

Next-generation sequencing: A new avenue to understand viral RNA–protein interactions

Received for publication, November 11, 2021, and in revised form, April 1, 2022. Published, Papers in Press, April 9, 2022.
<https://doi.org/10.1016/j.jbc.2022.101924>

Yiyang Zhou^{1,*}, Stephanea L. Sotcheff¹, and Andrew L. Routh^{1,2,3}

From the ¹Department of Biochemistry and Molecular Biology, and ²Sealy Center for Structural Biology and Molecular Biophysics, The University of Texas Medical Branch, Galveston, Texas, USA; ³Institute for Human Infections and Immunity, University of Texas Medical Branch, Galveston, Texas, USA

Edited by Karin Musier-Forsyth

The genomes of RNA viruses present an astonishing source of both sequence and structural diversity. From intracellular viral RNA–host interfaces to interactions between the RNA genome and structural proteins in virus particles themselves, almost the entire viral lifecycle is accompanied by a myriad of RNA–protein interactions that are required to fulfill their replicative potential. It is therefore important to characterize such rich and dynamic collections of viral RNA–protein interactions to understand virus evolution and their adaptation to their hosts and environment. Recent advances in next-generation sequencing technologies have allowed the characterization of viral RNA–protein interactions, including both transient and conserved interactions, where molecular and structural approaches have fallen short. In this review, we will provide a methodological overview of the high-throughput techniques used to study viral RNA–protein interactions, their biochemical mechanisms, and how they evolved from classical methods as well as one another. We will discuss how different techniques have fueled virus research to characterize how viral RNA and proteins interact, both locally and on a global scale. Finally, we will present examples on how these techniques influence the studies of clinically important pathogens such as HIV-1 and SARS-CoV-2.

RNA viruses and their genomes exhibit incredible sequence and structural diversity and a correspondingly complex range of RNA–protein interactions. Inside infectious virus particles (*in virio*), the specific and often programmed interactions of RNA with viral (nucleo-)capsids orchestrate virus particle assembly, genome packaging and release, structural stability, and can regulate RNA replication and transcription (1–4). In addition, RNA viruses and their genomes are dependent on host cellular factors and actively reprogram the host cellular environment to support RNA replication and/or prevent immune clearance. Much of this reprogramming occurs through numerous interactions between viral RNAs (vRNAs) and host factors (5). Thus, vRNAs have multiple simultaneous roles and overlapping interactions with both host and viral proteins that must be carefully coordinated during the viral lifecycle.

Studying and characterizing this complex choreography presents an important but major technical challenge.

Classical techniques that probe RNA–protein interactions, such as immunoprecipitation (IP) (6, 7), crosslinking (7–11), EMSA (12, 13), and affinity-capture (14, 15), offer reliable biochemical solutions both *in vitro* and *in vivo*. However, these methods typically rely on prior knowledge or prediction of RNA–protein interaction sites, which requires researchers to target specific hypothesized interactions. This may miss important and/or unanticipated interaction partners. Similarly, due to lack of high-throughput analytical power, molecular methods also are limited in their capacity to identify the many possible RNA targets and binding sites of a protein-of-interest.

The remarkable power of high-throughput next-generation sequencing (NGS) combined with traditional “tried and true” biochemical techniques provides a complementary approach to study vRNA–protein interactions in a high-throughput, specific, and unbiased fashion. This has broad applications in studying the role and activities of vRNA-binding proteins within the host, as well as in elucidating the structure of macromolecular RNA–protein complexes within virus particles themselves. Furthermore, NGS-based methods can often provide nucleotide or near-nucleotide resolution.

As such, NGS is often deployed as a hypothesis-generating instead of hypothesis-dependent tool, allowing for detection of novel RNA–protein interactions within the host transcriptome (*in vivo*) or the viral genomic material (*in virio*). This latter point is particularly important in the case of structural viral proteins, such as icosahedral capsid proteins or helical nucleocapsid (NC) proteins that engage in numerous simultaneous contacts with the viral genomic RNA during viral replication, particle assembly, and disassembly. Details of the molecular contacts of these structural proteins with genomic RNA are often lost in crystallographic or electron-microscopy based approaches due to the use of symmetrical averaging during image reconstruction. These techniques may be appropriate to resolve the symmetrical assemblies of the viral capsid proteins, but they can also obscure the inherently asymmetrically arranged genomic RNA within virus particles. Understanding the unique interactions with the numerous capsid proteins is important, as their relative binding positions and affinities with vRNA must be carefully choreographed to successfully guide virus particle assembly (16). Viral

* For correspondence: Yiyang Zhou, yizhou@utmb.edu.

structural proteins can also engage in numerous host RNA contacts. For example, capsid from flaviviruses is not only responsible for wrangling genomic RNA into virus particles but also is shuttled to the nucleus and nucleolus where it mediates diverse, perhaps nonspecific, host–RNA interactions that may be necessary to alter the host transcriptional program to provide a proviral environment (17).

Almost all NGS-based vRNA–protein interaction approaches are derived from or are at least inspired by classical biochemical techniques. In this review, we will discuss how cutting-edge NGS-based technologies evolved from classical methods. We will categorize these NGS approaches based upon whether they employ UV crosslinkers (*e.g.*, HITS-CLIP, PAR-CLIP, vPAR-CL, and *etc.*, Fig. 1), chemical (formaldehyde) crosslinkers (*e.g.*, Protein Interaction Profile Sequencing, PIP-seq), or no crosslinking at all (*e.g.*, native RNA immunoprecipitation and

sequencing, RIP-seq), as well as their methods for RNA–protein complex enrichment (IP or affinity capturing). We will also summarize notable applications in virus research (Table 1) and present how each has made an impact on the way we study numerous viruses including SARS-CoV-2 and HIV-1.

Photo-crosslinking methods

UV light–mediated crosslinking is a well-established tool for studying RNA–RNA-binding protein (RBP) interactions (18) due to its ability to elicit covalent bonds between adjacent (within covalent-bond distance) amino acid side chains and nitrogenous bases of nucleic acids (19). UV-induced crosslinking provides a means of stabilizing otherwise transient and less stable RNA–protein interactions, which usually consist of hydrogen bonds and electrostatic interactions (19). Zero distance (20) UV-induced RNA–protein crosslinking largely

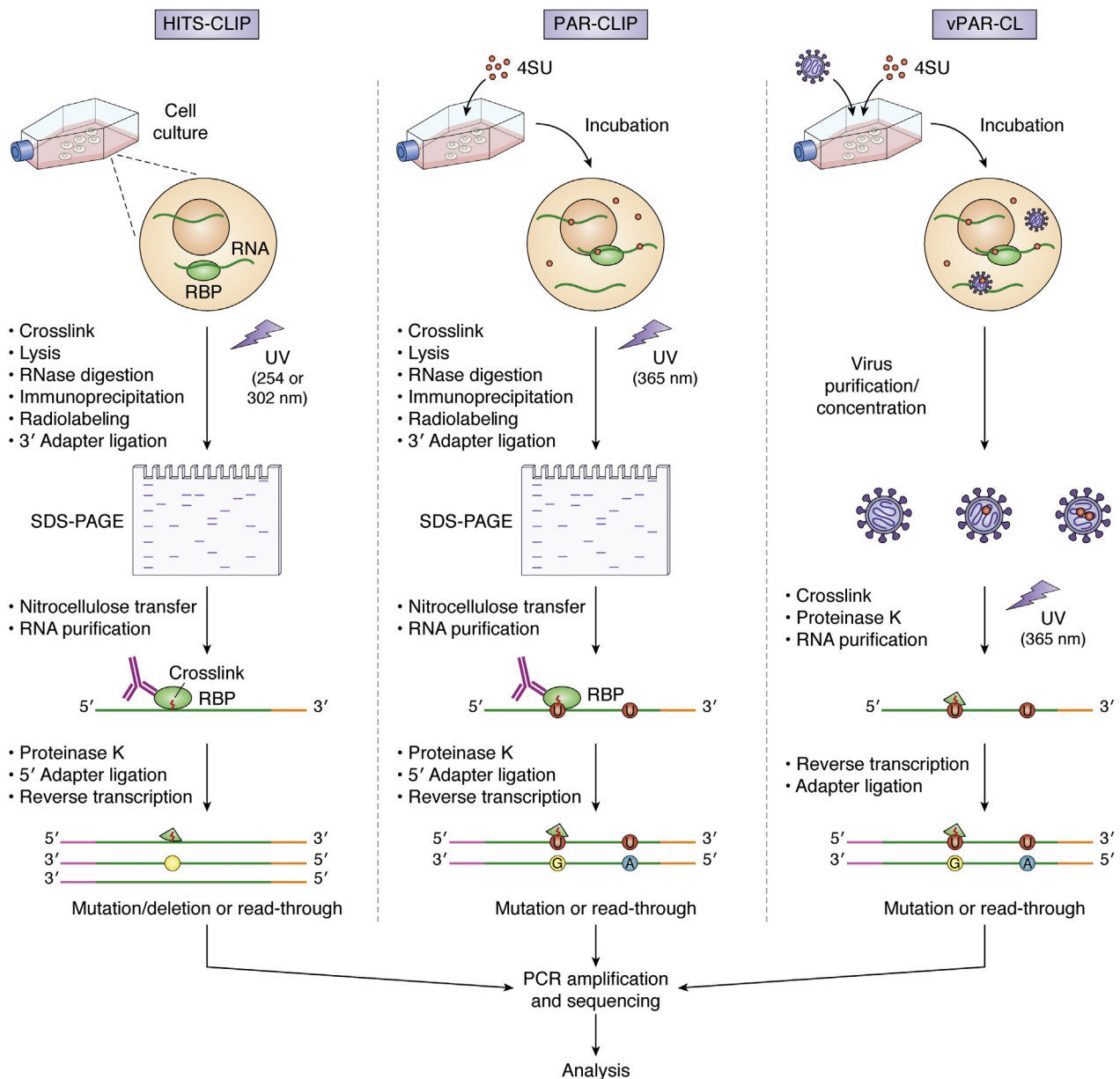


Figure 1. General methodologies of HITS-CLIP, PAR-CLIP, and vPAR-CL.

Table 1
NGS methods for studying RNA–protein interactions

Methods	Abbreviated procedures	Virus	Reference
Photo-crosslinking methods CLIP-seq/HITS-CLIP	UVB/C crosslink, IP, radiolabeling, SDS-PAGE, proteinase K, and NGS	Epstein-Barr virus	(46, 47)
		Kaposi's sarcoma-associated herpesvirus	(42, 69)
PAR-CLIP	Nucleotide analogs, UVA crosslink, IP, radiolabeling, SDS-PAGE, proteinase K, and NGS	Influenza virus	(54, 55)
		Hepatitis C Virus	(52)
		Simian gammaherpesviruses	(50, 51)
		Epstein-Barr virus	(69, 85)
		Kaposi's sarcoma-associated herpesvirus	(69, 85, 86)
		HIV	(89, 90, 102, 106, 111, 112, 114, 121, 122)
		Influenza	(205)
		Moloney leukemia virus 10	(206)
		Herpes Simplex Virus-1	(207, 208)
		Flaviviruses	(209)
Other CLIP-derived methods	iCLIP	Alphaviruses	(210)
		Hepatitis C virus	(211)
		HIV-1 (A3)	(212)
		SARS-CoV-2	(213)
		Flavivirus	(214)
vPAR-CL	Nucleotide analogs, UVA crosslink, proteinase K, NGS	Gammaherpesviruses	(215)
		Flock House Virus	(216)
Chemical crosslinking method PIP-seq	Formaldehyde crosslink, RNase footprinting, NGS	-	-
Affinity capturing methods APEX-seq/Proximity-CLIP	APEX-induced biotinylation, affinity capturing, NGS, and mass spectrometry	SARS-CoV-2	(160)
VIR-CLASP	4SU/photo-crosslinking, solid phase separation, mass spectrometry	CHIKV	(161)
		Influenza A virus	(163)
		Zika	(164)
		SARS-CoV-2	(165)
Non-crosslinking methods RIP-seq	Native IP, RNA extraction, NGS	HIV	(178)
		<i>Bombyx mori</i> nucleopolyhedrovirus	(181)
		EBV	(182–184, 217)
Chemical probing	SHAPE/DMS chemicals	SARS-CoV-2	(180)
		HIV-1	(199)

prevents long-distance, nonspecific crosslinks that are commonly induced by chemical crosslinkers such as formaldehyde (21). UV-induced crosslinks are also in general highly specific to nucleic acid–protein interactions.

In both classical biochemical and high-throughput NGS assays, the final analyte is an RNA fragment crosslinked to a protein of interest. UV crosslinking is typically followed by RNase digestion and enrichment of the RBP of interest using approaches such as IP or other affinity-capturing methods. Often, the RBP is digested by proteinase K to leave only small peptide adducts crosslinked to the RNA. Finally, the RNA–peptide complex is analyzed either directly or by reverse transcription into cDNA. Coupled with NGS, a number of high-throughput platforms such as high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP) (22), PhotoActivatable-Ribonucleoside–Enhanced Crosslinking and ImmunoPrecipitation (PAR-CLIP) (23), and viral PhotoActivatable Ribonucleoside CrossLinking (vPAR-CL) (24) have revolutionized the study of vRNA–protein interactions by providing unprecedented sequencing power and scale.

HITS-CLIP/CLIP-Seq

CLIP (crosslinking and immunoprecipitation) combines short wavelength UV irradiation and IP to identify RNA sequences

interacting with a protein of interest (25, 26). Short wavelength (~254 nm) UV irradiation has been extensively used to induce RNA–protein crosslinks for nearly half a century (27). In spite of this, the detailed photochemistry and biophysical mechanism of such reaction is not completely understood (28). It is generally accepted (18, 29–32) that when an aromatic ring (such as the nitrogen-containing aromatic base of nucleotides) is excited by UV irradiation, nucleobases are induced to a higher energetic state to exceed the ionization potential, which generates cation radicals. The consequence of such short-lived high energy nucleobases is either rapid thermal relaxation or the formation of a covalent bond with similar radicals in direct vicinity (such as UV-excited aromatic rings or other side chains in amino acids) (Fig. 2A). CLIP was first applied to investigate RNA targets that are bound and regulated by autoimmune neurologic disease antigens (Nova proteins, (33)) in mouse brain tissue (25). After UV crosslinking and anti-Nova IP, CLIP sequence-tags (which represent Nova-binding RNAs) were Sanger-sequenced to reveal RNA motifs associated with alternative spliced mRNAs regulated by Nova (34). These Nova-targeting RNA sites were further validated with Nova^{-/-} mice to demonstrate the advantage of CLIP methodology (25).

In the early days of NGS (35), CLIP quickly benefited and was adapted into HITS-CLIP (22). HITS-CLIP follows the same general scheme of CLIP (22) (Fig. 1): (1) RBP of interest is UV

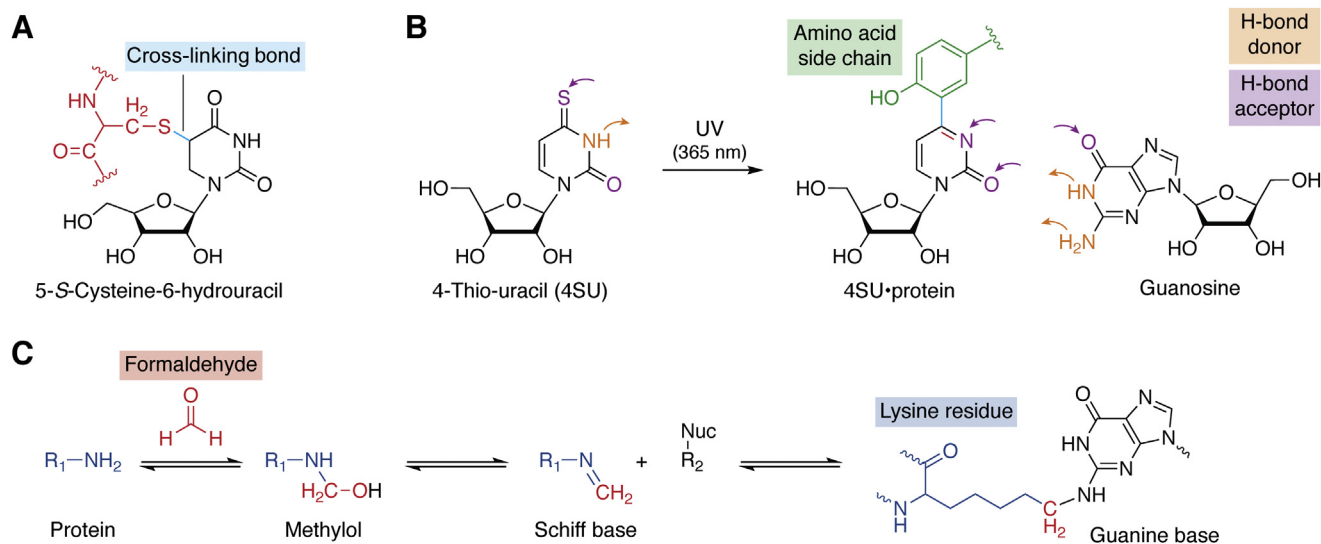


Figure 2. Mechanisms of common cross-linking methods. A, 5-S-cysteine-6-hydrouracil as an example of a cross-link product of 254-nm UV; the cyan line indicates the cross-linking bond. Adapted from the study by Smith and Aplin (218). B, 4-thio-uracil (4SU) cross-linked with amino acid side chains after 365-nm UV irradiation, which alters hydrogen bond donor/acceptor properties of 4SU, and subsequently results in 4SU-guanine mispairing during reverse transcription. Purple and orange arrows indicate hydrogen bond acceptors and donors, respectively. Adapted from the studies by Hafner *et al.* (23) Ascano *et al.* (80). C, molecular mechanism of formaldehyde cross-link: 1) protein; 2) Methylol; 3) Schiff Base; 4) example of lysine–guanine cross-link after formaldehyde cross-link. Adapted from the study by Hoffman *et al.* (21).

crosslinked with bound RNA *via* short wavelength UV irradiation (typically UVC \sim 254 nm or UVB \sim 300 nm) in cell culture; (2) cell lysate is treated with RNase to digest unprotected RNA; (3) IP with specific antibody of the RBP of interest enriches the targeted RNA–protein complex; (4) subsequent SDS-PAGE is used to remove noncrosslinked RNA (26); (5) proteinase K is used to cleave the crosslinked protein next to the carboxyl group of hydrophobic or aromatic amino acids (36); (6) the covalent interaction between polypeptide (or sometimes a single amino acid) and ribonucleotide is retained during RNA purification; (7) finally, sequencing adapters/linkers are ligated to the purified RNA–polypeptide complex and RT-PCR is used to reverse transcribe the RNA and generate dsDNA libraries for high-throughput sequencing.

NGS-coupled HITS-CLIP libraries can readily reveal hundreds of thousands of CLIP sequence-tags (22), which comprise RNA fragments \sim 50 to 200 nts in length, depending on the sequencing platform used (22, 26). This far exceeds the scale of traditional CLIP (usually only a few thousand CLIP tags (25)). The sequences of CLIP tags are thereafter aligned to genome to determine RBP-binding sites and RNA motifs associated with these interactions (22, 28).

Beyond its original application in mouse brain tissue, numerous applications of HITS-CLIP[†] have demonstrated its capacity for RNA-RBP discovery on a transcriptomic scale (reviewed in (28)). In virus research, HITS-CLIP shined in the fashion of AGO HITS-CLIP. Argonaute proteins (AGO) are essential members of the cellular RNA-induced silencing

complex. Guided *via* their interactions with numerous classes of small noncoding RNAs (such as miRNAs), AGO proteins cleave mRNA and repress translation (37). Due to the small size of AGO-bound miRNAs (\sim 20 nts.), it can be a challenge to correctly identify their mRNA partners from their sequence alone as the short miRNAs have many potential ambiguous cognate sequences in the transcriptome and miRNAs often bind to their mRNA targets in spite of multiple mismatches. This can be overcome by AGO HITS-CLIP, which demonstrated that AGO protein is sufficiently well associated with miRNAs and mRNAs to allow UV crosslinking of both AGO-miRNA and AGO-mRNA (38). After IP with specific AGO antibody, the discovered miRNA sequences can be used to seed match the codiscovered mRNA targets. AGO HITS-CLIP quickly empowered virus-encoded miRNA discovery in herpesviruses, which encode numerous miRNAs (39, 40). Kaposi's sarcoma-associated herpesvirus (KSHV, also known as human herpesvirus-8) is the tumorigenic cause of Kaposi's sarcoma (41). KSHV-encoded viral miRNAs are expressed in latently infected cells and are regarded to be associated with viral pathogenesis and tumorigenesis (41). Haecker *et al.* (42) performed AGO HITS-CLIP and recovered thousands of KSHV miRNA targets in KSHV-infected primary effusion lymphoma cells lines, which overlapped with important cellular pathways for KSHV pathogenesis and tumorigenesis. Importantly, AGO HITS-CLIP also recovered significantly different ratios of KSHV miRNAs to human-derived miRNAs, depending on whether the cell lines are postgerminal B cells (BCBL-1, (43)) or pre-B cells (BC-3, (44)). This suggests KSHV infection can potentially compromise the host RNA-induced silencing complex. Similar to KSHV, Epstein–Barr virus (EBV) is another dsDNA herpesvirus (human herpesvirus-4) that persists in humans with latent infection and tumorigenic potential in B cells (45). Riley *et al.* (46) used AGO HITS-CLIP to

[†] Interestingly, HITS-CLIP was renamed “CLIP-Seq” in a study of RbFOX2 splicing regulator targets in human embryonic stem cells (219). Notably, this alternative term triggered certain confusions, as a number of studies also referred such term to a collection of similar technologies that combine CLIP and NGS, such as PAR-CLIP (81). For clarity, in this review, CLIP-Seq refers to the alternative naming of HITS-CLIP, and other NGS techniques utilizing CLIP will be discussed separately.

identify a handful of EBV and human miRNA targets in EBV-transformed B cells (Jijoye cells) and revealed that EBV miRNA predominantly targets human mRNA 3' UTRs (46). Surprisingly, the highly expressed EBV miRNAs were also found to target human mRNAs that are involved in transcription regulation, apoptosis, cell cycle control, and signaling (46). In a follow-up study, Harold *et al.* (47) further investigated EBV miRNA's interaction with Caspase protein 3 protein (CASP3). CASP3 is a central host factor regulating apoptosis (48, 49) and has been speculated to be a target of EBV miRNAs for EBV-associated apoptosis repression (46). Using the same AGO HITS-CLIP and more advanced bioinformatics, Harold *et al.* discovered that EBV miRNAs specifically bind to the 3' UTR of CASP3 mRNA at 13 loci. A subsequent reporter assay confirmed that nine of the discovered EBV miRNAs exhibited significant repression of CASP3, validating the role of EBV miRNAs in targeting CASP3 protein (46). Together, the HITS-CLIP studies on EBV demonstrated the “general-to-specific” approach to understanding viral miRNA functions, in which discovery of protein-binding miRNA targets led to specific hypotheses and experimental validation. Such knowledge can ultimately contribute to defining molecular mechanisms and drug design.

Beside human herpesviruses, HITS-CLIP has been applied in numerous other viruses to study how vRNA and miRNA interact with cellular proteins, such as AGO. With simian gammaherpesviruses (*Herpesvirus saimiri*, HSV), Guo *et al.* used AGO HITS-CLIP to identify the enriched mRNA targets of miR-27 in T-cell receptor signaling pathway (50) and robust AGO-binding sites on both host and HSV genomes that are mediated by viral U-rich miRNAs (51). Luna *et al.* (52) used AGO HITS-CLIP to investigate the relation between hepatitis C virus (HCV) and microRNA-122 (miR-122). miR-122 is a highly expressed liver-specific miRNA, which is speculated to facilitate HCV replication (53). AGO binding of vRNA was observed at multiple regions of viral genome and particularly clustered at the HCV 5' UTR miR-122 sites, confirming AGO engagement of vRNA in a replication-dependent manner (52). In a subsequent study with recombinant virus, pharmacologic inhibition and a single-cell reporter assay further elucidated the ability of HCV RNA replication to derepress host cell miR-122 production that can lead to potential oncogenesis after viral infection (52).

Inside the virus particle, HITS-CLIP has also been demonstrated to be a reliable method to characterize interactions between the RNA genome and structural proteins. This was well illustrated for RNA–NC interactions in influenza A virus (IAV) by Lee *et al.* (54). Infected cells were crosslinked, lysed, nuclease-digested, and IP-ed with nucleocapsid antibodies for different IAV strains. HITS-CLIP sequence data revealed distinctive patterns of binding loci for each strain but that consistently favored G-rich and U-poor RNA regions. In a follow-up study, Le Sage *et al.* (55) extended the HITS-CLIP in-virion binding profiles to other strains of influenza virus and found that both IAV and influenza B virus exhibited “nonuniform and nonrandom” binding patterns between vRNA and NC and that there is uneven distribution of NC-binding sites among different vRNA segments.

In spite of the power of HITS-CLIP, several limitations still apply. While the high energy of short wavelength UV radiation (UVB, UVC) can efficiently stabilize RNA–protein complexes (usually within several seconds), UV radiation may also lead to mutagenesis in the target of interest (56–58). Additionally, unprotected RNA is prone to fragmentation after excessive UV irradiation and oxidation thereafter (59, 60). Although HITS-CLIP is readily applied in cell culture, short wavelength UV is limited in its penetration depth, which typically cannot exceed the second layer of epidermis (stratum lucidum) (61). This prevents HITS-CLIP's application to complex tissues or organs. It is also important to note that during short wavelength UV radiation, protein–RNA crosslinks are not exclusive (61–63) and extensive RNA–RNA crosslinking can also occur (64, 65), which can interfere with the interpretation of true RNA–protein interactions. During downstream analysis, bioinformatic tools have been developed to maximize the specificity and sensitivity of HITS-CLIP data (reviewed in (66, 67)). Nonetheless, HITS-CLIP is relatively coarse in terms of revealing nucleotide-resolution maps of RBP-binding sites (usually at a resolution of 30–60 nts). This is partially due to the lack of complete biophysical knowledge regarding UVB/UVC crosslinking specificity and their amino acid/nucleotide preferences (28). New bioinformatic tools allow for the detection of single nucleotide deletions and transitions in HITS-CLIP data (68). However, it remains complex to discern signals from background due to mutations occurring with different efficiencies for each nucleotide (69). Other technical artifacts such as mis-priming during reverse transcription (in the form of overrepresentation of sequences complementary to the primer) remain to be challenges for HITS-CLIP (70).

PAR-CLIP

PAR-CLIP provides a different path to high-throughput discovery of RNA–protein interactions (23). Based on the same principle as CLIP, PAR-CLIP integrates a significant variation through the use of photoactivatable nucleoside analogs. These enhance UV crosslinking by reducing the required excitation energy and also increasing crosslinking efficiency (23, 71–73). Analogs such as 5-azidouracil, 8-azidoadenine, 8-azidoguanine, 4-thiouracil, 5-bromouracil, 5-iodouracil, and 5-iodocytosine have all been successfully used (reviewed in (71)). Among these, thionucleobases such as 4-thiouracil (4SU) and 6-thioguanosine (6SG) have gained popularity for a number of reasons including the following: (1) sulfur is only 0.45 Å larger than oxygen (71, 74), allowing minimum structure perturbation; (2) 4SU or 6SG can be supplemented in cell culture at high concentrations without obvious cytotoxic effect (23, 72, 73, 75); (3) the sulfur substitution allows for UV excitation around 330 to 365 nm, avoiding the 260 nm excitation wavelength of native uracil and minimizing unwanted photochemistry and/or photodamage (71, 74); (4) free radicals of the thio-group can greatly enhance crosslinking efficiency and crosslinking yield can reach up to 90% (76); and (5) importantly, the 4SU/6SG incorporated RNA can lead to specific base mismatches during reverse transcription (U–G, and G–T) (77–79). This latter point is potentially due to alterations in hydrogen bond donor/acceptor within a

base pair as a result of the crosslinked peptide adduct (80) (Fig. 2B). These crosslink-specific mutations (U-C or G-A transitions) enable high-throughput screening for the precise nucleotide at the RNA–protein crosslink and therefore engaged in RNA–protein interactions.

PAR-CLIP takes advantage of thionucleobase-enhanced crosslinking with the following basic scheme (illustrated in Fig. 1): (1) modified nucleotides such as 4SU are supplemented in cell culture and converted to 4SUTP, which are subsequently incorporated into newly synthesized RNA; (2) 4SU-labeled cells are washed and crosslinked with long wavelength UV irradiation (typically UVA at 365 nm); (3) lysed cells are treated with RNase T1 and IPed on magnetic beads with an antibody specific to the protein of interest; (4) the enriched RNA–protein complex is then treated with RNase T1 again to ensure the removal of uncrosslinked or unprotected RNA, which is followed by radiolabeling of crosslinked RNA; (5) the recovered RNA–RBP from SDS-PAGE EMSA is then digested by proteinase K and the purified RNA–polypeptide complex is ready for reverse transcription and NGS library preparation (23, 81, 82). In the final sequence data, the RBP-interacting 4SU or 6SG is determined by U-to-C or G-to-A transitions, respectively. This mutation profile informs the crosslinking position and ultimately the RNA sites that interact with protein. Compared to HITS-CLIP, which only reveals the approximate binding site position, PAR-CLIP therefore provides a nucleotide-resolution map of RNA–RBP interaction sites.

PAR-CLIP provides a dependable method for identifying RNA/miRNA targets of important cellular proteins, such as Argonaut 2, embryonic lethal abnormal vision (ELAV) protein, pumilio homolog 2 (PUM2), insulin-like growth factor proteins (23, 83), as well as identifying actively transcribed tRNA genes by targeting pre-tRNA binding protein La (lupus antigen, a ubiquitous pre-tRNA–binding protein) (84). In virus research, like HITS-CLIP, PAR-CLIP was quickly adapted to identify the miRNA targets of human herpesviruses including KSHV and EBV. AGO PAR-CLIP uses 4SU to crosslink AGO with interacting miRNAs and mRNAs. Thousands of viral miRNA/mRNA targets were revealed to interact with AGO in latent cell lines infected with either KSHV or EBV (69, 85, 86). Among the discovered miRNA interaction sites, 7% were EBV miRNAs in infected lymphoblastoid cells lines (85), and 30%–27% were derived from KSHV miRNAs in infected BCBL-1 and primary effusion lymphoma cell line BC-1 (86), which is drastically different from the HITS-CLIP study of KSHV-infected BC-3 cells, where ~83% of miRNA reads were of viral origin (42). In comparison to previous studies (42, 46, 47, 69), both PAR-CLIP and HITS-CLIP demonstrated great reliability in determining miRNA targets, such as EBNA2, LMP1, and BHRF1 for EBV and miR-K10a, -K10b, and miR-142-3p for KSHV (69). As for AGO-mRNA targets, the number of genes regulated by viral or host miRNAs were found to be limited in both EBV and KSHV, while instead the majority of mRNA matches (of KSHV-encoded miRNAs) were targets within the host transcriptome (85, 86). This is expected as viral replication cycles are minimal in the selected latent cell lines (69). Among

the AGO mRNA targets characterized by PAR-CLIP, cellular pathways such as transcription regulation, intracellular signaling and transportation, protein localization, MAPKKK (Mitogen Activated Protein kinase kinase kinase), were found to be consistently influenced by both EBV and KSHV miRNAs (86). This suggests a functional similarity between KSHV and EBV miRNAs, despite their evolutionary distance.

PAR-CLIP has also been extensively utilized in HIV research. The HIV-1 Gag polyprotein contains viral structural proteins that coordinate numerous features of the viral life-cycle including genome selection, intracellular genome trafficking, virion assembly, budding, and maturation (reviewed in (87, 88)). This indicates that there are extensive interactions between Gag and numerous host and vRNAs at each of these different stages. Kutluay *et al.* (89, 90) used PAR-CLIP to enrich Gag–RNA complexes from both cells and virions to investigate the global Gag–RNA interactome during and after Gag-orchestrated genome assembly. This uncovered a surprising and drastic shift in profiles of Gag-interacting RNAs during HIV-1 intracellular virion assembly (90). The cytosolic monomeric Gag–RNA complexes occurred at discrete sequences within both the 5′ leader and 3′ Rev Response Element (RRE) of the HIV genomic RNA, suggesting these sites may be spatially adjacent (90). This was contrasted by the membrane-bound, oligomeric Gag–RNAs, which bound a range of sites across viral genome. Such dynamic RNA-binding properties of Gag are also observed during the viral maturation process. Mature virions, which comprise proteolyzed Gag, exhibited a Gag–RNA profile similar to that of the cytosolic fraction, whereas the immature virion exhibited significant similarity to membrane-bound Gag–RNA profile. In addition to vRNA, HIV-1 also packages cellular RNAs (91). Motif enrichment analysis (92) uncovered that cytoplasmic Gag binds to GU-rich cellular RNAs while membrane-bound Gag favors A-rich motifs (90). In the same study, individual domains of Gag were also investigated with PAR-CLIP to characterize RNA-binding specificity. The matrix domain (MA) of Gag was found to be devoid of vRNA but exclusively interacted with host tRNA (90). Interestingly, such MA–tRNA interactions showed strong preference for specific tRNA anticodons. Specific tRNAs such as Glu^{CTC}, Glu^{TTC}, and Gly^{GCC} were bound up to 100-fold more frequently than others (90). In contrast, NC was found to be primarily crosslinked to vRNA and largely resembled the interaction profile of full-length Gag. The detailed dissection of Gag domains, in combination with PAR-CLIP, addressed the question of whether MA can interact with viral or cellular RNAs (93). Together, Gag PAR-CLIP offered a valuable approach to determine the reciprocal and dynamic relationship between HIV-1 Gag and RNA in the context of membrane association during virion genesis.

HIV-1 integrase (IN) is a multi-domain enzyme and one of the cleaved products of Pol polyprotein (94). IN mediates the integration of viral DNA into host chromosomes following the production of double-stranded proviral DNA from reverse transcription (95, 96). In addition to integration, IN has long been suggested to coordinate viral replication and virion maturation. Mutated IN can lead to the eccentric “exile” of

ribonucleoprotein complexes (RNPs) outside of capsid shell and ultimately impairs virion maturation (97–101). Using PAR-CLIP, Kessl *et al.* identified IN-binding RNA targets in virions (102). IN showed strong binding preference for the trans-activation response (TAR) element (103) and RRE, but not for the packaging element ψ , suggesting IN and NC have both shared and unique roles in HIV-1 genome assembly and particle maturation (102). *In vitro* biochemical experiments and mass spectrometry–based protein footprinting (104) were conducted to validate IN-RRE/IN-TAR binding and to evaluate their affinities with deletions/changes of structural elements. The quinoline-based allosteric HIV-1 integrase inhibitors are a class of anti-HIV agents that bind noncatalytic sites of IN, preventing IN–vRNA interactions (105). In one study, Madison *et al.* (106) used allosteric HIV-1 integrase inhibitors (107) to promote eccentric maturation of HIV-1 and applied PAR-CLIP (alternatively named “CLIP-seq” in this study) to reveal conservation of NC–vRNA interaction sites, despite vRNA being mislocated outside the capsid shell. Together, PAR-CLIP unveiled the correlation of IN–vRNA interactions and particle maturation.

In addition to important HIV-1 viral proteins, PAR-CLIP has been used to investigate several HIV-1-related cellular factors, antiviral proteins, and RNA modifications (reviewed in (108)). For example, APOBEC3 (A3) is a family of cytidine deaminases that can suppress a broad range of viruses, including HIV-1 (109, 110). PAR-CLIP has been used to uncover the preferential binding of A3 proteins to cellular and virion vRNAs, suggesting binding specificity is influenced by nucleotide composition (G-rich and/or A-rich) instead of sequence (111). In a following study, PAR-CLIP revealed the virion-encapsidated human A3H haplotype II (huA3H) protein favors interaction with short RNA duplexes (7 nucleotides) (112). Another host protein related to HIV-1 pathogenesis is the Zinc finger antiviral protein (ZAP). ZAP represents another class of antiviral host factors targeting a broad range of viruses by promoting the degradation of viral mRNAs (113). PAR-CLIP provided evidence that ZAP binds highly specifically to CpG dinucleotides (114). This allows for ZAP to differentially target viral but not host mRNA, as the latter is typically depleted of frequent CpG dinucleotides (115). It is also interesting to note that some have speculated that HIV has similarly begun evolving to evade this mechanism by altering dinucleotide content (116, 117).

In host cells, DNA methylation of provirus promoter regions has been shown to regulate HIV-1 latency and transcription activation (118). In addition, RNA methylation has recently been recognized as an important factor in HIV-1 RNA metabolism and replication (119). Methylation of adenosine at the N6 position (m6A) is the most prevalent mRNA modification, which are ‘read’ by the cytoplasmic YTH domain family (YTHDF) proteins (120). One study used PAR-CLIP to enrich YTHDF–vRNA complexes and unveiled that m6A modifications were clustered at HIV-1 3′ UTR exclusively (121). Interestingly, another study applied HITS-CLIP instead of PAR-CLIP (122) and determined that binding sites of YTHDF1–3 proteins are located in both 5′ and 3′ UTRs. This

discrepancy provides an interesting comparison of the two methods and suggests that HITS-CLIP allows for a more permissive and therefore sensitive detection of protein-binding sites, while PAR-CLIP provides a more stringent but specific readout.

PAR-CLIP applications in virus studies are far beyond the above-mentioned human herpesviruses and HIV. Influenza virus, flavivirus, and alphavirus research have all benefited from PAR-CLIP and its extensions. We listed more selected applications in Table 1.

Similar to HITS-CLIP, PAR-CLIP also has several intrinsic limitations. The introduction of a nucleoside analog to nascent RNA transcripts or vRNAs limits PAR-CLIP applications within the scope of cell culture and *in vitro* systems instead of *in vivo* research. Photoactivatable nucleoside-induced cytotoxicity (123), as well as other cellular stress and inhibitory effects (124), should also be critically evaluated for each cell line or experimental model. The concentration and uptake efficiency of analogs also require optimization. Typically, only one analog is applied in PAR-CLIP experiment. This dictates that the observable interacting sites are limited to the crosslink events at that particular nucleotide (*e.g.*, only U crosslinks can be revealed with 4SU supplement in cultured cells). Although it is possible to supplement with a different analog in parallel experiment (*e.g.*, 6SG), they each exhibit different crosslink efficiencies and signal intensities (23). Similarly, thio/UVA–induced crosslinks only favor reactive amino acid side chains (mainly phenylalanine, tyrosine, and tryptophan, while lysine and cysteine can also be crosslinked to a lesser extent) (77, 79), hence certain RNA–protein interaction lacking these favorable amino acids may be missed. The long and technical experimental procedures of PAR-CLIP (Fig. 1) also demand a large quantity of starting material (usually starting between 10^8 and 10^9 cells) (81). Additional concerns such as antibody specificity, IP efficiency, as well as whether a crosslink can interfere with antibody binding, should also be considered for PAR-CLIP, as with HITS-CLIP and other IP-based techniques.

vPAR-CL

vPAR-CL represents a technique that is different from other CLIP-based methods. vPAR-CL is specifically designed to investigate the *in virio* RNA–protein interactions of an RNA virus (24). In contrast to the complex cellular micro-environment, vPAR-CL takes advantages of the unique and highly confined enclosure of an assembled virus, which only consists of few well-defined components (*e.g.*, vRNA and capsid). In return, vPAR-CL obviates the time- and resource-consuming IP process but instead relies on the standard purification of the given virus (24). vPAR-CL starts in an analogous fashion to PAR-CLIP (Fig. 1), whereas a photoactivatable nucleoside analog (*e.g.*, 4SU) is supplemented to cell culture during virus infection to label newly synthesized vRNA genomes. After cell lysis, the virus particles are purified (typically with PEG precipitation and/or sucrose gradient ultracentrifugation) and RNase treated to remove any copurified unpackaged viral or host RNAs. The intact, 4SU-containing viruses are crosslinked with 365 nm UV and then are subjected

to proteinase K digestion, RNA extraction, reverse transcription, and NGS library construction. In comparison to PAR-CLIP (23, 81, 82), vPAR-CL avoids the need for labor-intensive procedures such as IP of RBP, dephosphorylation, radio-labeling, SDS-PAGE and electro-elution, and therefore significantly shortens the time frame of experiments and the requirement of starting materials. When coupled with the fragmentation-free NGS technology such as ClickSeq (125, 126), vPAR-CL can yield results with less than 2 μg of viral particles (equivalent to 250 ng of vRNA in the case of Flock House virus (FHV)) (24), while typical PAR-CLIP starts with a large number of cells (*e.g.*, 10^8 cells or approximately 10–50 15 cm cell culture plates for HEK293 cells) (81, 82). Removing the IP in vPAR-CL mitigates the bias generated due to specificity of antibodies and efficiency of IP and the associated procedures. This also prevents any potential interference of IP that is caused by RNA–protein crosslinking. Without any artificial enrichment, vPAR-CL largely minimizes the background noise, as any random, nonspecific crosslinked signals and intrinsic viral mutations will be diluted, while specific and consistent RNA–capsid interactions can readily be discerned (24). An important and novel feature of vPAR-CL is derived from the fact that all purified vRNA is sequenced (rather than just the IP-enriched crosslinked fragments) from both cross-linked RNA (4SU+/UV+) and noncrosslinked control (4SU+/UV-). As a result, vPAR-CL directly compares the U-to-C transition rates between conditions. This eliminates intrinsic mutational events (127) from interaction site interpretation (24) (Fig. 3) but also provides a ratiometric value for RNA–crosslinking, allowing for both the identification of RNA regions that are bound to RBPs in a conserved and structured manner as well as providing information on which sites do not interact with RBPs.

vPAR-CL has been demonstrated with FHV to reveal highly conserved RNA–capsid interactions across the entire encapsidated viral genome. This suggests a structural tropism of the RNA inside virion (24) that likely reflects the dodecahedral RNA cage previously observed for FHV RNA in x-ray and cryo-EM studies (128). The distribution of RNA–capsid interaction clusters also suggests that the packaged FHV genome has multiple conserved RNA–capsid–binding sites. This indicates that the genome packaging mechanism of FHV could resemble that of other +ssRNA viruses (16). Combined

with dimethyl sulfate mutational profiling with sequencing (129), it was also uncovered that FHV RNA–capsid interactions are enriched in dsRNA regions, and the disruption of base pairing at these vPAR-CL sites interfered with viral fitness (24).

vPAR-CL represents a rapid and labor-friendly method to specifically study *in virio* RNA–protein interaction sites. Both vPAR-CL and PAR-CLIP utilize photo-activatable nucleoside analogs to achieve efficient crosslinking and investigate nucleotide-resolution interactions sites. Therefore, vPAR-CL shares the same biochemical limitations as PAR-CLIP related to thio-based photo-crosslinking. Additionally, due to the lack of enrichment of crosslinked 4SU-containing RNA–protein complexes, the incorporation rate of 4SU is critical in vPAR-CL experiments, which determines whether the vPAR-CL signal can stand out above the background. This requires optimization on 4SU dosage and virus harvest timing. vPAR-CL is demonstrated to be reliable to resolve the protein–interaction sites in icosahedral particles such as FHV, whose RNA genome is in direct contact with single protein component—the capsid. It remains an open question whether vPAR-CL can be adapted to other enveloped viruses such as flavivirus, in which the vRNA genome may be in contact with multiple protein components (17).

Other CLIP-derived methods

There are many techniques derived from the same rationale as CLIP but differ in downstream procedures and applications to identify specific protein–RNA interactions in cells, viruses, and *in vitro* (reviewed in (130–132)). Nonetheless, the principles behind these protocols remain largely the same, as do many of the steps necessary to conduct these experiments and analyze the resulting sequencing data. Here, we will only provide a brief overview to compare and contrast a handful of these CLIP-derived methods based on their hallmarks.

In individual nucleotide resolution CLIP (iCLIP) (133), crosslinked RNA–RBP sites are identified *via* cDNA chain termination at the introduced covalent bonds during reverse transcription, which is expected to occur in more than 80% cases (133). This contrasts HITS-CLIP which typically characterizes the approximate crosslinking sites after cDNA read-through. To capture these truncated cDNA, the reverse transcription primer contains both the 3' and 5' adapters for sequencing. The transcribed cDNA is then intramolecularly

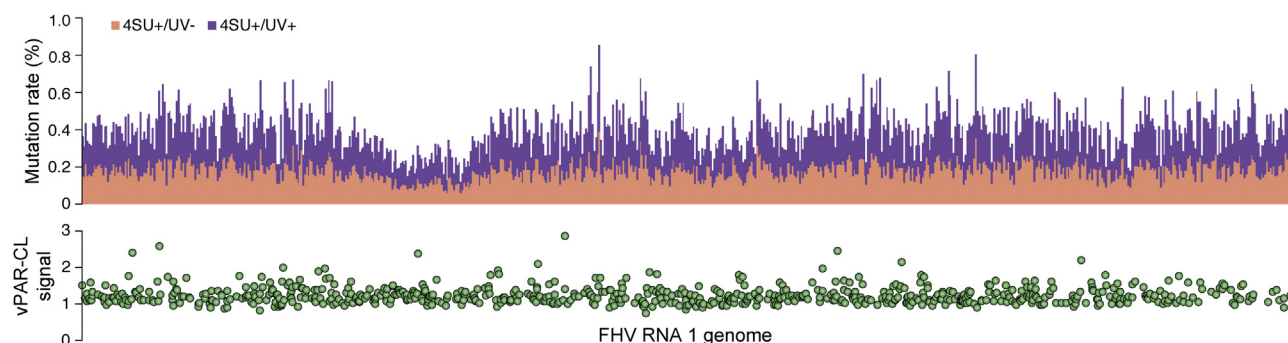


Figure 3. Example of vPAR-CL signals across full-length RNA1 genome of FHV. The U–C transition rates (*upper*) were compared between cross-linked virus (4SU+/UV+) and non-cross-linked control virus (4SU+/UV-) to yield vPAR-CL signals (*lower*), which represents the fold change of U–C transitions. Adapted from the study by Zhou and Routh (24).

circularized. This, importantly, allows for nucleotide resolution mapping of the interaction site, which is located one nucleotide upstream of the truncation site.

Infrared-CLIP (134) utilizes antibody-conjugated beads to IP RNA–RBP complex followed by on-bead RNase digestion to maximize retained RNA fragments. This is followed by the ligation of an IR800-biotin adapter to the 3' end of IP-ed RNAs to avoid the standard radioisotope labeling (of HITS-CLIP and PAR-CLIP) at the 5' ends of RNA molecules. The infrared-biotin adapter not only prevents the inconsistent autoradiography signals due to radioisotope decay but also largely reduces the time required for protein–RNA complex visualization with equivalent efficiency.

In simplified CLIP (135), after IP, crosslinked RNA is biotin labeled and subsequently visualized *via* streptavidin-horseradish peroxidase to avoid radiolabeling. This is followed by proteinase K digestion and polyadenylation of RNA to allow reverse transcription with a modified oligo-d(T) primer. This obviates the ineffective RNA ligation of sequencing adapters, allows for low input materials during reverse transcription, and omits size-selection of cDNA products. Therefore, as the name suggests, simplified CLIP presents a simplified and efficient procedure compared to traditional HITS-CLIP.

Enhanced CLIP (136) builds upon the iCLIP protocol but instead incorporates two separate adapter ligation steps: (1) after IP, the enriched RNA is dephosphorylated and the first ssRNA adapter is ligated to the 3' end of crosslinked RNA. This first adapter contains an “in-line-barcode” to enable the pooling of similar molecular weighted samples from multiple experiments; (2) The second ssDNA adapter, which contains a random nucleotide sequence (a random-mer), is ligated to the 3' end of cDNA after RT. This ssDNA adapter serves to preserve the single-nucleotide resolution of terminated cDNA reads. The random nucleotide sequence of the second adapter (also commonly referred as unique molecular identifier) allows for demultiplexing and to determine whether identical sequenced reads represent two unique RNA fragments (bearing different random-mer sequences) or the PCR duplicates of the same RNA fragments (bearing the same random-mer sequences). Together, enhanced CLIP provides an approach to greatly shorten the hands-on time of typical CLIP-based techniques and to reduce nonspecific RNA–protein interaction artifacts and PCR bias during data interpretation.

In terms of virus research, we list examples of these CLIP-derived methods that have been used to study protein–RNA interactions in response to viral infections (Table 1).

Chemical crosslinking

Although photo- (UV-) crosslinking methods offer reliable solutions to characterize direct RNA-protein contacts with nucleotide or near-nucleotide resolution, methods such as HITS-CLIP, PAR-CLIP, vPAR-CL, and other similar technologies are only effective for *in vitro*, in virion, and cell culture studies. When it comes to a whole tissue or whole organism

scale, UV-light often falls short in providing a viable crosslinking solution, due to its deficiency in evenly penetrating dense and complex tissues. Instead, crosslinking methods using chemicals such as formaldehyde are widely used due to their ability to permeate different tissues.

Formaldehyde elicits crosslinking of a broad range of biomolecules. Between protein and nucleotides, formaldehyde first reacts with protein nucleophiles to yield an imine (Schiff Base), whose subsequent interaction with amino groups of DNA/RNA bases results in covalent crosslinking (Fig. 2C) (reviewed in (21)). In comparison to “zero-distance” photocrosslinking, the additional carbon of formaldehyde extends the crosslinkable range to ~ 2 to 3 \AA (21) (2 N-C bonds between nucleic acid and amino acid) but still allows for crosslinking of macromolecules within close proximity (137, 138). It has been suggested that the numerous formaldehyde aggregates in commercial products can further extend the distance-spanning capability to much greater distances (21, 139). Therefore, formaldehyde is able to crosslink and capture certain RNA–protein interactions that are otherwise excluded from UV crosslinking (140, 141). An important feature of formaldehyde crosslinking is its reversibility. The heat and salt conditions allowing for formaldehyde crosslink reversal have been well characterized (138, 142, 143) and are critical in ChIP/ChIP-seq experiments (144).

PIP-seq (141, 145, 146) was therefore developed, which combines crosslinking, RNase-mediated protein footprinting, and high-throughput nuclease-sensitivity sequencing assays (147, 148). Although PIP-seq can also utilize UV to crosslink RNA–RBP, formaldehyde remains the most common reagent in this application. In PIP-seq, formaldehyde-crosslinked cells are lysed and total cellular RNA–protein complexes are separated into two pools. Pool 1 (experimental group) undergoes RNase digestion and subsequent heat reversal to enrich for protein-bound RNA footprints (crosslinked and hence protected from RNase digestion). Pool 2 (control group) is first subjected to proteinase K treatment and subsequent RNase digestion. This is followed by heat reversal of formaldehyde crosslinks. The recovered RNAs from both pools are used as input for NGS library synthesis. Comparison of the read coverage across the whole host transcriptome is drawn to identify protein-protected RNA sites, which is represented by a differential signal between pool 1 and pool 2 (141, 145, 146). The use of ssRNA- or dsRNA-specific RNases allows identification of ssRNA or dsRNA sites that are associated with proteins (145). PIP-seq has been applied in human cell (HeLa) (141) and plant cell (145, 149, 150) transcriptomes and identified numerous novel RPB-binding motifs, including enriched RBP-reacting polymorphisms that are associated with diseases (141).

PIP-seq is not accompanied with any conventional IP methods to allow specific enrichment of RBPs of interest. Unlike vPAR-CL which is only applied in virions with highly selected RNA/protein components, PIP-seq instead searches the entire transcriptomic length in a complex cellular or *in vivo* setting. This reveals the global state of protein-bound RNA sites, without informing on the counterpart (RBP). To

address this, ‘RNA:protein immunoprecipitation in tandem (RIPiT)-Seq’ can be performed (151). RIPiT-Seq resembles a chimeric form of PIP-seq and RIP (Native RNA IP)-seq, which combines formaldehyde crosslinking, RNase footprinting, and IP to allow discovery of RNA-binding sites of RBP of interest. Of note, in RIPiT-Seq, the RBP of interest is typically additionally labeled (*e.g.*, FLAG-tag). After formaldehyde crosslinking, two sequential IP steps (with anti-FLAG and anti-RBP, respectively) are involved and separated by RNase treatment to generate RNP footprints. This is followed by crosslink reversal, RNA extraction, dephosphorylation, size selection, and NGS library construction (151). As a major advantage of RIPiT-Seq, the two sequential IPs (which is enabled by additional FLAG tag) significantly deplete intracellular RNA species (*e.g.*, rRNA fragments, tRNAs), which therefore improves signal-to-noise ratio in the downstream NGS assays.

Although formaldehyde is commonly regarded as a more efficient and powerful crosslinker to UV irradiation, formaldehyde readily induces protein-protein crosslinking, which can interfere with the specificity of identification of desired RNA–protein interaction as well as yielding indirect signal (149). It is also important to recognize the formaldehyde aggregates in commercial formalin solution (21, 139) can potentially introduce significant amount of long-range, nonspecific multi-component crosslinks in biological samples. The required sample-by-sample optimization of crosslinking conditions (to avoid over-crosslink) and the RNA thermo-damage during crosslink reversal also remain as challenges of PIP-seq method and other formaldehyde-based approaches.

Affinity-capture methods and proteomics

Besides IP, affinity-capture (152) provides an alternative solution to specifically purifying protein contents of interest as well as to enrich crosslinked RNA sequences. There are numerous adaptations of HITS-CLIP or PAR-CLIP, in which IP was replaced with affinity-capture methods. One innovative application that combines affinity-capture and NGS is proximity-CLIP, which utilizes the unique abilities of engineered ascorbic acid peroxidase protein 2 (APEX2) to biotinylation proximal endogenous proteins with biotin-phenol moieties (153–155). In proximity-CLIP (156), APEX2 is fused with a localization element to specifically target a subcellular compartment, and cells are infused with 4SU. APEX2 is activated by biotin-phenol and subsequent hydrogen peroxide addition to covalently tag proteins of proximity with biotin. This is followed by 4SU-induced UVA crosslinking (as described for PAR-CLIP) to capture nascent transcripts interacting with biotinylated proteins. Compartment-specific RNPs and proteins are then captured by streptavidin affinity chromatography. Quite uniquely, proximity-CLIP allows for concurrent compartment-specific analysis of both proteome (*via* mass spectrometry) and interacting RNAs (*via* RNA-seq) (156). A similar technology, APEX-seq (157–159), is based on the same rationale and provides yet another solution to depict both the proteome and its interacting RNA profiles with specificity to a certain cellular compartment. In contrast to the

4SU-induced crosslinking of proximity-CLIP, APEX-seq typically crosslinks RNA to proteins *via* formaldehyde (159) or even bypasses crosslinking altogether (157, 158). In virus research, APEX-seq data was referenced in a study investigating localizing signals in SARS-CoV-2 RNA, predicting that both genomic and subgenomic vRNAs are localized to the host mitochondria and nucleolus (160).

As a recent development, VIR-CLASP (VIRal Crosslinking And Solid-phase Purification) also presents a method that is designed to investigate the protein components of the vRNA-cellular RBP interactions (161). In VIR-CLASP, in a similar fashion to vPAR-CL (24), cells are 4SU labeled and infected with virus of choice, such as chikungunya virus (161). This is followed by UVA crosslinking and cell lysis. The denatured RNA–protein complexes are then recovered with solid-phase purification with solid phase reversible immobilization (162) beads to selectively enrich for nucleic acids. RNA is digested with nucleases and the crosslinked protein components are identified by LC–MS/MS. In addition to chikungunya virus, VIR-CLASP also successfully identified host proteins interacting with viral genomes of IAV (163), Zika virus (164), and SARS-CoV-2 (165). Notably, similar techniques have been applied previously to study the mRNA-binding proteins in mammalian cell lines (166), with the main difference being the use of oligo(dT) beads to harvest crosslinked RNA–RBP complexes. Interestingly, in this study, the use of 4SU-induced crosslinking and mass spectrometry for protein identification is informally termed “PAR-CL”, which is not to be confused with vPAR-CL.

There are also numerous RNA-centric methods that utilize affinity-capture to investigate RNA–protein interactions. For example, RNA Bind-n-Seq (167, 168) is based on the same principle of RIP-seq where no crosslink was induced. However, in RNA Bind-n-Seq, streptavidin-binding peptide tag was used to purify the targeted protein with streptavidin magnetic beads. Similarly, crosslinking and affinity purification (iCLAP) (169) purifies double-tagged (Strep/His) RBPs using streptavidin beads.

Non-crosslinking methods

RIP-seq

Native RNA immunoprecipitation (RIP) was first applied to isolate and purify proteins interacting with the XIST RNA that controls X chromosome inactivation (170, 171). RIP quickly partnered with NGS to form RIP-seq to allow high-throughput screening of Polycomb repressive complex 2 (PRC2) interacting RNAs in embryonic stem cells (172). The rationale of RIP-seq is largely similar to HITS-CLIP, with the main difference of the lack of crosslinking (171, 172). Instead, RIP-seq relies on the native high affinity between certain RNAs and RBPs to withstand the subsequent IP and purification. In RIP-seq, cell or nuclear lysates are directly subjected to specific antibody binding and bead pull-down. Strongly bound RNAs are retained after extraction and NGS libraries and bioinformatics assays are conducted thereafter (172, 173). RIP-seq is demonstrated to be a versatile method of transcriptomic

miRNA/mRNA-RBP profiling in the fields of stem cell research, RNA epigenetics, and alternative splicing (172, 174–176). Alternatively to NGS, RIP has also been combined with microarray analysis to yield another high-throughput application: RIP-Chip (177).

In virus research, Lichinchi *et al.* combined m⁶A IP and RIP-seq (MeRIP-seq) to study the HIV-1 induced m⁶A increase in both host and viral mRNAs and uncovered the connection between HIV-1 RRE methylation and nuclear export efficiency of RNA (178). Similarly, a recent study combined m⁶A-RIP-seq and m⁶A-PAR-CLIP (miCLIP) (179) to reveal eight putative m⁶A sites in the SARS-CoV-2 genome (180). Importantly, single-nucleotide variants associated with these identified m⁶A sites allowed the phylogenetic clustering of specific US epidemic strains of SARS-CoV-2. In silkworms infected with *Bombyx mori* nucleopolyhedrovirus (BmNPV), Nie *et al.* conducted AGO2 RIP-seq and identified numerous small noncoding RNAs, including highly enriched rRNA-derived fragments (181). With EBV, AGO2-RIP-seq depicted the viral and cellular miRNA landscape in diffuse large B-cell lymphoma cell lines (182) and lymphoblastoid cell lines (183). Another important EBV protein is Epstein-Barr nuclear antigen 1 (EBNA1), which plays a vital role in viral replication and the partitioning of viral genomic DNA during latent viral infection but it has also been shown to alter splicing of host mRNAs (184). EBNA1-RIP-seq clearly showed that EBNA1 bound specific cellular targets such certain mRNAs and non-coding RNA (184). Interestingly, the discovered EBNA1-RIP-seq targets did not include any gene whose splicing was modulated by EBNA1, suggesting that EBNA1's ability to modulate splicing does not require direct interaction between EBNA1 and target RNAs (184).

In RIP-seq, the obviation of crosslinking largely shortens the experimental procedure and reduces the requirement of starting material. However, the exclusion of crosslinking also limits the application of RIP-seq. The requirement for strong interactions within an RNA–RBP complex that can withstand purification procedures is unguaranteed and the IP of non-crosslinked RNA inevitably precludes stringent washing during purification that would remove contaminants. Similarly, RNase digestion must be controlled carefully in RIP-seq. RIP-based techniques are typically suitable for stable RNP, such as those interacting with noncoding RNAs (185). RIP therefore provides a suboptimal solution for the study of transient or less stable RNA–protein interactions. Furthermore, false positive signals can arise from the reassociation of RNA (particularly mRNA) and proteins in the lysate (185–188).

SHAPE/DMS flexibility probing

Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) and dimethyl sulfate (DMS) chemistries have been extensively used as RNA secondary structural probing methods, due to their extraordinary abilities to methylate flexible nucleotides in biological samples (SHAPE methylates 2'-OH of RNA ribose while DMS methylates N1 of adenine and N3 of cytosine) (189–193). More recently, SHAPE or DMS

probing has been empowered by high-throughput NGS and advanced bioinformatics to generate mutational profiles (MaPs) to offer incredible depth and single nucleotide resolution to transcriptomic scale RNA structure discovery (129, 194–198). As great as their power in ssRNA probing, it is often overlooked that SHAPE and DMS are, at their core, nucleotide flexibility probes. Both structural (*e.g.*, base pairing) and functional restrictions (*e.g.*, RNA–protein interaction, DNA nucleosome formation) can preclude methylation and allow unveiling of the structural/functional relations. Indeed, Smola *et al.* (197) used SHAPE-MaP to investigate intracellular RNA–protein interactions: SHAPE chemical 1M7 was applied under native conditions to living cells as well as extracted and refolded RNAs. The differences in SHAPE reactivity can hence be compared as the *ex vivo* RNA samples are deproteinized and lack numerous RNA–protein interactions compared to *in cellulo*. Similar to PIP-seq (141), SHAPE-MaP and related chemical probing methods provide a high-throughput transcriptomic scan of RNA–RBP interaction sites, with little prerequisite for existing knowledge of involved proteins or RNAs. In addition, Kenyon *et al.* also combined SHAPE-MaP with photo-crosslinking to fully take advantage of both worlds (199). XL-SHAPE (crosslink and SHAPE) utilizes SHAPE reagent N-methylisatoic anhydride in conjunction with 254 nm UV irradiation to concurrently capture RNA structural changes as well as protein-binding sites. With *in vitro* crosslinking, XL-SHAPE successfully identified the RNA-interacting site of HIV-1 Tat/TAR (200) complex and Gag-binding site on HIV-1 leader sequence. This was further complimented by differential SHAPE reactivities of bound/unbound RNAs (199). Similar “function-to-structure” investigation has also been conducted on FHV, in which vPAR-CL and dimethyl sulfate mutational profiling with sequencing were used to identify clustered RNA–capsid interactions on dsRNA (24).

Perspective

Virus research has been fueled by rapidly emerging high-throughput and quantitative tools to study RNA–protein interactions. Since the development of HITS-CLIP, RNA research as well as the RNA virus field has evolved to transcriptomic and genomic scale studies. It is apparent that the NGS-enabled RNA–protein interaction technologies are an ever-growing field and that more and more such techniques will continue to be applied in an expanding range of viral and biological systems. It is inevitable that many of these technologies have been adapted or developed to study pressing issues such as the ongoing SARS-CoV-2 pandemic.

With such a great availability of NGS-enhanced technologies to investigate RNA–protein interactions, it could be overwhelming to determine the most appropriate scenarios to apply each technique. We hereby provide a quick reference flowchart (Fig. 4), in which we categorized the above-mentioned approaches based on their discovery goals (RNA-centric or protein-centric discoveries), their starting materials, and procedural requirement (*e.g.*, whether antibody is available

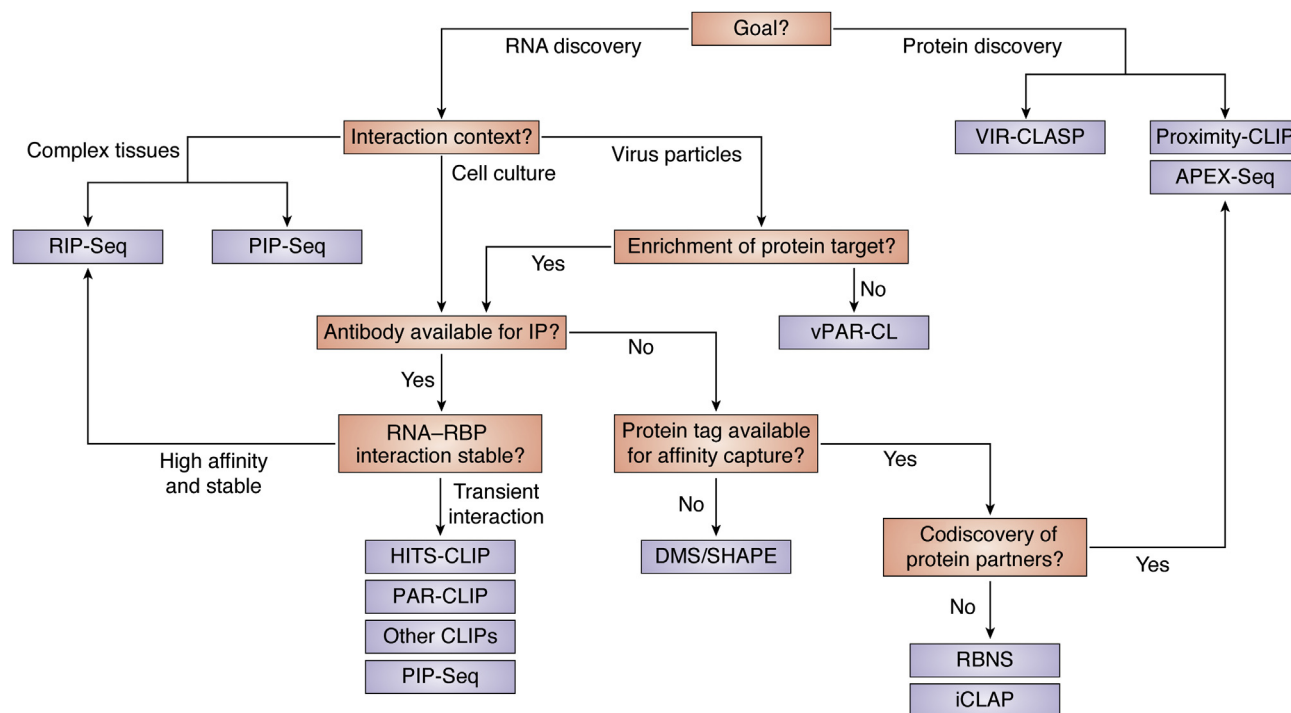


Figure 4. A quick reference flowchart based on different methods' discovery goals, starting materials, and procedural requirements.

for IP). This flowchart is intended for brief guidance, while many other factors should be considered during experimental design.

Certainly, many challenges remain. Although novel methods such as vPAR-CL reduce experimental burden, many techniques are still labor-intensive and require handling of hazardous materials such as radioisotopes. Both technological and computational advances are necessary to improve RNA-binding site discovery to have a higher sensitivity, lower bias, and better specificity. It is also challenging to adapt these methods to applications in more complex scenarios such as in animal models or higher plants. Finally, as many techniques only focus on end-point analysis, it is important but remains difficult to understand how the RNA–protein interactions and the associated regulatory networks dynamically change at different timescales. This is especially important to virus research as vRNA genomes encounter a myriad of protein partners throughout their replication cycles.

It is also tempting to cross-compare different platforms for their reliability, sensitivity, and the biological relevance of the discovered sites. However, it remains challenging to draw such conclusions, as different experimentalists will inevitably employ different cell lines, culturing conditions, virus strains, infection methods to address their own research need. Numerous crosslinking methods, enzymes (83), sequencing adapters (201), and bioinformatic pipelines (67, 202) can all dictate the final output of experiments. To the best of our knowledge, there is no systemic study available thus far to compare different approaches in detail. Nevertheless, a few studies managed to provide a glimpse at how certain methods overlapped or contrasted with others, in their own specific

experimental scenarios. In one forementioned example, two studies tackled HIV-1 mRNA methylation sites with PAR-CLIP and HITS-CLIP (122). PAR-CLIP revealed (121) the exclusively enriched m6A sites at 3' UTR whereas HITS-CLIP identified sites at both 5' and 3' UTRs (122). The exact nucleotide sequences of identified sites in these two studies are also different. In another study (83), both HITS-CLIP and PAR-CLIP were used with the same bioinformatic framework to investigate RNA-binding sites of HuR protein (203). It is shown that each method delivered reproducible results with similar correlation coefficient between replicates (83). However, there was no direct cross-comparison between HITS-CLIP and PAR-CLIP in regard to the proportion of overlapped discoveries. These limited examples suggest divergencies of the different approaches: each highlights its own advantages and limitations; each has its own focus (e.g., HITS-CLIP focuses on a wide range of binding sites, while PAR-CLIP focuses on higher specificity of the binding sites); each will discover different sets of RBP-binding sites on a given transcriptome or genome. It also demonstrates that the NGS and bioinformatic discoveries resulting from any of these approaches should be considered as a highly efficient hypothesis-generating approach, which ultimately necessitate experimental validation.

Looking ahead, several rapidly growing new technologies also help depict a bright picture of virus RNA–protein interaction research. For example, single-cell RNA sequencing allows for comprehensive understanding of an individual cell in the context of its original microenvironment. Additionally, Oxford Nanopore Technologies' long-read nanopore sequencing is advantageous in generating long and continuous

single-molecule sequencing reads, which can effectively cover the entire virus genome to understand the coevolution of distant virus mutations and other features (204). Furthermore, nanopore sequencing can directly sequence RNA in its native state, allowing for the detection and resolution of modified RNA bases which may be extended to detect peptide adducts generated due to RNA-protein crosslinking. Altogether, it is foreseeable that both single-cell RNA sequencing and Nanopore sequencing, among other new technologies, will further lead the vRNA-protein interaction research to new and exciting avenues in the near future.

Author contributions—Y. Z. writing—original draft; Y. Z., S. L. S., and A. L. R. writing—review and editing; S. L. S. visualization; A. L. R. supervision.

Funding and additional information—This work was funded by the National Institute of Allergy and Infectious Diseases, NIH, U54AI150472 subcontract to A. L. R., National Institute of Allergy and Infectious Diseases, NIH, R21AI151725 to A. L. R., National Institute of Allergy and Infectious Diseases, NIH, P01AI150585 to A. L. R. Y. Z. is supported by the Institute for Human Infections & Immunity, Sealy & Smith Foundation, UTMB (CON29094).

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: CLIP, crosslinking and immunoprecipitation; DMS, dimethyl sulfate; EBV, Epstein-Barr virus; FHV, Flock House virus; HCV, hepatitis C virus; HITS-CLIP, high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation; IAV, influenza A virus; iCLIP, individual nucleotide resolution CLIP; IN, integrase; IP, immunoprecipitation; KSHV, Kaposi's sarcoma-associated herpesvirus; MA, matrix domain; MaP, mutational profile; NC, nucleocapsid; NGS, next-generation sequencing; PAR-CLIP, PhotoActivatable-Ribonucleoside-Enhanced CrossLinking and ImmunoPrecipitation; PIP-seq, Protein Interaction Profile Sequencing; RBP, RNA-binding protein; RIP, RNA immunoprecipitation; RIPit, RNA:protein immunoprecipitation in tandem; RIP-seq, native RNA immunoprecipitation and sequencing; RNP, ribonucleoprotein; RRE, Rev Response Element; SHAPE, selective 2'-hydroxyl acylation analyzed by primer extension; TAR, trans-activation response; VIR-CLASP, VIRal Crosslinking And Solid-phase Purification; vPAR-CL, viral PhotoActivatable-Ribonucleoside CrossLinking; vRNA, viral RNA; YTHDF, YTH domain family.

References

- Hsin, W. C., Chang, C. H., Chang, C. Y., Peng, W. H., Chien, C. L., Chang, M. F., and Chang, S. C. (2018) Nucleocapsid protein-dependent assembly of the RNA packaging signal of Middle East respiratory syndrome coronavirus. *J. Biomed. Sci.* **25**, 47
- McBride, R., van Zyl, M., and Fielding, B. C. (2014) The coronavirus nucleocapsid is a multifunctional protein. *Viruses* **6**, 2991–3018
- Verheije, M. H., Hagemeijer, M. C., Ulasli, M., Reggiori, F., Rottier, P. J., Masters, P. S., and de Haan, C. A. (2010) The coronavirus nucleocapsid protein is dynamically associated with the replication-transcription complexes. *J. Virol.* **84**, 11575–11579
- Chen, C. Y., Chang, C. K., Chang, Y. W., Sue, S. C., Bai, H. L., Riang, L., Hsiao, C. D., and Huang, T. H. (2007) Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. *J. Mol. Biol.* **368**, 1075–1086
- Nagy, P. D., and Pogany, J. (2011) The dependence of viral RNA replication on co-opted host factors. *Nat. Rev. Microbiol.* **10**, 137–149
- Gilbert, C., and Svejstrup, J. Q. (2006) RNA immunoprecipitation for determining RNA-protein associations *in vivo*. *Curr. Protoc. Mol. Biol.* **Chapter 27**. Unit 27 24
- Niranjanakumari, S., Lasda, E., Brazas, R., and Garcia-Blanco, M. A. (2002) Reversible cross-linking combined with immunoprecipitation to study RNA-protein interactions *in vivo*. *Methods* **26**, 182–190
- Wetz, K., and Habermehl, K.-O. (1982) Specific cross-linking of capsid proteins to virus RNA by ultraviolet irradiation of poliovirus. *J. Gen. Virol.* **59**, 397–401
- Liu, Z.-R., and Smith, C. W. J. (1999) The methylene blue mediated photocrosslinking method for detection of proteins that interact with double-stranded RNA. In: Haynes, S. R., ed. *RNA-protein Interaction Protocols*, Humana Press, Totowa, NJ: 35–47
- Harris, M. E., and Christian, E. L. (2009) RNA crosslinking methods 2009; **468**, 127–146
- Jensen, K. B., and Darnell, R. B. (2008) Clip: Crosslinking and immunoprecipitation of *in vivo* RNA targets of RNA-binding proteins. *Methods Mol. Biol.* **488**, 85–98
- Fillebeen, C., Wilkinson, N., and Pantopoulos, K. (2014) Electrophoretic mobility shift assay (EMSA) for the study of RNA-protein interactions: The IRE/IRP example. *J. Vis. Exp.* <https://doi.org/10.3791/52230>
- Li, Y., Jiang, Z., Chen, H., and Ma, W. J. (2004) A modified quantitative EMSA and its application in the study of RNA-protein interactions. *J. Biochem. Biophys. Methods* **60**, 85–96
- McHugh, C. A., Russell, P., and Guttman, M. (2014) Methods for comprehensive experimental identification of RNA-protein interactions. *Genome Biol.* **15**, 203
- Iioka, H., Loisel, D., Haystead, T. A., and Macara, I. G. (2011) Efficient detection of RNA-protein interactions using tethered RNAs. *Nucl. Acids Res.* **39**, e53
- Rolfsson, O., Middleton, S., Manfield, I. W., White, S. J., Fan, B., Vaughan, R., Ranson, N. A., Dykeman, E., Twarock, R., Ford, J., Kao, C. C., and Stockley, P. G. (2016) Direct evidence for packaging signal-mediated assembly of bacteriophage MS2. *J. Mol. Biol.* **428**, 431–448
- Sotcheff, S., and Routh, A. (2020) Understanding flavivirus capsid protein functions: The tip of the iceberg. *Pathogens* **9**, 42
- Greenberg, J. R. (1979) Ultraviolet light-induced crosslinking of mRNA to proteins. *Nucl. Acids Res.* **6**, 715–732
- Poria, D. K., and Ray, P. S. (2017) RNA-protein UV-crosslinking assay. *Bio Protoc.* **7**, e2193
- Rivera-Santiago, R. F., Sriswasdi, S., Harper, S. L., and Speicher, D. W. (2015) Probing structures of large protein complexes using zero-length cross-linking. *Methods* **89**, 99–111
- Hoffman, E. A., Frey, B. L., Smith, L. M., and Auble, D. T. (2015) Formaldehyde crosslinking: A tool for the study of chromatin complexes. *J. Biol. Chem.* **290**, 26404–26411
- Licatalosi, D. D., Mele, A., Fak, J. J., Ule, J., Kayikci, M., Chi, S. W., Clark, T. A., Schweitzer, A. C., Blume, J. E., Wang, X., Darnell, J. C., and Darnell, R. B. (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr, Jungkamp, A. C., Munschauer, M., Ulrich, A., Wardle, G. S., Dewell, S., Zavolan, M., and Tuschl, T. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141
- Zhou, Y., and Routh, A. (2020) Mapping RNA-capsid interactions and RNA secondary structure within virus particles using next-generation sequencing. *Nucl. Acids Res.* **48**, e12
- Ule, J., Jensen, K. B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R. B. (2003) CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212–1215
- Ule, J., Jensen, K., Mele, A., and Darnell, R. B. (2005) Clip: A method for identifying protein-RNA interaction sites in living cells. *Methods* **37**, 376–386

27. Schoemaker, H. J., and Schimmel, P. R. (1974) Photo-induced joining of a transfer RNA with its cognate aminoacyl-transfer RNA synthetase. *J. Mol. Biol.* **84**, 503–513
28. Darnell, R. B. (2010) HITS-CLIP: Panoramic views of protein-RNA regulation in living cells. *Wiley Interdiscip. Rev. RNA* **1**, 266–286
29. Urdaneta, E. C., and Beckmann, B. M. (2020) Fast and unbiased purification of RNA-protein complexes after UV cross-linking. *Methods* **178**, 72–82
30. Pashev, I. G., Dimitrov, S. I., and Angelov, D. (1991) Crosslinking proteins to nucleic acids by ultraviolet laser irradiation. *Trends Biochem. Sci.* **16**, 323–326
31. Gorner, H. (1994) Photochemistry of DNA and related biomolecules: Quantum yields and consequences of photoionization. *J. Photochem. Photobiol. B* **26**, 117–139
32. Shetlar, M. D. (1980) Cross-Linking of proteins to nucleic acids by ultraviolet light. In Smith, K. C., ed., *Photochemical and Photobiological Reviews* **5**. Springer US, Boston, MA: 105–197
33. Musunuru, K., and Darnell, R. B. (2001) Paraneoplastic neurologic disease antigens: RNA-binding proteins and signaling proteins in neuronal degeneration. *Annu. Rev. Neurosci.* **24**, 239–262
34. Jensen, K. B., Dredge, B. K., Stefani, G., Zhong, R., Buckanovich, R. J., Okano, H. J., Yang, Y. Y. L., and Darnell, R. B. (2000) Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron* **25**, 359–371
35. Voelkerding, K. V., Dames, S. A., and Durtschi, J. D. (2009) Next-generation sequencing: From basic research to diagnostics. *Clin. Chem.* **55**, 641–658
36. Betzel, C., Pal, G. P., and Saenger, W. (1988) Three-dimensional structure of proteinase K at 0.15-nm resolution. *Eur. J. Biochem.* **178**, 155–171
37. Jonas, S., and Izaurralde, E. (2015) Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* **16**, 421–433
38. Chi, S. W., Zang, J. B., Mele, A., and Darnell, R. B. (2009) Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps. *Nature* **460**, 479–486
39. Boss, I. W., Plaisance, K. B., and Renne, R. (2009) Role of virus-encoded microRNAs in herpesvirus biology. *Trends Microbiol.* **17**, 544–553
40. Pfeffer, S., Zavolan, M., Grasser, F. A., Chien, M., Russo, J. J., Ju, J., John, B., Enright, A. J., Marks, D., Sander, C., and Tuschl, T. (2004) Identification of virus-encoded microRNAs. *Science* **304**, 734–736
41. Mesri, E. A., Cesarman, E., and Boshoff, C. (2010) Kaposi's sarcoma and its associated herpesvirus. *Nat. Rev. Cancer* **10**, 707–719
42. Haecker, I., Gay, L. A., Yang, Y., Hu, J., Morse, A. M., McIntyre, L. M., and Renne, R. (2012) Ago HITS-CLIP expands understanding of Kaposi's sarcoma-associated herpesvirus miRNA function in primary effusion lymphomas. *PLoS Pathog.* **8**, e1002884
43. Fais, F., Gaidano, G., Capello, D., Ghiohni, A., Ghiotto, F., Roncella, S., Carbone, A., Chiorazzi, N., and Ferrarini, M. (1999) Immunoglobulin V region gene use and structure suggest antigen selection in AIDS-related primary effusion lymphomas. *Leukemia* **13**, 1093–1099
44. Matolcsy, A., Nádor, R. G., Cesarman, E., and Knowles, D. M. (1998) Immunoglobulin VH gene mutational analysis suggests that primary effusion lymphomas derive from different stages of B cell maturation. *Am. J. Pathol.* **153**, 1609–1614
45. Houldcroft, C. J., and Kellam, P. (2015) Host genetics of Epstein-Barr virus infection, latency and disease. *Rev. Med. Virol.* **25**, 71–84
46. Riley, K. J., Rabinowitz, G. S., Yario, T. A., Luna, J. M., Darnell, R. B., and Steitz, J. A. (2012) EBV and human microRNAs co-target oncogenic and apoptotic viral and human genes during latency. *EMBO J.* **31**, 2207–2221
47. Harold, C., Cox, D., and Riley, K. J. (2016) Epstein-Barr viral microRNAs target caspase 3. *Virol. J.* **13**, 145
48. Porter, A. G., and Janicke, R. U. (1999) Emerging roles of caspase-3 in apoptosis. *Cell Death Differ.* **6**, 99–104
49. Green, D. R., and Llambi, F. (2015) Cell death signaling. *Cold Spring Harb. Perspect. Biol.* **7**, a006080
50. Guo, Y. E., Riley, K. J., Iwasaki, A., and Steitz, J. A. (2014) Alternative capture of noncoding RNAs or protein-coding genes by herpesviruses to alter host T cell function. *Mol. Cell* **54**, 67–79
51. Guo, Y. E., Oei, T., and Steitz, J. A. (2015) Herpesvirus saimiri MicroRNAs preferentially target host cell cycle regulators. *J. Virol.* **89**, 10901–10911
52. Luna, J. M., Scheel, T. K., Danino, T., Shaw, K. S., Mele, A., Fak, J. J., Nishiuchi, E., Takacs, C. N., Catanese, M. T., de Jong, Y. P., Jacobson, I. M., Rice, C. M., and Darnell, R. B. (2015) Hepatitis C virus RNA functionally sequesters miR-122. *Cell* **160**, 1099–1110
53. Jopling, C. L., Yi, M., Lancaster, A. M., Lemon, S. M., and Sarnow, P. (2005) Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science* **309**, 1577–1581
54. Lee, N., Le Sage, V., Nanni, A. V., Snyder, D. J., Cooper, V. S., and Lakdawala, S. S. (2017) Genome-wide analysis of influenza viral RNA and nucleoprotein association. *Nucl. Acids Res.* **45**, 8968–8977
55. Le Sage, V., Nanni, A. V., Bhagwat, A. R., Snyder, D. J., Cooper, V. S., Lakdawala, S. S., and Lee, N. (2018) Non-uniform and non-random binding of nucleoprotein to influenza A and B viral RNA. *Viruses* **10**, 522
56. Ikehata, H., and Ono, T. (2011) The mechanisms of UV mutagenesis. *J. Radiat. Res.* **52**, 115–125
57. Lawrence, C. W., Christensen, R., and Schwartz, A. (1982) Mechanisms of UV mutagenesis in yeast. *Basic Life Sci.* **20**, 109–120
58. Kato, T., Shinoura, Y., Templin, A., and Clark, A. J. (1980) Analysis of ultraviolet light-induced suppressor mutations in the strain of *Escherichia coli* K-12 AB1157: An implication for molecular mechanisms of UV mutagenesis. *Mol. Gen. Genet.* **180**, 283–291
59. Ariza-Mateos, A., Prieto-Vega, S., Diaz-Toledano, R., Birk, A., Szeto, H., Mena, I., Berzal-Herranz, A., and Gomez, J. (2012) RNA self-cleavage activated by ultraviolet light-induced oxidation. *Nucl. Acids Res.* **40**, 1748–1766
60. Wurtmann, E. J., and Wolin, S. L. (2009) RNA under attack: Cellular handling of RNA damage. *Crit. Rev. Biochem. Mol. Biol.* **44**, 34–49
61. Tundup, S., Akhter, Y., Thiagarajan, D., and Hasnain, S. E. (2006) Clusters of PE and PPE genes of *Mycobacterium tuberculosis* are organized in operons: Evidence that PE Rv2431c is co-transcribed with PPE Rv2430c and their gene products interact with each other. *FEBS Lett.* **580**, 1285–1293
62. Itri, F., Monti, D. M., Della Ventura, B., Vinciguerra, R., Chino, M., Gesuele, F., Lombardi, A., Velotta, R., Altucci, C., Birolo, L., Piccoli, R., and Arciello, A. (2016) Femtosecond UV-laser pulses to unveil protein-protein interactions in living cells. *Cell Mol. Life Sci.* **73**, 637–648
63. Schwartz, J. C., Wang, X., Podell, E. R., and Cech, T. R. (2013) RNA seeds higher-order assembly of FUS protein. *Cell Rep.* **5**, 918–925
64. Brimacombe, R., Stiege, W., Kyriatsoulis, A., and Maly, P. (1988) Intra-RNA and RNA-protein cross-linking techniques in *Escherichia coli* ribosomes. *Methods Enzymol.* **164**, 287–309
65. Zwieb, C., Ross, A., Rinke, J., Meinke, M., and Brimacombe, R. (1978) Evidence for RNA-RNA cross-link formation in *Escherichia coli* ribosomes. *Nucl. Acids Res.* **5**, 2705–2720
66. Bottini, S., Pratella, D., Grandjean, V., Repetto, E., and Trabucchi, M. (2018) Recent computational developments on CLIP-seq data analysis and microRNA targeting implications. *Brief Bioinform.* **19**, 1290–1301
67. Wang, T., Xiao, G., Chu, Y., Zhang, M. Q., Corey, D. R., and Xie, Y. (2015) Design and bioinformatics analysis of genome-wide CLIP experiments. *Nucl. Acids Res.* **43**, 5263–5274
68. Zhang, C., and Darnell, R. B. (2011) Mapping *in vivo* protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat. Biotechnol.* **29**, 607–614
69. Haecker, I., and Renne, R. (2014) HITS-CLIP and PAR-CLIP advance viral miRNA targetome analysis. *Crit. Rev. Eukaryot. Gene Expr.* **24**, 101–116
70. Gillen, A. E., Yamamoto, T. M., Kline, E., Hesselberth, J. R., and Kabos, P. (2016) Improvements to the HITS-CLIP protocol eliminate widespread mispriming artifacts. *BMC Genomics* **17**, 338
71. Meisenheimer, K. M., and Koch, T. H. (1997) Photocross-linking of nucleic acids to associated proteins. *Crit. Rev. Biochem. Mol. Biol.* **32**, 101–140
72. Favre, A., Moreno, G., Blondel, M. O., Kliber, J., Vinzens, F., and Salet, C. (1986) 4-thiouridine photosensitized RNA-protein crosslinking in mammalian cells. *Biochem. Biophys. Res. Commun.* **141**, 847–854

73. Favre, A., Saintome, C., Fourrey, J. L., Clivio, P., and Laugaa, P. (1998) Thionucleobases as intrinsic photoaffinity probes of nucleic acid structure and nucleic acid protein interactions. *J. Photoch. Photobiol B* **42**, 109–124
74. Favre, A. (1990) 4-thiouridine as an intrinsic photoaffinity probe of nucleic acid structure and interactions. *Bioorg. Photochem.* **1**, 379–425
75. Melvin, W. T., Milne, H. B., Slater, A. A., Allen, H. J., and Keir, H. M. (1978) Incorporation of 6-thioguanosine and 4-thiouridine into Rna - application to isolation of newly synthesized Rna by affinity chromatography. *Eur. J. Biochem.* **92**, 373–379
76. Tanner, N. K., Hanna, M. M., and Abelson, J. (1988) Binding interactions between yeast tRNA ligase and a precursor transfer ribonucleic acid containing two photoreactive uridine analogs. *Biochemistry* **27**, 8852–8861
77. Meisenheimer, K. M., Meisenheimer, P. L., and Koch, T. H. (2000) Nucleoprotein photo-cross-linking using halopyrimidine-substituted RNAs. *Methods. Enzymol.* **318**, 88–104
78. Testa, S. M., Disney, M. D., Turner, D. H., and Kierzek, R. (1999) Thermodynamics of RNA–RNA duplexes with 2- or 4-thiouridines: Implications for antisense design and targeting a group I intron. *Biochemistry* **38**, 16655–16662
79. Ascano, M., Hafner, M., Cekan, P., Gerstberger, S., and Tuschl, T. (2012) Identification of RNA–protein interaction networks using PAR-CLIP. *Wiley Interdiscip. Rev. RNA* **3**, 159–177
80. Ascano, M., Hafner, M., Cekan, P., Gerstberger, S., and Tuschl, T. (2012) Identification of RNA–protein interaction networks using PAR-CLIP. *WIREs RNA* **3**, 159–177
81. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A. C., Munschauer, M., Ulrich, A., Wardle, G. S., Dewell, S., Zavolan, M., and Tuschl, T. (2010) PAR-CLIP—a method to identify transcriptome-wide the binding sites of RNA binding proteins. *J. Vis. Exp.* <https://doi.org/10.3791/2034>
82. Spitzer, J., Hafner, M., Landthaler, M., Ascano, M., Farazi, T., Wardle, G., Nusbaum, J., Khorshid, M., Burger, L., Zavolan, M., and Tuschl, T. (2014) PAR-CLIP (photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation): A step-by-step protocol to the transcriptome-wide identification of binding sites of RNA-binding proteins. *Methods Enzymol.* **539**, 113–161
83. Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M., and Zavolan, M. (2011) A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat. Methods* **8**, 559–564
84. Gogakos, T., Brown, M., Garzia, A., Meyer, C., Hafner, M., and Tuschl, T. (2017) Characterizing expression and processing of precursor and mature human tRNAs by Hydro-tRNAseq and PAR-CLIP. *Cell Rep.* **20**, 1463–1475
85. Skalsky, R. L., Corcoran, D. L., Gottwein, E., Frank, C. L., Kang, D., Hafner, M., Nusbaum, J. D., Feederle, R., Delecluse, H. J., Luftig, M. A., Tuschl, T., Ohler, U., and Cullen, B. R. (2012) The viral and cellular microRNA targetome in lymphoblastoid cell lines. *PLoS Pathog.* **8**, e1002484
86. Gottwein, E., Corcoran, D. L., Mukherjee, N., Skalsky, R. L., Hafner, M., Nusbaum, J. D., Shamulailatpam, P., Love, C. L., Dave, S. S., Tuschl, T., Ohler, U., and Cullen, B. R. (2011) Viral microRNA targetome of KSHV-infected primary effusion lymphoma cell lines. *Cell Host & Microbe* **10**, 515–526
87. Bell, N. M., and Lever, A. M. (2013) HIV gag polyprotein: Processing and early viral particle assembly. *Trends Microbiol.* **21**, 136–144
88. Sundquist, W. L., and Krausslich, H. G. (2012) HIV-1 assembly, budding, and maturation. *Cold Spring Harb. Perspect. Med.* **2**, a006924
89. Kutluay, S. B., and Bieniasz, P. D. (2016) Analysis of HIV-1 gag-RNA interactions in cells and virions by CLIP-seq. *Methods Mol. Biol.* **1354**, 119–131
90. Kutluay, S. B., Zang, T., Blanco-Melo, D., Powell, C., Jannain, D., Errando, M., and Bieniasz, P. D. (2014) Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. *Cell* **159**, 1096–1109
91. Rulli, S. J., Jr, Hibbert, C. S., Mirro, J., Pederson, T., Biswal, S., and Rein, A. (2007) Selective and nonselective packaging of cellular RNAs in retrovirus particles. *J. Virol.* **81**, 6623–6631
92. Georgiev, S., Boyle, A. P., Jayasurya, K., Ding, X., Mukherjee, S., and Ohler, U. (2010) Evidence-ranked motif identification. *Genome Biol.* **11**, R19
93. Alfadhli, A., and Barklis, E. (2014) The roles of lipids and nucleic acids in HIV-1 assembly. *Front. Microbiol.* **5**, 253
94. Craigie, R. (2012) The molecular biology of HIV integrase. *Future Virol.* **7**, 679–686
95. Delelis, O., Carayon, K., Saib, A., Deprez, E., and Mouscadet, J. F. (2008) Integrase and integration: Biochemical activities of HIV-1 integrase. *Retrovirology* **5**, 114
96. Chiu, T. K., and Davies, D. R. (2004) Structure and function of HIV-1 integrase. *Curr. Top Med. Chem.* **4**, 965–977
97. Engelman, A. (1999) *In vivo* analysis of retroviral integrase structure and function. *Adv. Virus Res.* **52**, 411–426
98. Johnson, B. C., Metifiot, M., Ferris, A., Pommier, Y., and Hughes, S. H. (2013) A homology model of HIV-1 integrase and analysis of mutations designed to test the model. *J. Mol. Biol.* **425**, 2133–2146
99. Mohammed, K. D., Topper, M. B., and Muesing, M. A. (2011) Sequential deletion of the integrase (Gag-Pol) carboxyl terminus reveals distinct phenotypic classes of defective HIV-1. *J. Virol.* **85**, 4654–4666
100. Lu, R., Limon, A., Devroe, E., Silver, P. A., Cherepanov, P., and Engelman, A. (2004) 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.* **78**, 12735–12746
101. Engelman, A., Englund, G., Orenstein, J. M., Martin, M. A., and Craigie, R. (1995) Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. *J. Virol.* **69**, 2729–2736
102. Kessl, J. J., Kutluay, S. B., Townsend, D., Rebusburg, S., Slaughter, A., Larue, R. C., Shkriabai, N., Bakouche, N., Fuchs, J. R., Bieniasz, P. D., and Kvaratskhelia, M. (2016) HIV-1 integrase binds the viral RNA genome and is essential during virion morphogenesis. *Cell* **166**, 1257–1268. e1212
103. Bannwarth, S., and Gatignol, A. (2005) HIV-1 TAR RNA: The target of molecular interactions between the virus and its host. *Curr. HIV Res.* **3**, 61–71
104. Kvaratskhelia, M., and Grice, S. F. (2008) Structural analysis of protein-RNA interactions with mass spectrometry. *Methods Mol. Biol.* **488**, 213–219
105. Sharma, A., Slaughter, A., Jena, N., Feng, L., Kessl, J. J., Fadel, H. J., Malani, N., Male, F., Wu, L., Poeschla, E., Bushman, F. D., Fuchs, J. R., and Kvaratskhelia, M. (2014) A new class of multimerization selective inhibitors of HIV-1 integrase. *PLoS Pathog.* **10**, e1004171
106. Madison, M. K., Lawson, D. Q., Elliott, J., Ozanturk, A. N., Koneru, P. C., Townsend, D., Errando, M., Kvaratskhelia, M., and Kutluay, S. B. (2017) Allosteric HIV-1 integrase inhibitors lead to premature degradation of the viral RNA genome and integrase in target cells. *J. Virol.* **91**
107. Engelman, A., Kessl, J. J., and Kvaratskhelia, M. (2013) Allosteric inhibition of HIV-1 integrase activity. *Curr. Opin. Chem. Biol.* **17**, 339–345
108. Bieniasz, P. D., and Kutluay, S. B. (2018) CLIP-related methodologies and their application to retrovirology. *Retrovirology* **15**, 35
109. Stavrou, S., and Ross, S. R. (2015) APOBEC3 proteins in viral immunity. *J. Immunol.* **195**, 4565–4570
110. Harris, R. S., and Dudley, J. P. (2015) APOBECs and virus restriction. *Virology* **479–480**, 131–145
111. York, A., Kutluay, S. B., Errando, M., and Bieniasz, P. D. (2016) The RNA binding specificity of human APOBEC3 proteins resembles that of HIV-1 nucleocapsid. *PLoS Pathog.* **12**, e1005833
112. Bohn, J. A., Thummar, K., York, A., Raymond, A., Brown, W. C., Bieniasz, P. D., Hatziioannou, T., and Smith, J. L. (2017) APOBEC3H structure reveals an unusual mechanism of interaction with duplex RNA. *Nat. Commun.* **8**, 1021
113. Zhu, Y., Chen, G., Lv, F., Wang, X., Ji, X., Xu, Y., Sun, J., Wu, L., Zheng, Y. T., and Gao, G. (2011) Zinc-finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs for degradation. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 15834–15839
114. Takata, M. A., Goncalves-Carneiro, D., Zang, T. M., Soll, S. J., York, A., Blanco-Melo, D., and Bieniasz, P. D. (2017) CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature* **550**, 124–127

115. Karlin, S., and Mrazek, J. (1997) Compositional differences within and between eukaryotic genomes. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 10227–10232
116. van der Kuyl, A. C., and Berkhout, B. (2012) The biased nucleotide composition of the HIV genome: A constant factor in a highly variable virus. *Retrovirology* **9**, 92
117. Antzin-Anduetza, I., Mahiet, C., Granger, L. A., Odendall, C., and Swanson, C. M. (2017) Increasing the CpG dinucleotide abundance in the HIV-1 genomic RNA inhibits viral replication. *Retrovirology* **14**, 49
118. Kint, S., Trypsteen, W., De Spiegelaere, W., Malatinkova, E., Kinloch-de Loes, S., De Meyer, T., Van Criekeing, W., and Vandekerckhove, L. (2020) Underestimated effect of intragenic HIV-1 DNA methylation on viral transcription in infected individuals. *Clin. Epigenetics* **12**, 36
119. Riquelme-Barrios, S., Pereira-Montecinos, C., Valiente-Echeverria, F., and Soto-Rifo, R. (2018) Emerging roles of N(6)-methyladenosine on HIV-1 RNA metabolism and viral replication. *Front. Microbiol.* **9**, 576
120. Zaccara, S., and Jaffrey, S. R. (2020) A unified model for the function of YTHDF proteins in regulating m(6)a-modified mRNA. *Cell* **181**, 1582–1595.e1518
121. Kennedy, E. M., Bogerd, H. P., Kornepati, A. V., Kang, D., Ghoshal, D., Marshall, J. B., Poling, B. C., Tsai, K., Gokhale, N. S., Horner, S. M., and Cullen, B. R. (2016) Posttranscriptional m(6)A editing of HIV-1 mRNAs enhances viral gene expression. *Cell Host & Microbe* **19**, 675–685
122. Tirumuru, N., Zhao, B. S., Lu, W., Lu, Z., He, C., and Wu, L. (2016) N(6)-methyladenosine of HIV-1 RNA regulates viral infection and HIV-1 Gag protein expression. *Elife* **5**, e15528
123. Lozzio, C. B., and Wigler, P. W. (1971) Cytotoxic effects of thiopyrimidines. *J. Cell Physiol.* **78**, 25–32
124. Burger, K., Muhl, B., Kellner, M., Rohrmoser, M., Gruber-Eber, A., Windhager, L., Friedel, C. C., Dolken, L., and Eick, D. (2013) 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. *RNA Biol.* **10**, 1623–1630
125. Jaworski, E., and Routh, A. (2018) ClickSeq: Replacing fragmentation and enzymatic ligation with click-chemistry to prevent sequence chimeras. In: Head, S. R., Ordoukhanian, P., Salomon, D. R., eds. *Next Generation Sequencing: Methods and Protocols*, Springer New York, New York, NY: 71–85
126. Routh, A., Head, S. R., Ordoukhanian, P., and Johnson, J. E. (2015) ClickSeq: Fragmentation-Free next-generation sequencing *via* click ligation of adaptors to stochastically terminated 3'-azido cDNAs. *J. Mol. Biol.* **427**, 2610–2616
127. Sanjuan, R., and Domingo-Calap, P. (2016) Mechanisms of viral mutation. *Cell Mol. Life Sci.* **73**, 4433–4448
128. Tihova, M., Dryden, K. A., Le, T. V. L., Harvey, S. C., Johnson, J. E., Yeager, M., and Schneemann, A. (2004) Nodavirus coat protein imposes dodecahedral RNA structure independent of nucleotide sequence and length. *J. Virol.* **78**, 2897–2905
129. Zubradt, M., Gupta, P., Persad, S., Lambowitz, A. M., Weissman, J. S., and Rouskin, S. (2017) DMS-MaPseq for genome-wide or targeted RNA structure probing *in vivo*. *Nat. Methods* **14**, 75–82
130. Lee, F. C. Y., and Ule, J. (2018) Advances in CLIP technologies for studies of protein-RNA interactions. *Mol. Cell* **69**, 354–369
131. Le Sage, V., Kanarek, J. P., Snyder, D. J., Cooper, V. S., Lakdawala, S. S., and Lee, N. (2020) Mapping of influenza virus RNA-RNA interactions reveals a flexible network. *Cell Rep.* **31**, 107823
132. Lin, C., and Miles, W. O. (2019) Beyond CLIP: Advances and opportunities to measure RBP-RNA and RNA-RNA interactions. *Nucl. Acids Res.* **47**, 5490–5501
133. Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L. E., Sibley, C. R., Sugimoto, Y., Tajnik, M., Konig, J., and Ule, J. (2014) iCLIP: protein-RNA interactions at nucleotide resolution. *Methods* **65**, 274–287
134. Zarnegar, B. J., Flynn, R. A., Shen, Y., Do, B. T., Chang, H. Y., and Khavari, P. A. (2016) irCLIP platform for efficient characterization of protein-RNA interactions. *Nat. Methods* **13**, 489–492
135. Kargapolova, Y., Levin, M., Lackner, K., and Danckwardt, S. (2017) sCLIP-an integrated platform to study RNA-protein interactomes in biomedical research: identification of CSTF2tau in alternative processing of small nuclear RNAs. *Nucl. Acids Res.* **45**, 6074–6086
136. Van Nostrand, E. L., Pratt, G. A., Shishkin, A. A., Gelboin-Burkhardt, C., Fang, M. Y., Sundaraman, B., Blue, S. M., Nguyen, T. B., Surka, C., Elkins, K., Stanton, R., Rigo, F., Guttman, M., and Yeo, G. W. (2016) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514
137. Quievryn, G., and Zhitkovich, A. (2000) Loss of DNA-protein crosslinks from formaldehyde-exposed cells occurs through spontaneous hydrolysis and an active repair process linked to proteasome function. *Carcinogenesis* **21**, 1573–1580
138. Solomon, M. J., and Varshavsky, A. (1985) Formaldehyde-mediated DNA-protein crosslinking: A probe for *in vivo* chromatin structures. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 6470–6474
139. Hahnenstein, I., Albert, M., Hasse, H., Kreiter, C. G., and Maurer, G. (1995) NMR spectroscopic and densimetric study of reaction kinetics of formaldehyde polymer formation in water, deuterium oxide, and methanol. *Ind. Eng. Chem. Res.* **34**, 440–450
140. Au, P. C., Helliwell, C., and Wang, M. B. (2014) Characterizing RNA-protein interaction using cross-linking and metabolite supplemented nuclear RNA-immunoprecipitation. *Mol. Biol. Rep.* **41**, 2971–2977
141. Silverman, I. M., Li, F., Alexander, A., Goff, L., Trapnell, C., Rinn, J. L., and Gregory, B. D. (2014) RNase-mediated protein footprint sequencing reveals protein-binding sites throughout the human transcriptome. *Genome Biol.* **15**, R3
142. Wisniewski, J. R., Dus, K., and Mann, M. (2013) Proteomic workflow for analysis of archival formalin-fixed and paraffin-embedded clinical samples to a depth of 10 000 proteins. *Proteomics Clin. Appl.* **7**, 225–233
143. Shi, S. R., Taylor, C. R., Fowler, C. B., and Mason, J. T. (2013) Complete solubilization of formalin-fixed, paraffin-embedded tissue may improve proteomic studies. *Proteomics Clin. Appl.* **7**, 264–272
144. Jackson, V. (1978) Studies on histone organization in the nucleosome using formaldehyde as a reversible cross-linking agent. *Cell* **15**, 945–954
145. Foley, S. W., and Gregory, B. D. (2016) Protein interaction profile sequencing (PIP-seq). *Curr. Protoc. Mol. Biol.* **116**. <https://doi.org/10.1002/cpmb.21>
146. Silverman, I. M., and Gregory, B. D. (2015) Transcriptome-wide ribonuclease-mediated protein footprinting to identify RNA-protein interaction sites. *Methods* **72**, 76–85
147. Li, F., Zheng, Q., Ryvkin, P., Dragomir, I., Desai, Y., Aiyer, S., Valladares, O., Yang, J., Bambina, S., Sabin, L. R., Murray, J. I., Lamitina, T., Raj, A., Cherry, S., Wang, L. S., et al. (2012) Global analysis of RNA secondary structure in two metazoans. *Cell Rep.* **1**, 69–82
148. Zheng, Q., Ryvkin, P., Li, F., Dragomir, I., Valladares, O., Yang, J., Cao, K., Wang, L. S., and Gregory, B. D. (2010) Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in Arabidopsis. *PLoS Genet.* **6**, e1001141
149. Foley, S. W., Vandivier, L. E., Kuksa, P. P., and Gregory, B. D. (2015) Transcriptome-wide measurement of plant RNA secondary structure. *Curr. Opin. Plant Biol.* **27**, 36–43
150. Anderson, S. J., Willmann, M. R., and Gregory, B. D. (2016) Protein interaction profile sequencing (PIP-seq) in plants. *Curr. Protoc. Plant Biol.* **1**, 163–183
151. Singh, G., Ricci, E. P., and Moore, M. J. (2014) RIPit-seq: A high-throughput approach for footprinting RNA:protein complexes. *Methods* **65**, 320–332
152. LaCava, J., Fernandez-Martinez, J., Hakhverdyan, Z., and Rout, M. P. (2016) Protein complex purification by affinity capture. *Cold Spring Harb. Protoc.* <https://doi.org/10.1101/pdb.top077545>
153. Bosch, J. A., Chen, C. L., and Perrimon, N. (2021) Proximity-dependent labeling methods for proteomic profiling in living cells: An update. *Wiley Interdiscip. Rev. Dev. Biol.* **10**, e392
154. Chen, C. L., and Perrimon, N. (2017) Proximity-dependent labeling methods for proteomic profiling in living cells. *Wiley Interdiscip. Rev. Dev. Biol.* **6**. <https://doi.org/10.1002/wdev.272>
155. Hung, V., Udeshi, N. D., Lam, S. S., Loh, K. H., Cox, K. J., Pedram, K., Carr, S. A., and Ting, A. Y. (2016) Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nat. Protoc.* **11**, 456–475

156. Benhalevy, D., Anastasakis, D. G., and Hafner, M. (2018) Proximity-CLIP provides a snapshot of protein-occupied RNA elements in sub-cellular compartments. *Nat. Methods* **15**, 1074–1082
157. Padron, A., Iwasaki, S., and Ingolia, N. T. (2019) Proximity RNA labeling by APEX-seq reveals the organization of translation initiation complexes and repressive RNA granules. *Mol. Cell* **75**, 875–887.e5
158. Fazal, F. M., Han, S., Parker, K. R., Kaewsapsak, P., Xu, J., Boettiger, A. N., Chang, H. Y., and Ting, A. Y. (2019) Atlas of subcellular RNA localization revealed by APEX-seq. *Cell* **178**, 473–490.e26
159. Kaewsapsak, P., Shechner, D. M., Mallard, W., Rinn, J. L., and Ting, A. Y. (2017) Live-cell mapping of organelle-associated RNAs via proximity biotinylation combined with protein-RNA crosslinking. *Elife* **6**, e29224
160. Wu, K. E., Fazal, F. M., Parker, K. R., Zou, J., and Chang, H. Y. (2020) RNA-GPS predicts SARS-CoV-2 RNA residency to host mitochondria and nucleolus. *Cell Syst.* **11**, 102–108.e3
161. Kim, B., Arcos, S., Rothamel, K., and Ascano, M. (2021) Viral cross-linking and solid-phase purification enables discovery of ribonucleo-protein complexes on incoming RNA virus genomes. *Nat. Protoc.* **16**, 516–531
162. DeAngelis, M. M., Wang, D. G., and Hawkins, T. L. (1995) Solid-phase reversible immobilization for the isolation of PCR products. *Nucl. Acids Res.* **23**, 4742–4743
163. Kim, B., Arcos, S., Rothamel, K., Jian, J., Rose, K. L., McDonald, W. H., Bian, Y., Reasoner, S., Barrows, N. J., Bradrick, S., Garcia-Blanco, M. A., and Ascano, M. (2020) Discovery of widespread host protein interactions with the pre-replicated genome of CHIKV using VIR-CLASP. *Mol. Cell* **78**, 624–640.e627
164. Barrows, N. J., Anglero-Rodriguez, Y., Kim, B., Jamison, S. F., Le Sommer, C., McGee, C. E., Pearson, J. L., Dimopoulos, G., Ascano, M., Bradrick, S. S., and Garcia-Blanco, M. A. (2019) Dual roles for the ER membrane protein complex in flavivirus infection: Viral entry and protein biogenesis. *Sci. Rep.* **9**, 9711
165. Yamada, T., Sato, S., Sotoyama, Y., Orba, Y., Sawa, H., Yamauchi, H., Sasaki, M., and Takaoka, A. (2021) RIG-I triggers a signaling-abortive anti-SARS-CoV-2 defense in human lung cells. *Nat. Immunol.* **22**, 820–828
166. Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., Davey, N. E., Humphreys, D. T., Preiss, T., Steinmetz, L. M., Krijgsvelde, J., and Hentze, M. W. (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406
167. Lambert, N. J., Robertson, A. D., and Burge, C. B. (2015) RNA bind-n-seq: Measuring the binding affinity landscape of RNA-binding proteins. *Methods Enzymol.* **558**, 465–493
168. Lambert, N., Robertson, A., Jangi, M., McGeary, S., Sharp, P. A., and Burge, C. B. (2014) RNA bind-n-seq: Quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol. Cell* **54**, 887–900
169. Wang, Z., Kayikci, M., Briese, M., Zarnack, K., Luscombe, N. M., Rot, G., Zupan, B., Curk, T., and Ule, J. (2010) iCLIP predicts the dual splicing effects of TIA-RNA interactions. *PLoS Biol.* **8**, e1000530
170. Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S., and Brockdorff, N. (1996) Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131–137
171. Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J., and Lee, J. T. (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756
172. Zhao, J., Ohsumi, T. K., Kung, J. T., Ogawa, Y., Grau, D. J., Sarma, K., Song, J. J., Kingston, R. E., Borowsky, M., and Lee, J. T. (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell* **40**, 939–953
173. Zambelli, F., and Pavesi, G. (2015) RIP-Seq data analysis to determine RNA-protein associations. *Methods Mol. Biol.* **1269**, 293–303
174. Wang, Z., and Xie, Y. (2020) Chapter Twenty-Four - RIP-seq in epigenetic analyses. In: Tollesbol, T., ed. *Epigenetics Methods*, Academic Press, Birmingham, AL: **18**, 495–506
175. Jain, R., Devine T Fau - George, A. D., George Ad Fau - Chittur, S. V., Chittur Sv Fau - Baroni, T. E., Baroni Te Fau - Penalva, L. O., Penalva Lo Fau - Tenenbaum, S. A., and Tenenbaum, S. A. RIP-chip Analysis: RNA-Binding Protein Immunoprecipitation-Microarray (Chip) Profiling.
176. Wessels, H.-H., Hirsekorn, A., Ohler, U., and Mukherjee, N. (2016) Identifying RBP targets with RIP-seq. In: Dassi, E., ed. *Post-transcriptional Gene Regulation*, Springer New York, New York, NY: 141–152
177. Dahm, G. M., Gubin, M. M., Magee, J. D., Techasintana, P., Calaluze, R., and Atasoy, U. (2012) Method for the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts using RIP-Chip. *J. Vis. Exp.* <https://doi.org/10.3791/3851>
178. Lichinchi, G., Gao, S., Saletore, Y., Gonzalez, G. M., Bansal, V., Wang, Y., Mason, C. E., and Rana, T. M. (2016) Dynamics of the human and viral m(6)A RNA methylomes during HIV-1 infection of T cells. *Nat. Microbiol.* **1**, 16011
179. Linder, B., Grozhik, A. V., Orlarerin-George, A. O., Meydan, C., Mason, C. E., and Jaffrey, S. R. (2015) Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* **12**, 767–772
180. Liu, J., Xu, Y. P., Li, K., Ye, Q., Zhou, H. Y., Sun, H., Li, X., Yu, L., Deng, Y. Q., Li, R. T., Cheng, M. L., He, B., Zhou, J., Li, X. F., Wu, A., et al. (2021) The m(6)A methylome of SARS-CoV-2 in host cells. *Cell Res.* **31**, 404–414
181. Nie, Z., Zhou, F., Li, D., Lv, Z., Chen, J., Liu, Y., Shu, J., Sheng, Q., Yu, W., Zhang, W., Jiang, C., Yao, Y., Yao, J., Jin, Y., and Zhang, Y. (2013) RIP-seq of BmAgo2-associated small RNAs reveal various types of small non-coding RNAs in the silkworm, *Bombyx mori*. *BMC Genomics* **14**, 661
182. Ayoubian, H., Ludwig, N., Fehlmann, T., Menegatti, J., Groger, L., Anastasiadou, E., Trivedi, P., Keller, A., Meese, E., and Grasser, F. A. (2019) Epstein-barr virus infection of cell lines derived from diffuse large B-cell lymphomas alters MicroRNA loading of the Ago2 complex. *J. Virol.* **93**, e01297-18
183. Skalsky, R. L., Kang, D., Linnstaedt, S. D., and Cullen, B. R. (2014) Evolutionary conservation of primate lymphocryptovirus microRNA targets. *J. Virol.* **88**, 1617–1635
184. Boudreault, S., Armero, V. E. S., Scott, M. S., Perreault, J. P., and Bisailon, M. (2019) The Epstein-Barr virus EBNA1 protein modulates the alternative splicing of cellular genes. *Virology* **16**, 29
185. Riley, Kasandra J., and Steitz, Joan A. (2013) The “observer effect” in genome-wide surveys of protein-RNA interactions. *Mol. Cell* **49**, 601–604
186. Riley, K. J., Yario, T. A., and Steitz, J. A. (2012) Association of Argonaute proteins and microRNAs can occur after cell lysis. *RNA* **18**, 1581–1585
187. Mili, S., and Steitz, J. A. (2004) Evidence for reassociation of RNA-binding proteins after cell lysis: Implications for the interpretation of immunoprecipitation analyses. *RNA* **10**, 1692–1694
188. Baltimore, D., and Huang, A. S. (1970) Interaction of HeLa cell proteins with RNA. *J. Mol. Biol.* **47**, 263–273
189. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J. S. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures *in vivo*. *Nature* **505**, 701–705
190. Sukosd, Z., Swenson, M. S., Kjems, J., and Heitsch, C. E. (2013) Evaluating the accuracy of SHAPE-directed RNA secondary structure predictions. *Nucl. Acids Res.* **41**, 2807–2816
191. Cordero, P., Kladwang, W., VanLang, C. C., and Das, R. (2012) Quantitative dimethyl sulfate mapping for automated RNA secondary structure inference. *Biochemistry* **51**, 7037–7039
192. Low, J. T., and Weeks, K. M. (2010) SHAPE-directed RNA secondary structure prediction. *Methods* **52**, 150–158
193. Deigan, K. E., Li, T. W., Mathews, D. H., and Weeks, K. M. (2009) Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 97–102
194. Mailler, E., Paillart, J. C., Marquet, R., Smyth, R. P., and Vivet-Boudou, V. (2019) The evolution of RNA structural probing methods: From gels to next-generation sequencing. *Wiley Interdiscip. Rev. RNA* **10**, e1518
195. Strobel, E. J., Yu, A. M., and Lucks, J. B. (2018) High-throughput determination of RNA structures. *Nat. Rev. Genet.* **19**, 615–634

196. Smola, M. J., and Weeks, K. M. (2018) In-cell RNA structure probing with SHAPE-MaP. *Nat. Protoc.* **13**, 1181–1195
197. Smola, M. J., Rice, G. M., Busan, S., Siegfried, N. A., and Weeks, K. M. (2015) Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.* **10**, 1643–1669
198. Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A., and Weeks, K. M. (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* **11**, 959–965
199. Kenyon, J. C., Prestwood, L. J., and Lever, A. M. (2015) A novel combined RNA-protein interaction analysis distinguishes HIV-1 Gag protein binding sites from structural change in the viral RNA leader. *Sci. Rep.* **5**, 14369
200. Rice, A. P. (2017) The HIV-1 Tat protein: Mechanism of action and target for HIV-1 cure strategies. *Curr. Pharm. Des.* **23**, 4098–4102
201. Jayaprakash, A. D., Jabado, O., Brown, B. D., and Sachidanandam, R. (2011) Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucl. Acids Res.* **39**, e141
202. De, S., and Gorospe, M. (2017) Bioinformatic tools for analysis of CLIP ribonucleoprotein data. *Wiley Interdiscip. Rev. RNA* **8**. <https://doi.org/10.1002/wrna.1404>
203. Atasoy, U., Watson, J., Patel, D., and Keene, J. D. (1998) ELAV protein HuA (HuR) can redistribute between nucleus and cytoplasm and is upregulated during serum stimulation and T cell activation. *J. Cell Sci.* **111**, 3145–3156
204. Jaworski, E., and Routh, A. (2017) Parallel ClickSeq and Nanopore sequencing elucidates the rapid evolution of defective-interfering RNAs in Flock House virus. *PLoS Pathog.* **13**, e1006365
205. Williams, G. D., Townsend, D., Wylie, K. M., Kim, P. J., Amarasinghe, G. K., Kutluay, S. B., and Boon, A. C. M. (2018) Nucleotide resolution mapping of influenza A virus nucleoprotein-RNA interactions reveals RNA features required for replication. *Nat. Commun.* **9**, 465
206. Sievers, C., Schlumpf, T., Sawarkar, R., Comoglio, F., and Paro, R. (2012) Mixture models and wavelet transforms reveal high confidence RNA-protein interaction sites in MOV10 PAR-CLIP data. *Nucl. Acids Res.* **40**, e160
207. Pan, D., Flores, O., Umbach, J. L., Pesola, J. M., Bentley, P., Rosato, P. C., Leib, D. A., Cullen, B. R., and Coen, D. M. (2014) A neuron-specific host microRNA targets herpes simplex virus-1 ICP0 expression and promotes latency. *Cell Host & Microbe* **15**, 446–456
208. Sun, B., Yang, X., Hou, F., Yu, X., Wang, Q., Oh, H. S., Raja, P., Pesola, J. M., Vanni, E. A. H., McCarron, S., Morris-Love, J., Ng, A. H. M., Church, G. M., Knipe, D. M., Coen, D. M., *et al.* (2021) Regulation of host and virus genes by neuronal miR-138 favours herpes simplex virus 1 latency. *Nat. Microbiol.* **6**, 682–696
209. Gokhale, N. S., McIntyre, A. B. R., McFadden, M. J., Roder, A. E., Kennedy, E. M., Gandara, J. A., Hopcraft, S. E., Quicke, K. M., Vazquez, C., Willer, J., Ilkayeva, O. R., Law, B. A., Holley, C. L., Garcia-Blanco, M. A., Evans, M. J., *et al.* (2016) N6-Methyladenosine in flaviviridae viral RNA genomes regulates infection. *Cell Host & Microbe* **20**, 654–665
210. Brown, R. S., Anastasakis, D. G., Hafner, M., and Kielian, M. (2020) Multiple capsid protein binding sites mediate selective packaging of the alphavirus genomic RNA. *Nat. Commun.* **11**, 4693
211. Flynn, R. A., Martin, L., Spitale, R. C., Do, B. T., Sagan, S. M., Zarnegar, B., Qu, K., Khavari, P. A., Quake, S. R., Sarnow, P., and Chang, H. Y. (2015) Dissecting noncoding and pathogen RNA-protein interactomes. *RNA* **21**, 135–143
212. Apolonia, L., Schulz, R., Curk, T., Rocha, P., Swanson, C. M., Schaller, T., Ule, J., and Malim, M. H. (2015) Promiscuous RNA binding ensures effective encapsidation of APOBEC3 proteins by HIV-1. *PLoS Pathog.* **11**, e1004609
213. Schmidt, N., Lareau, C. A., Keshishian, H., Ganskih, S., Schneider, C., Hennig, T., Melanson, R., Werner, S., Wei, Y., Zimmer, M., Ade, J., Kirschner, L., Zielinski, S., Dolken, L., Lander, E. S., *et al.* (2021) The SARS-CoV-2 RNA-protein interactome in infected human cells. *Nat. Microbiol.* **6**, 339–353
214. Ooi, Y. S., Majzoub, K., Flynn, R. A., Mata, M. A., Diep, J., Li, J. K., van Buuren, N., Rumachik, N., Johnson, A. G., Puschnik, A. S., Marceau, C. D., Mlera, L., Grabowski, J. M., Kirkegaard, K., Bloom, M. E., *et al.* (2019) An RNA-centric dissection of host complexes controlling flavivirus infection. *Nat. Microbiol.* **4**, 2369–2382
215. Bullard, W. L., Kara, M., Gay, L. A., Sethuraman, S., Wang, Y., Nirmalan, S., Esemeli, A., Feswick, A., Hoffman, B. A., Renne, R., and Tibbetts, S. A. (2019) Identification of murine gammaherpesvirus 68 miRNA-mRNA hybrids reveals miRNA target conservation among gammaherpesviruses including host translation and protein modification machinery. *PLoS Pathog.* **15**, e1007843
216. Jaworski, E., Langsjoen, R. M., Judy, B., Newman, P., Plante, J. A., Plante, K. S., Miller, A. L., Zhou, Y., Swetnam, D., Dong, J., Ren, P., Pyles, R. B., Ksiazek, T., Menachery, V. D., Weaver, S. C., *et al.* (2021) Tiled-ClickSeq for targeted sequencing of complete coronavirus genomes with simultaneous capture of RNA recombination and minority variants. *Elife* **10**, e68479
217. Skalsky, R. L. (2017) Analysis of viral and cellular MicroRNAs in EBV-infected cells. *Methods Mol. Biol.* **1532**, 133–146
218. Smith, K. C., and Aplin, R. T. (1966) A mixed photoproduct of uracil and cysteine (5-S-Cysteine-6-hydrouracil). A possible model for the *in Vivo* cross-linking of deoxyribonucleic acid and protein by ultraviolet light*. *Biochemistry* **5**, 2125–2130
219. Yeo, G. W., Coufal, N. G., Liang, T. Y., Peng, G. E., Fu, X. D., and Gage, F. H. (2009) An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat. Struct. Mol. Biol.* **16**, 130–137