlile TM, Rojas-Duran MF, Zinshteyn B, Shin H, Bartoli KM, Gilbert WV. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature*. 2014;515:143–6.
158. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M, Sorek R, Rechavi G. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*. 2012;485:201–6.
159. Dominissini D, Nachtergaele S, Moshitch-Moshkovitz S, Peer E, Kol N, Ben-Haim MS, Dai Q, Di Segni A, Salmon-Divon M, Clark WC, Zheng G, Pan T, Solomon O, Eyal E, Hershkovitz V, Han D, Dore LC, Amariglio N, Rechavi G, He C. The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature*. 2016;530:441–6.
160. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*. 2012;149:1635–46.
161. Schwartz S, Bernstein DA, Mumbach MR, Jovanovic M, Herbst RH, Leon-Ricardo BX, Engreitz JM, Guttman M, Satija R, Lander ES, Fink G, Regev A. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell*. 2014;159:148–62.
162. Squires JE, Patel HR, Nusch S, Sibbritt T, Humphreys DT, Parker BJ, Suter CM, Preiss T. Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res*. 2012;40:5023–33.
163. Yue Y, Liu J, He C. RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. *Genes Dev*. 2015;29:1343–55.
164. Canaani D, Kahana C, Lavi S, Groner Y. Identification and mapping of N6-methyladenosine containing sequences in simian virus 40 RNA. *Nucleic Acids Res*. 1979;6:2879–99.
165. Hashimoto SI, Green M. Multiple methylated cap sequences in adenovirus type 2 early mRNA. *J Virol*. 1976;20:425–35.
166. Sommer S, Salditt-Georgieff M, Bachenheimer S, Darnell JE, Furuichi Y, Morgan M, Shatkin AJ. The methylation of adenovirus-specific nuclear and cytoplasmic RNA. *Nucleic Acids Res*. 1976;3:749–65.
167. Krug RM, Morgan MA, Shatkin AJ. Influenza viral mRNA contains internal N6-methyladenosine and 5'-terminal 7-methylguanosine in cap structures. *J Virol*. 1976;20:45–53.
168. Dimock K, Stoltzfus CM. Sequence specificity of internal methylation in B77 avian sarcoma virus RNA subunits. *Biochemistry*. 1977;16:471–8.
169. Beemon K, Keith J. Localization of N6-methyladenosine in the Rous sarcoma virus genome. *J Mol Biol*. 1977;113:165–79.
170. Kane SE, Beemon K. Precise localization of m6A in Rous sarcoma virus RNA reveals clustering of methylation sites: implications for RNA processing. *Mol Cell Biol*. 1985;5:2298–306.
171. Kennedy EM, Bogerd HP, Kornepati AV, Kang D, Ghoshal D, Marshall JB, Poling BC, Tsai K, Gokhale NS, Horner SM, Cullen BR. Posttranscriptional m(6)A editing of HIV-1 mRNAs enhances viral gene expression. *Cell Host Microbe*. 2016;19:675–85.
172. Lichinchi G, Gao S, Saletore Y, Gonzalez GM, Bansal V, Wang Y, Mason CE, Rana TM. Dynamics of the human and viral m(6)A RNA methylomes during HIV-1 infection of T cells. *Nat Microbiol*. 2016;1:16011.
173. Tirumuru N, Zhao BS, Lu W, Lu Z, He C, Wu L. N(6)-methyladenosine of HIV-1 RNA regulates viral infection and HIV-1 Gag protein expression. *Elife*. 2016. <https://doi.org/10.7554/eLife.15528>.