ANNUAL REVIEWS

Annual Review of Genomics and Human Genetics

RNA Conformation Capture by Proximity Ligation

Grzegorz Kudla,[1] Yue Wan,[2,3,4] and Aleksandra Helwak[5]

[1]MRC Human Genetics Unit, University of Edinburgh, Edinburgh EH4 2XU, United Kingdom; email: gkudla@gmail.com

[2]Stem Cell and Regenerative Medicine, Genome Institute of Singapore, Singapore 138672

[3]School of Biological Sciences, Nanyang Technological University, Singapore 637551

[4]Department of Biochemistry, National University of Singapore, Singapore 117596

[5]Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom

Annu. Rev. Genom. Hum. Genet. 2020. 21:81–100

First published as a Review in Advance on April 22, 2020

The *Annual Review of Genomics and Human Genetics* is online at genom.annualreviews.org

https://doi.org/10.1146/annurev-genom-120219-073756

Keywords

CLASH, PARIS, SPLASH, Hi-C, RNA–RNA interactome, miRNA, RNA structure, RNA viruses, bioinformatics

Abstract

RNA proximity ligation is a set of molecular biology techniques used to analyze the conformations and spatial proximity of RNA molecules within cells. A typical experiment starts with cross-linking of a biological sample using UV light or psoralen, followed by partial fragmentation of RNA, RNA–RNA ligation, library preparation, and high-throughput sequencing. In the past decade, proximity ligation has been used to study structures of individual RNAs, networks of interactions between small RNAs and their targets, and whole RNA–RNA interactomes, in models ranging from bacteria to animal tissues and whole animals. Here, we provide an overview of the field, highlight the main findings, review the recent experimental and computational developments, and provide troubleshooting advice for new users. In the final section, we draw parallels between DNA and RNA proximity ligation and speculate on possible future research directions.

1. INTRODUCTION

Although structures of DNA and proteins have been resolved experimentally since the 1950s, it took another 20 years for scientists to observe the first three-dimensional structure of an RNA molecule at atomic resolution (31, 65). There are multiple reasons for this long delay, but one of them is the intrinsic flexibility of RNA: With a few important exceptions, most naturally occurring RNAs do not adopt a single specific structure. Instead, a typical RNA adopts an ensemble of conformations so diverse that common techniques for molecular structure determination, such as X-ray crystallography, nuclear magnetic resonance, or cryo-electron microscopy, become useless. The degree of conformational flexibility differs among classes of RNA; for example, transfer RNAs (tRNAs) typically retain a similar structure throughout their lifetime, whereas riboswitches, messenger RNAs (mRNAs), and long noncoding RNAs typically do not. Nevertheless, transient RNA structures can still play important physiological roles, for example, in posttranscriptional regulation (17).

The high degree of RNA conformational flexibility, its functional importance, and the inadequacy of traditional structural biology methods call for the development of alternative ways to determine RNA structures. Although diverse, these alternative methods have common points: They focus on the identification of secondary rather than tertiary structures, and they often comprise an experiment that identifies structural constraints for an RNA or group of RNAs and a computational analysis that integrates these constraints with a thermodynamic model of RNA folding. A good example of this approach is the combination of thermodynamic modeling with analysis of RNA sequence coevolution, which has allowed the determination of secondary structures of many RNAs, including tRNAs and ribosomal RNAs (rRNAs). Chemical and enzymatic probing methods have long been used to probe structures of individual RNAs in vitro, and in the past decade many have been adapted for transcriptome-wide investigations through next-generation sequencing (62, 71). A wide variety of reagents can identify various aspects of RNA structure, and the experimental pipelines have become well integrated with bioinformatic tools for RNA structure prediction. Methods are being developed for correlated structure probing, which reports on the coexistence of structured or unstructured regions in an RNA molecule (27, 47). However, one limitation of structural probing is that it usually generates one-dimensional information, which does not directly identify both partners of each interaction.

In 2011, Kudla et al. (34) described a novel approach to experimentally map RNA structure and intermolecular RNA–RNA interactions, known as CLASH (cross-linking, ligation, and sequencing of hybrids). CLASH relies on the principle of proximity ligation, whereby RNAs are partially fragmented and pairs of interacting RNA fragments are ligated together and sequenced. In the past decade, RNA proximity ligation techniques have evolved and been applied to a growing set of questions in models ranging from bacteria to animal tissues. There are two widely used strategies for sample preparation prior to proximity ligation: (*a*) protein pull-down, where the RNA duplexes bound to a specific protein or protein complex are isolated, typically after cross-linking with UV light, and (*b*) psoralen cross-linking, where derivatives of a photoactivated compound, psoralen, are used to stabilize and enrich RNA–RNA duplexes. Some approaches emphasize the depth of coverage for a specific subclass of RNA–RNA interaction, while others prioritize breadth of coverage, aiming to resolve the entire transcriptome, though potentially missing less abundant or less stably associated RNA species. There is also a trade-off in the range of interactions detected, as some methods identify pairs of RNAs that are in direct contact or close proximity with each other, while others recover pairs of RNAs that are part of the same ribonucleoprotein (RNP) complex, or even the same subcellular compartment. Here, we review published work on RNA proximity ligation, highlighting the main conclusions and technical details of experimental and

computational methods. We give advice for troubleshooting experiments and computation. In a final section, we speculate on the main challenges and future directions of the RNA proximity ligation field.

2. SUMMARY OF RNA PROXIMITY LIGATION STUDIES TO DATE

2.1. Protein Cross-Linking

The idea that proximity ligation and sequencing can be used to map RNA–RNA interactions originated around 2009 in the Tollervey lab at the University of Edinburgh. The initial impulse came from the development of CRAC [cross-linking and analysis of complementary DNAs (cDNAs)], a high-throughput method to map protein interaction sites in RNA molecules that is related to CLIP-seq (cross-linking and immunoprecipitation followed by sequencing) (20, 69). In CRAC, RNA–protein complexes are stabilized by UV cross-linking, followed by protein purification in denaturing conditions, partial enzymatic digestion of RNA, cDNA library preparation, and sequencing. During the analysis of a CRAC experiment, a chimeric read was found between a box C/D short nucleolar RNA (snoRNA) and the rRNA (5). Interestingly, the mapping locations of the two fragments of the chimera corresponded to the known binding site of the snoRNA to the preribosome. A bioinformatic pipeline was developed to automatically identify and annotate chimeras in yeast Nop1/fibrillarin, Nop56, and Nop58 CRAC data sets and found thousands of chimeras that corresponded to known rRNA–snoRNA interactions (34). In addition to intermolecular interactions, the first set of CLASH experiments generated chimeric reads in which both fragments could be mapped to the same transcript, suggestive of intramolecular interactions. CLASH with Brr2, a helicase associated with the spliceosome, identified abundant chimeras between pairs of fragments of U2 short nuclear RNA (snRNA). These chimeras were consistent with a novel local interaction within the 3′ domain of U2 and a long-range interaction that spanned approximately 900 nucleotides between the central domain and the 3′ end of yeast U2 (34). Gratifyingly, the revised U2 structure derived from these analyses was later confirmed by cryo-electron microscopy analysis of the spliceosome (16).

Initially, CLASH experiments were not specifically designed to generate RNA–RNA chimeras. Rather, chimeras were formed fortuitously, either by an endogenous RNA ligase in the cell lysate or during the 5′- or 3′-linker ligation. Because these steps had not been optimized for the ligation of cellular RNAs to each other, the proportion of chimeric reads was low (~1% of total reads). The original protocol was later modified by changing the order of enzymatic reactions and including a separate RNA–RNA ligation step, which increased the proportion of chimeras to 3% of total mapped reads and to more than 10% of chimeras among reads longer than 70 nucleotides (25, 68). The first set of AGO1 CLASH experiments performed in human HEK293 cells recovered 18,000 distinct microRNA (miRNA)–mRNA interactions (25) and validated a set of newly discovered miRNAs by showing that they can bind specific targets in a seed-dependent manner (15). Although most interactions included the expected canonical base-pairing, which consists of six or more consecutive Watson–Crick pairs between the miRNA 5′-seed region and the target RNA, many interactions were noncanonical and included G–U base pairs, mismatched and bulged nucleotides, no seed binding, or extensive base-pairing between both 5′ and 3′ ends of the miRNA and the target. This observation naturally raised a question: What fraction of the interaction sites represent background created during the experimental procedure? The scarcity of interspecies chimeras in mixed human–yeast experiments supported the reliability of the data but showed that stringent purification of RNA–protein complexes is important for avoiding spurious interactions. Although many studies support the relevance of noncanonical miRNA–target interactions

(6, 57, 75), the functional significance and even the existence of such interactions remain a matter of debate (1, 3).

Proximity ligation was soon adapted for the analysis of miRNA targets in animal tissues and organisms. In 2014, the Rajewsky group (21) added a proximity ligation step to the established PAR-CLIP (photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation) protocol and applied this method to identify miRNA targets in living *Caenorhabditis elegans* using GFP-tagged ALG-1 as a bait. They also identified binding sites for viral miRNAs from Kaposi's sarcoma–associated herpesvirus and Epstein–Barr virus in previously published CLIP data sets and validated the regulatory role of some of these interactions in a reporter assay. In 2015, a new method called CLEAR-CLIP (covalent ligation of endogenous Argonaute-bound RNAs with cross-linking and immunoprecipitation) allowed for the recovery of 130,000 miRNA targets directly from the UV-irradiated mouse neocortex using endogenous AGO as a bait (45). In CLEAR-CLIP, AGO–RNA complexes were enriched by immunoprecipitation, and the purity of the sample was ensured by multiple high-salt and high-detergent washes. A Gene Ontology analysis of the data found associations of miRNAs with specific regulatory functions, such as miR-26 with axon development and locomotion and miR-138 with neurotransmitter functions. The establishment of protocols that can be used with endogenous proteins significantly expanded the applicability of RNA proximity ligation methods. A similar approach, known as qCLASH (quick cross-linking, ligation, and sequencing of hybrids), was also used to study viral infection strategies: Upon infection of B cells with murine gammaherpesvirus 68, 1,500 interactions were found between virus miRNA and host mRNA during lytic infection, latent infection, and reactivation from latency (8). The significant overlap with miRNA–target interactions found in cells infected with different gammaherpesviruses indicates that conserved cellular pathways are preferentially targeted during gammaherpesvirus infection.

Beyond the miRNA interactome, CLASH has been used to study the interactomes of other classes of small RNAs, including Piwi-interacting RNAs (piRNAs) and short RNAs (sRNAs). piRNAs, a class of small RNA implicated in the silencing of transposons in animal germlines, base-pair with all germline mRNAs and many noncoding RNAs, pinpointing coordinated pathways of transcriptome surveillance in *C. elegans* (56). Similar to miRNA–target interactions, piRNA–target interactions depend on a seed sequence near the piRNA 5′ end and supplemental pairing near the 3′ end. In bacteria, sRNAs associated with the RNA chaperone Hfq base-pair with their target RNAs, leading to target degradation or translational inhibition. Variations of CLASH have been applied to recover the complement of sRNA–target interactions in *Escherichia coli* using Hfq (41) or RNase E (74) as a bait. In addition to the discovery of large numbers of previously unknown interactions, these experiments established the seed regions responsible for target binding in multiple sRNAs and revealed the complexity of sRNA regulation following environmental changes, such as iron limitation or reaching the stationary phase.

Although it is most commonly used for the analysis of small RNA interactions, protein cross-linking followed by proximity ligation has also been applied to other questions, from the structural analysis of individual RNAs to large-scale studies of RNA interactomes. The Ule group (63) developed a variation of their iCLIP (individual-nucleotide-resolution cross-linking and immunoprecipitation) method, known as hiCLIP, to draw a map of interactions within mRNAs bound by the double-stranded RNA-binding protein Staufen. In RIPPLiT (RNA immunoprecipitation and proximity ligation in tandem), sequential pull-down of protein components of the exon junction complex provided an atlas of mRNA conformations when bound to this complex, i.e., prior to translation (43). An approach known as MARIO (mapping RNA interactome in vivo) detected RNA–RNA interactions in the vicinity of all RNA-binding proteins using a biotin-linked reagent coupled with cysteine amino acids (48). These studies revealed global properties of the RNA

interactome, such as the degree of connectivity of the RNA–RNA interaction network and the distribution of interaction distances within structured RNAs. Prior to translation, intramolecular interactions were broadly distributed throughout the entire length of mRNA transcripts, with an inverse relationship between interaction density and distance along the transcript, which indicates a lack of locus-specific structure and a rod-like three-dimensional organization of messenger RNPs (43). Consistent with the unfolding of mRNA by translating ribosomes, Staufen-bound RNA duplexes were depleted within the coding sequences of highly translated mRNAs, whereas long-range interactions were prevalent in 3′ untranslated regions (63).

2.2. Psoralen Cross-Linking

In 2016, we and others reported transcriptome-wide maps of RNA–RNA interactions by cross-linking with psoralen, a plant-derived chemical that intercalates into double-stranded nucleic acids and forms reversible covalent bonds between the two strands after UV irradiation (2, 24, 39, 53). Unlike CLASH, psoralen-based methods do not rely on the pull-down of RNA–RNA interactions with a specific protein, and thus they can, in principle, yield transcriptome-wide RNA interactomes. All of the methods utilize similar principles of cross-linking RNAs in vivo but use different strategies for enriching for cross-linked fragments: two-dimensional gel electrophoresis in PARIS (psoralen analysis of RNA interactions and structures) (39), digestion by RNase R in LIGR-seq (ligation of interacting RNA followed by high-throughput sequencing) (53), and biotin–streptavidin enrichment in SPLASH (sequencing of psoralen cross-linked, ligated, and selected hybrids) (2). The cross-linked fragments are then joined together by proximity ligation, reverse cross-linked, and converted into a cDNA library, similar to the protein-based methods.

The psoralen cross-linking experiments uncovered general properties of RNA–RNA interactomes in mammalian cells. PARIS showed extensive intramolecular interactions in human HeLa and 293T cells and in mouse embryonic stem cells, many of which show alternative base-pairing, suggesting substantial structural heterogeneity in the cell (39). This study also showed that some of these alternative base-pairings are conserved between humans and mice, suggesting that they could be functionally important. In addition, PARIS elucidated the structure of repeat A of *Xist* in vivo, indicating that this repeat can base-pair with other repeats to form complex structures, complementing studies of XIST structures obtained with biochemical and bioinformatic methods (40, 52). LIGR-seq experiments in human 293T cells captured known and new snRNA–snRNA and snoRNA–rRNA intermolecular interactions. In particular, they showed that SNORD83B can bind to mRNAs to regulate their gene expression levels, identifying an unexpected role for this snoRNA (53). SPLASH experiments in human lymphoblastoid cells found thousands of intramolecular interactions and showed that long-range intramolecular interactions are associated with higher translation, in agreement with the mRNA circularization model for ribosome recycling. In addition, SPLASH identified thousands of intermolecular RNA–RNA interactions and showed that interacting RNAs tend to have similar biological functions and that the interactions could be remodeled depending on cellular states (2).

SPLASH, PARIS, and COMRADES (cross-linking of matched RNAs and deep sequencing), a new method that uses clickable psoralen and two-step enrichment of cross-linked RNA (78), were also applied to study the RNA genome structures of dengue and Zika viruses (28, 35, 78). All three methods detected dense networks of RNA–RNA interactions within viral genomes inside infected cells. Some of these interactions were functionally important, and their disruption influenced virus fitness. The PARIS study compared lineage-specific pairwise interactions between Asian- and African-specific Zika viruses. An interaction between the 5′ untranslated region and the

E protein–coding sequence was specific to the Asian strain and affected virus fitness when mutated, suggesting that RNA structures could underlie differences in virus pathogenicity (35). SPLASH was used to interrogate four serotypes of dengue viruses and four geographically different strains of Zika viruses inside their virions in order to identify conserved long-range interactions. Many unique and alternative base-pairings were similar across the dengue and Zika genomes, indicating that the overall architecture of these viruses is conserved (28). Comparing in-virion and in-cell structures showed that longer-range virus interactions tend to be disrupted inside cells, which is consistent with studies showing that RNAs tend to be less structured in vivo. In addition to detecting intramolecular virus RNA interactions, COMRADES showed that Zika interacts with numerous small host RNAs, including tRNAs and miRNAs (78). One of the miRNAs, miR-21, was found to interact with the 5′ circularization sequence of Zika to regulate Zika infectivity, deepening our understanding of virus–host RNA interactions. However, given the extensive remodeling of the host cell upon virus entry, it is likely that many more virus–host interactions and their functions are waiting to be uncovered.

Unlike the single positive-strand RNA genome in dengue and Zika viruses, the influenza genome consists of eight negative-strand RNA segments that interact with viral RNPs. Using SPLASH in influenza virions, the Bauer group (10) showed extensive intra- and intermolecular interactions between different genomic segments, leading to redundancy in the interaction network. In addition, this group showed that different strains could have different interactions that drive cosegregation of the genomic segments, deepening our understanding of how segments are associated with each other during virus reassortment. Current published studies include RNA interactomes of dengue, Zika, and influenza viruses, and it is likely that proximity ligation methods will be useful to understand genome organization and virus–host interactions in many different virus systems.

2.3. Other Methods

While most other studies have relied on cross-linking to stabilize RNA–protein or RNA–RNA interactions, the RPL (RNA proximity ligation) approach from the Shendure laboratory omitted the cross-linking step and instead relied on RNA digestion and religation in partially solubilized cells (49). Despite the lack of cross-linking, the interactions were remarkably consistent with known base-pairing patterns in stable RNAs, including the ribosome, snRNAs, and snoRNAs.

Very recently, an approach known as proximity RNA sequencing (RNA-seq) has been developed to measure the spatial association between pairs of RNAs in formaldehyde cross-linked lysates of mammalian cell nuclei (46). Rather than using RNA–RNA ligation, proximity RNA-seq relies on co-barcoding of pairs of RNA molecules in emulsion PCR and detects long-distance associations, such as the association of snoRNAs and a subset of mRNAs with a nucleolar and perinucleolar compartment.

3. A MIX-AND-MATCH TOOLBOX FOR RNA PROXIMITY LIGATION

Although many RNA proximity ligation protocols have been described, most of them follow the same general scheme: The biological sample is first stabilized by cross-linking, followed by optional targeted enrichment of RNA, RNase treatment, proximity ligation, library preparation, optional enrichment of chimeric reads, and sequencing (**Figure 1**). Each of these steps may be combined with others in a mix-and-match fashion that best suits the needs of a particular experiment. In the following sections, we summarize the various options available at each experimental stage.

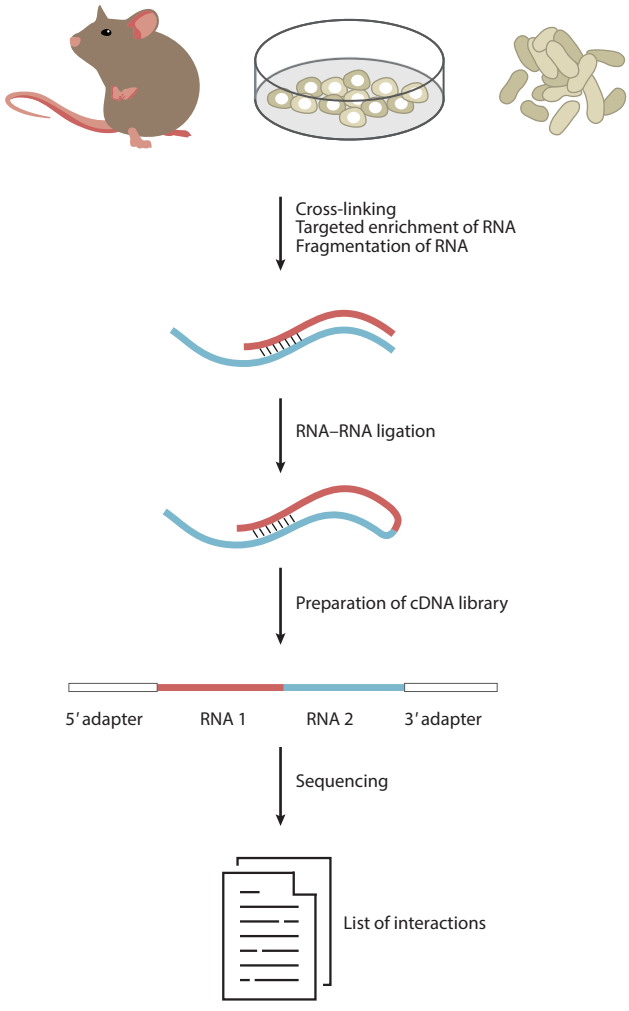

Figure 1

Mix-and-match proximity ligation. Although many protocols for RNA proximity ligation have been developed, most of them follow the scheme shown here, and individual steps can be chosen based on the needs of the specific experiment.

3.1. Stabilization of In Vivo Interactions by Cross-Linking

Most proximity ligation protocols include stabilization of in vivo RNA–RNA or RNA–protein interactions by cross-linking. According to some authors, cross-linking is essential for the recovery of chimeras (45) or reduces the recovery of spurious interactions (41). The most popular approach for the stabilization of RNA duplexes bound by a bait protein is UVC (254 nm) or UVB (~300 nm) cross-linking, which can be applied to a wide range of biological samples, including cell lysates, cell fractions (58), living cells (25, 34), tissues (69), and multicellular organisms (77). The alternative is UVA (365 nm) cross-linking of living cells pretreated with 4-thiouridine (4SU). 4SU/UVA cross-linking is most frequently used with cell lines (22) but has also been used with *C. elegans* (30) and *Saccharomyces cerevisiae* (29). 4SU/UVA cross-linking gives a better signal-to-noise ratio than UVC (55), and unlike UVC, it is not mutagenic, but 4SU treatment is toxic to cells (9).

Whereas UV cross-linking is commonly used to identify RNA–RNA duplexes associated with a specific bait protein, cross-linking with the psoralen derivative 4′-aminomethyl trioxsalen (AMT) has been used in studies of the global RNA–RNA interactome. AMT intercalates into double-stranded RNA and creates covalent bonds primarily between pyrimidines of two RNA strands after exposure to UVA (24). AMT cross-linking is reversible upon exposure to high doses of UVC, which can introduce mutations. A variation introduced in SPLASH (2) is cross-linking with the ThermoFisher Scientific reagent EZ-Link Psoralen-PEG3-Biotin, which allows for specific selection of cross-linked RNA duplexes. Unfortunately, biotinylated psoralen is not easily taken up by mammalian cells and requires pretreatment of cells with digitonin. The intake problem has been solved in COMRADES by the use of psoralen-triethylene glycol azide, which easily enters cells and can be biotinylated using click chemistry at later stages of the protocol (78). So far, psoralen cross-linking has been used in prokaryotic and eukaryotic cells and in virion particles (2, 28, 35, 38, 39, 53, 78).

The choice of cross-linking method depends on the timescale of the process under study. UV cross-linking can be done in a matter of seconds and is appropriate for the analysis of dynamic interactions and to investigate new interactions and changes of RNA conformation induced by rapid changes in the environment. Because 4SU can be used to label nascent transcripts in pulse–chase experiments, 4SU/UVA is the method of choice for studying cotranscriptional interactions. By contrast, published psoralen protocols require 10 minutes of incubation and 20 minutes of cross-linking time and are more appropriate for the study of steady-state interactions.

The cross-linking method also affects the type and spatial range of interactions detected. UV and psoralen cross-linking have a known preference for uracil bases (60). Following UV cross-linking, most protein molecules are cross-linked to only one strand of the RNA duplex, which can skew the results toward interactions that are more stable in vitro. By contrast, psoralen cross-linking introduces covalent bonds into RNA duplexes, allowing recovery of interactions that would otherwise be lost during sample preparation. Both methods are expected to recover direct interactions, though data analysis also reveals some enrichment of indirect interactions. To recover both direct and indirect interactions (e.g., for the characterization of larger RNA–protein complexes), the nonspecific cross-linkers formaldehyde and ethylene glycol bis(succinimidyl succinate) have been used instead (48). Long-range interactions were also detected by RIPPLiT performed on the exon junction complexes in the absence of cross-linking (43).

3.2. Targeted RNA Enrichment

RNA proximity ligation experiments cover a continuum from targeted analysis of individual RNAs (28, 35, 78) to subsets of RNAs and entire transcriptomes (2, 39). The scale on which each study operates is determined mainly by its strategy for targeted enrichment of RNA.

Robust methods exist for specific enrichment of nucleic acids via biotinylated oligonucleotide pull-down (13), and some have been used in proximity ligation studies. To enrich the RNA genome of the Zika virus, one study extracted the genomic RNA from the cell lysate by hybridization to immobilized DNA probes (78). To specifically characterize interactions involving all polyA-tailed mRNAs, a pull-down with oligo d(T) beads has been used (58).

Many experiments have focused on interactions of a subset of RNAs bound to a specific protein of interest. These interactions can be enriched by immunoprecipitation of a protein of choice using an antibody (45), a combination of two antibodies (43), or pull-down of tagged bait proteins. Immunoprecipitation is more flexible, does not require the preparation of transgenic cell lines, and is also suitable for tissues, but a good antibody is essential. Antibody pull-down allows for moderately stringent purification of RNPs using high concentrations of salt and detergents to

reduce the background. To further increase the stringency of the purification, interactions among the antibody, epitope, and beads can be stabilized with formaldehyde, followed by washes with denaturing agents (4). Alternatively, protein tagging is commonly used. Common tags include the FLAG-tag, which allows for specific elution with the FLAG peptide (41), and the His-tag, which allows for stringent purification of complexes under denaturing conditions. Denaturation was reported to reduce the background (25) but also reduced the recovery of actual interactions (45). The tandem tags GFP-Tev-FLAG (56) and ProteinA-TEV-His (34) have also been used to achieve higher sample purity. After protein pull-down, proximity ligation and other enzymatic reactions are usually performed on beads when RNA is still attached to the protein, which allows for efficient washes of samples and quick replacement of reaction mixtures.

Experiments aiming to recover the entire RNA interactome (2, 39, 53) do not require a targeted RNA enrichment step (although these techniques employ RNA duplex enrichment following proximity ligation, as described below). After cross-linking RNA duplexes with psoralen, total RNA is extracted with phenol-based reagents. Alternatively, MARIO (48) extracts all protein-bound RNAs simultaneously through global protein modification with biotin (EZ-Link Iodoacetyl-PEG2-Biotin), followed by streptavidin pull-down of proteins with bound RNA duplexes. To increase coverage of non-ribosomal RNAs, some methods additionally include an rRNA depletion step (48, 53).

3.3. RNase Treatment and Proximity Ligation

RNA fragmentation prior to proximity ligation is important and needs to be optimized carefully. The method of RNA fragmentation may influence the efficiency of ligation because of intrinsic nucleotide preferences of RNA ligases and because ligation requires single-stranded nucleotides at the end of each duplex. The hiCLIP (63) and MARIO (48) methods facilitate the ligation of fully base-paired regions by introducing a linker between the ligated fragments. In addition, fragments that are too long may promote background ligation between different RNA duplexes, reduce the resolution of mapped interactions, and require longer sequencing reads to cover the entire chimeric RNA, whereas fragments that are too short may lead to problems with unambiguous identification of interacting RNA molecules during bioinformatic analysis.

RNA fragmentation typically relies on treatment with RNases, some of which can be inactivated by incubation with inhibitors, while others must be washed off before ligation. Some protocols apply chemical fragmentation using Mg^{2+} ions at high temperature (2, 58), which allows for quick quenching by lowering the reaction temperature. Fragmentation of RNA by sonication is a possible alternative and has been used in efforts to map RNA interactions without the use of proximity ligation, though the resulting fragment sizes are longer than those obtained with other methods. The choice of fragmentation method determines the order of the subsequent steps to prepare RNA 5′ and 3′ ends for proximity ligation.

The most common fragmentation methods (RNase A, RNase T1, RNase I, or chemical fragmentation) generate 5′-hydroxyl and 3′-phosphate or 2′3′-cyclic phosphate RNA termini. Because these termini are incompatible with standard ligation, they must be regenerated by hydrolysis of 3′-phosphate with alkaline phosphatase to 3′-hydroxyl and phosphorylation of 5′-hydroxyl with polynucleotide kinase to 5′-phosphate. The 5′-phosphate and 3′-hydroxyl ends are compatible with proximity ligation catalyzed by T4 RNA ligase I. Omission of the RNA–RNA ligation step reduces the number of chimeras, but chimeras can still be recovered, either because they are formed during linker ligation steps (34) or because of endogenous ligase activity (21, 45).

Because proximity ligation is crucial but inefficient, there is great interest in optimizing this step, and various groups have proposed alternative solutions. One is to use RNases that leave

ligation-compatible ends (RNase H, RNase III, or nuclease S1); unfortunately, these RNases have specific properties that limit their applicability to narrow subsets of RNA. Other possibilities include using two steps of polynucleotide kinase treatment for both the 3′-hydroxyl and 5′-phosphate regeneration steps (2, 41, 45), thus increasing the chance of ligation for molecules that contained a 2′3′-cyclic phosphate; using the tRNA ligase AtRNL instead of T4 RNA ligase I, which allows for ligation of the original ends after standard RNase treatment and thus limits the number of required enzymatic steps (32); performing RNA–RNA ligation using Lucigen's thermostable CircLigase at a high temperature (60°C) when RNA structures are more relaxed, potentially increasing ligation efficiency (53); and ligating an additional RNA linker between the two strands of the RNA duplex (48, 63). The use of the linker allows for better control over the ligation reaction and helps to unambiguously define interacting RNA molecules during bioinformatic analysis.

3.4. Preparation of Sequencing Libraries

Following proximity ligation, there are two ways to convert chimeric RNA molecules into a cDNA sequencing library: (*a*) the CLIP approach, which includes two independent ligations of sequencing adapters to the 5′ and 3′ ends of RNA, followed by reverse transcription and PCR amplification of the library (25, 26), and (*b*) the iCLIP approach, which includes ligation of a 3′ linker followed by reverse transcription, circularization of cDNA using Lucigen's CircLigase II (which is reportedly more efficient than using an RNA ligase), linearization, and PCR amplification of the library (63). Both strategies have been used extensively, and there is no reason to think that one is better than the other. Alternatively, a commercial kit for cDNA library preparation can be used, such as Takara Bio's SMARTer smRNA-Seq Kit (43) or Lexogen's QuantSeq 3′ mRNA-Seq Library Prep Kit REV (58).

3.5. Chimera Enrichment

Because proximity ligation is inefficient, only a minority of RNA duplexes are ligated among a large background of nonligated single RNAs. To increase the coverage of chimeras, several strategies have been developed for the selection of chimeric molecules from nonchimeric molecules.

Psoralen cross-linking gives an opportunity to directly select cross-linked RNA duplexes. In PARIS, slower-migrating cross-linked RNAs are separated from unmodified RNAs using two-dimensional gel electrophoresis (39); in LIGR-seq, non-cross-linked RNAs are depleted with the exonuclease RNase R (53); and in SPLASH (2) and COMRADES (78), cross-linked RNA duplexes are associated with biotin-modified psoralen, allowing streptavidin pull-down of duplexes. It is also possible to enrich a specific subset of chimeras: In ChimP (chimera PCR), specific miRNA–target interactions are detected by PCR with a pair of sequence-specific primers (7). Similarly, it should be possible to amplify a subset of chimeras by using one sequence-specific primer and one primer matching the sequencing adapter.

However, cross-linked fragments are not necessarily ligated, and additional methods have been used to enrich ligated RNA–RNA hybrids. The easiest and most common approach relies on size selection of RNA or cDNA fragments based on their migration in an agarose or polyacrylamide gel. This selection can be performed at various stages of the protocol: after RNase fragmentation (58), after reverse transcription (63), or (in most methods) after PCR amplification of the library; some methods use more than one size selection step. As an alternative to gel electrophoresis, size selection on AMPure beads has been used (7). In all cases, longer fragments are preferentially selected for sequencing because they are deemed to be enriched for chimeras, as confirmed by

bioinformatic analysis of CLASH data (68). By analogy to Hi-C experiments, duplexes might also be enriched using a biotinylated linker ligated between the two strands of the duplex.

4. BIOINFORMATIC METHODS

4.1. Calling Chimeras

The first step in the bioinformatic analysis of RNA proximity ligation experiments is the identification (calling) of chimeric reads. This can be done with general-purpose mapping programs [e.g., STAR (Spliced Transcripts Alignment to a Reference) (12) or TopHat (67)] using settings originally designed for the detection of spliced reads, while taking care to allow noncanonical splice junctions (since proximity ligation can link arbitrary sequences) (49). If paired-end sequencing is used, chimeras can simply be identified as instances where both reads from a pair are mapped to distinct positions along the genome or transcriptome (41), though such an approach might miss a fraction of local interactions in which the reads map close to each other. Alternatively, dedicated pipelines for calling chimeras have been developed, such as hyb (68), ChimeraTie (43), Aligater (53), MARIO tools (48), and CLAN (Cross-Linked Reads Analysis) (76). Dedicated chimera-mapping programs can improve the sensitivity and specificity of chimera detection, facilitate the manipulation of relevant mapping parameters, and provide some level of downstream analysis.

When mapping to the genome, many chimeric reads are expected to represent splice junctions rather than proximity ligation products. To avoid these, reads can be mapped against a spliced transcriptome database (25, 49) or mapped to the genome followed by filtering of known splice junctions (39). To avoid the mapping of reads containing homopolymeric repeats, reads can be filtered as a preprocessing step or during mapping—for example, using BLAST (Basic Local Alignment Search Tool), which includes homopolymer filtering by default.

On some occasions, the two fragments of a chimeric read may not encompass the entire base-paired region, either because of excessive RNase digestion during sample preparation or because the sequencing read does not cover the entire cDNA. In such cases, reads should be bioinformatically extended in either direction (25, 56). Failure to extend the reads may result in underestimation of association strength.

UV and psoralen cross-linking are known to introduce chemical changes in the RNA, which can result in nucleotide substitutions and short deletions when reverse transcribed into cDNA. These mutations, which arise during sequencing library preparation, can lead to errors in RNA folding prediction. To avoid such artifacts, read sequences can be substituted with their corresponding transcript sequences taken from the genome or transcriptome database, prior to downstream analysis of chimeras (68).

4.2. Calling Interactions

Following the calling of chimeras, the next step is to integrate the information from multiple reads to call candidate RNA–RNA interactions. One way to do this is to identify clusters of chimeric reads, also known as duplex groups (39). This approach is similar to the clustering of reads in a CLIP experiment, except that the clusters are bipartite, and both fragments of a chimera are required to overlap with either part of the cluster.

Unfortunately, the clustering approach breaks down with increasing sequencing depth, because separate clusters are merged together, and in extreme cases they can cover the entire transcript. To overcome this problem, interactions can be identified from a two-dimensional contact probability map as regions with high coverage of chimeras or a high density of ligation junctions (49, 78). To account for PCR and ligation biases that can potentially inflate chimera coverage in specific

regions of the transcript, the contact map can be normalized using approaches borrowed from the Hi-C analysis tool kit (49).

After calling, the interactions must be scored to evaluate their statistical and/or biological significance. Interactions supported by larger numbers of chimeric reads are often considered more reliable, though one study removed the most frequent ligation junctions because it was hypothesized that they represented deletions or unannotated splice junctions (49). Some studies have used ad hoc scoring methods based on a combination of coverage, reproducibility between replicate experiments, identification of chimeras in forward and reverse orientation, and folding energy, as well as criteria specific to the RNA under study (25, 34, 74). For example, box C/D snoRNAs are known to base-pair with target RNA with their guide region, and this property can be used to recover interactions likely to guide target methylation. Because read coverage is expected to increase proportionally to the abundance of each interaction partner, the relative abundance of interaction partners and chimeras can be used to score interactions, and the statistical enrichment of chimeras can be quantified by Fisher's exact test or other statistical methods (41).

4.3. Resolving RNA Structures

The task of inferring RNA secondary structures (i.e., intra- or intermolecular base-pairing patterns) from proximity ligation data ranges from straightforward to potentially unsolvable. Among the simple cases are chimeras between small RNAs (miRNAs, piRNAs, and sRNAs) and their targets, which can typically be folded in silico in an unambiguous way and are often supported by known rules of interaction, such as miRNA seed base-pairing. On the other hand, proximity ligation of large RNAs, such as the genomes of RNA viruses, reveals complicated patterns of mutually incompatible local and long-range interactions, consistent with the notion that RNA exists as a complex structural ensemble rather than a unique, well-defined structure. Bioinformatic methods to recover these structural ensembles from proximity ligation data have only just begun to emerge (28, 35, 78).

An especially difficult problem concerns the phasing of RNA structural elements—that is, deciding which individual interactions coexist within one RNA molecule. This matters, for example, for the discovery of pseudoknots: Proximity ligation may indicate a pattern of intersecting base pairs consistent with the presence of a pseudoknot, but without additional information it is impossible to determine whether the base pairs originated from the same molecule (indicating a pseudoknot) or from different molecules (indicating alternative conformations). The phasing problem might be resolved by performing proximity ligation in conditions that favor a particular RNA conformation or by incorporating additional data, such as information from correlated chemical probing (27, 47) or long-read sequencing strategies (e.g., nanopore sequencing) (18).

4.4. Integrated Analysis for Determining RNA Structures

While RNA proximity ligation is useful for determining pairwise RNA interactions, the low chimeric ratio, relatively sparse data, and non-single-nucleotide resolution can result in challenges in accurate modeling of RNA structures. Theoretically, pairwise interaction information could be integrated with other single-nucleotide structure-probing strategies, such as SHAPE-MaP (selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling) or icSHAPE (in vivo click selective 2-hydroxyl acylation and profiling experiment) (59, 61), to create a more accurate structure model; however, doing so is nontrivial in practice. In modeling of dengue and Zika genome structures, SHAPE-MaP data have been used to generate pseudoenergies and create an ensemble of structure conformations, and structures that best fit high-confidence SPLASH

interactions were then selected as the most probable conformation (28). Alternatively, the degree of fit, F, between a particular candidate structure and an entire proximity ligation data set can be calculated as

$$F = \sum_{i,j} b_{ij} \times C_{ij},$$

where b_{ij} equals 1 if bases (i,j) are paired in the candidate structure and 0 otherwise, and C_{ij} is the number of reads in the proximity ligation data set that are consistent with base-pairing between i and j. In one analysis, a proximity ligation read was deemed consistent with base pair (i,j) if that base pair was included in the predicted minimum folding energy structure for that read (78).

4.5. Data Representation

As technologies mature, the research community gradually agrees on standard ways to share and visualize data: RNA-seq results are stored as binary alignment map (BAM) files, structures are stored as Protein Data Bank (PDB) files, and Hi-C is visualized as topologically associating domain (TAD) diagrams. Standard data formats facilitate both the reuse of data and the interpretation of results. Alas, there is not yet a widely accepted format for sharing RNA proximity ligation data. Until a consensus emerges, it is essential to share both the raw sequencing data and well-annotated read-level mapping data, together with unambiguous identifiers of reference sequences. Appropriate formats include BAM, browser extensible data (BED), and hyb (68). The most common ways to visualize RNA proximity ligation results are shown in **Figure 2**.

5. DID MY EXPERIMENT WORK?

You have completed your RNA proximity ligation experiment and obtained sequencing data. How can you tell if your experiment worked?

5.1. Proportion of Chimeric Reads

Depending on the experimental protocol, 0.5–10% of the reads are normally expected to be chimeras (25, 34, 39, 56, 63, 78), though some studies have reported as many as 30% (41, 48). Fewer chimeras could indicate a problem with the experiment (RNA–RNA ligation), analysis (which can be checked by testing the analysis pipeline on a previously published data set), or sequencing (are the reads long enough to recover both fragments of chimeras?). If the experiment includes a negative control (e.g., without ligase or without cross-linking), there should be a clear difference in the proportion of chimeras between the experiment and control.

5.2. Comparison with Known RNA Structures

Benchmarking against known structures should be used wherever possible to validate the results. rRNA is the most abundant species of RNA in a cell, and the three-dimensional structures of prokaryotic and eukaryotic ribosomes are known. Multiple RNA proximity ligation papers have reported a correspondence between the location of chimeric reads and the structure of the ribosome: 50–80% of reproducible chimeras agreed with known interactions within large and small ribosomal subunits (25, 49, 63). A cross-linking-free study showed statistically significant enrichment of chimeras between physically proximal fragments of the ribosome (43). Some of the experimental protocols include an rRNA depletion step, which increases the coverage of

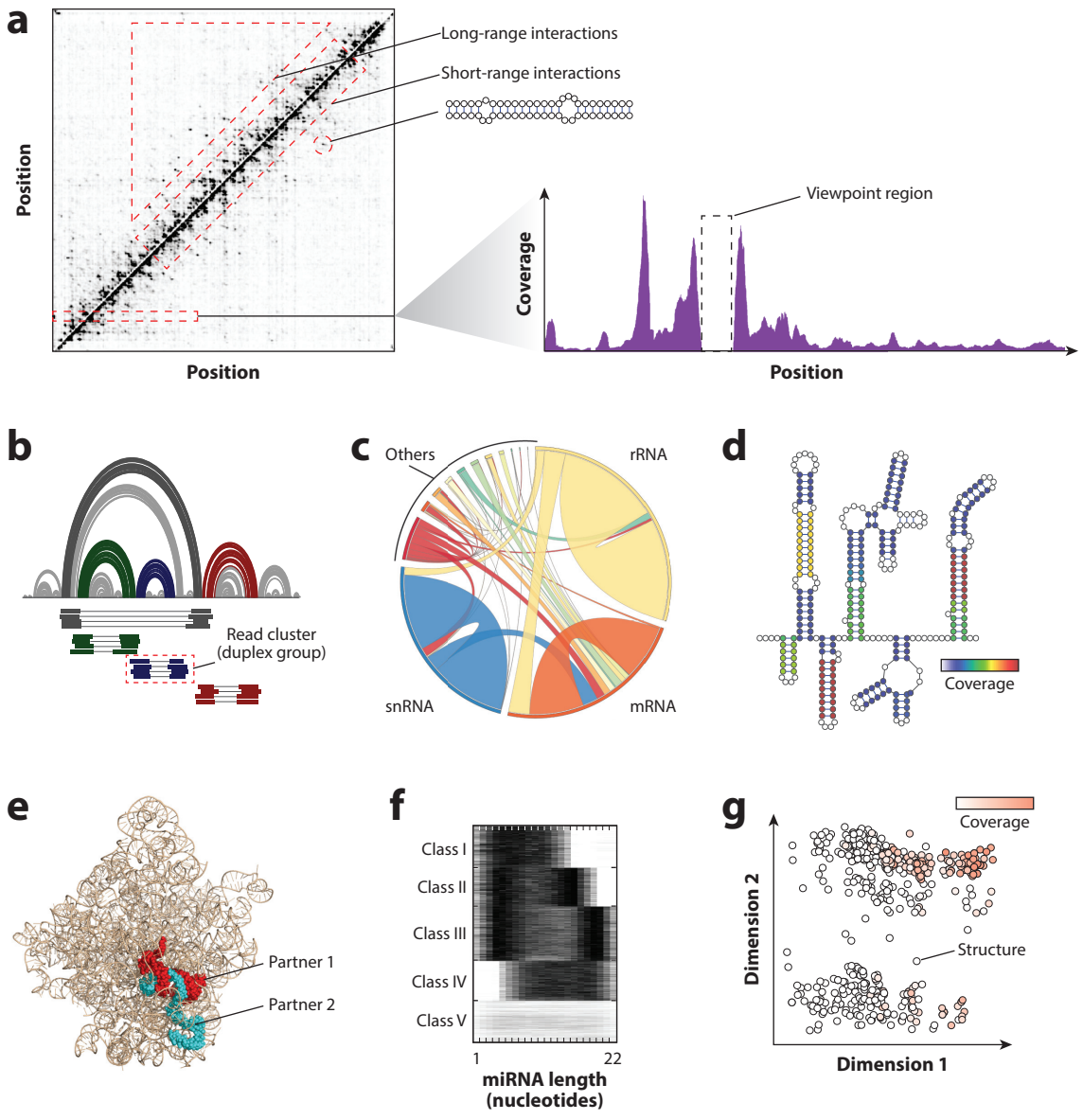

Figure 2

Visualization of RNA proximity ligation data. The ensemble of intramolecular interactions in a specific transcript can be shown as a two-dimensional contact density map (panel *a*, *left*). The positions of the first and second fragments of each chimera are shown on the *x* and *y* axes, respectively, so that 5′–3′ chimeras are above the diagonal and 3′–5′ chimeras are below it. To highlight interactions involving a particular position along a transcript, a viewpoint graph can be used (panel *a*, *right*). Alternatively, interactions within and between transcripts can be displayed in a genome browser as arcs and duplex groups, allowing the integration of interaction data with additional annotations (panel *b*), or as a Circos plot, emphasizing interactions between multiple transcripts (panel *c*). Diagrams of consensus RNA secondary structures can be displayed with a color code that represents support for individual base pairs (panel *d*), and interactions can be overlaid onto known secondary or tertiary structures (panel *e*). Bespoke graphs have been used to visualize the ensemble of miRNA–target interactions (panel *f*) and the ensemble of conformations of a viral genome fragment (panel *g*). Abbreviations: miRNA, microRNA; mRNA, messenger RNA; rRNA, ribosomal RNA; snRNA, short nuclear RNA. Left side of panel *a* and panel *g* adapted from Reference 78; panel *c* adapted from Reference 53; panel *e* adapted from Reference 34; panel *f* adapted from Reference 25.

non-ribosomal reads but makes it difficult to benchmark the data against the ribosome structure. In these cases, it may be possible to benchmark the data against other abundant structural RNAs, such as snRNAs or snoRNAs (53), or against sets of known or predicted interactions derived from alternative sources.

5.3. Cross-Species Controls

Typically, 1–10% of interspecies chimeras have been found, even when the experimental sample was mixed with an excess of RNA from a different species (2, 25, 45, 48). In one experiment, 14–15% of chimeras were cross-species with UV cross-linking and 19–25% without cross-linking, but these proportions dropped to 0.8–3.8% when only statistically significant interactions were considered (41). The background can also be estimated by quantifying interactions between same-species transcripts, which are unlikely to interact in physiological conditions. For example, an experiment performed in living *C. elegans* showed that germline piRNAs were nearly 10 times more likely to form chimeras with germline mRNAs than with somatic mRNAs, indicating that few interactions were formed during sample preparation after lysis (56). In a psoralen cross-linking experiment, few intermolecular chimeras between cytoplasmic and mitochondrial ribosomes were found, even though intramolecular chimeras were abundant (78).

5.4. Folding Energy of Chimeras

If the chimeras represent genuine Watson–Crick interactions, they should show a strong base-pairing propensity when analyzed with an RNA folding program. Many experiments have compared the predicted minimum folding energy distributions of chimeras against the background of randomly reassigned chimera fragments, against fragments with randomly shuffled bases, or against RNA fragments drawn from random positions in the transcriptome (25, 34, 39, 48, 63). The interaction energies of chimeric reads were typically much stronger than the controls, strongly supporting the notion that chimeras were formed by ligation of base-paired RNA fragments. Folding energy has been shown to correlate with read coverage, perhaps reflecting the lower propensity of stable chimeras to dissociate during sample preparation. Some proximity ligation protocols aim to recover indirect interactions mediated through proteins (43). In such cases, chimeras are expected to show little or no skew toward low folding energy.

5.5. Orientation of Chimeras

In almost all reported RNA proximity ligation studies, intramolecular chimeras were recovered in the 5′–3′ orientation (also known as in-line, forward, regularly gapped, etc.) and 3′–5′ orientation (also known as inverted, reverse, chiastic, etc.). A strong correspondence between coverage in 5′–3′ and 3′–5′ orientations is often expected and can be inspected visually on a coverage heat-map graph (**Figure 2**) or by correlating the coverage in the two orientations. In some cases, the counts of chimeras in the two orientations are expected to be different. For example, the structure of the RNA-induced silencing complex, composed of an Ago protein, a miRNA, and a target RNA, makes the 3′ end of the miRNA more accessible for RNA–RNA ligation than the 5′ end. As a result, most chimeras are found in the 5′–miRNA–mRNA–3′ orientation rather than the 5′–mRNA–miRNA–3′ orientation (25, 45). Similarly, in an Hfq pull-down experiment, most chimeras were ligated in the 5′–mRNA–sRNA–3′ orientation, consistent with the 3′ end of sRNA being unavailable for ligation (41). In the RIPPLiT exon junction complex pull-down study, short- and long-range 5′–3′ and 3′–5′ chimeras were found in the sample treated with

ligase, whereas only short-range 5′–3′ chimeras were found in the no-ligase control (43). Such chimeras were deemed to be artifacts, and the authors filtered them by optimizing Bowtie2's read and reference gap penalties to disfavor short fragment junction gaps and by disallowing multimapping fragments. Some chimera mapping strategies implicitly filter away short-range intramolecular 5′–3′ chimeras, since such chimeras are called as a gapped nonchimeric read (68).

6. FUTURE DIRECTIONS

The field of RNA proximity ligation has been developing in parallel to similar studies of DNA, with little interaction between the two. The DNA field appears to be several years ahead: The first application of DNA proximity ligation, known as chromosome conformation capture (3C), was published in 2002 (11), and the most popular genome-wide method, Hi-C, was developed in 2009 (36). Since then, the chromosome conformation capture field has been growing steadily, with more than 100 papers published in 2018 alone and entire conferences dedicated to the subject. An analysis of recent trends in DNA proximity ligation might therefore provide insights into future developments in RNA proximity ligation.

In recent Hi-C studies, researchers were excited to document dynamic changes of genome conformation in mitosis (19), spermatogenesis (70), oocyte-to-zygote transition (14), and differentiation (44, 66). There is ample evidence that DNA conformations are context dependent and functionally important, and the same holds true for RNA; examples include antiterminator sequences in bacterial RNAs, RNA thermometers, and natural RNA aptamers, and high-throughput probing shows condition-dependent remodeling of RNA structures (2, 64, 72). However, few proximity ligation studies have addressed the structural dynamics of RNA. In bacteria, one study showed that the network of interactions between sRNAs and mRNAs depends on growth conditions (41), and SPLASH experiments in human embryonic stem cells and differentiated cells showed that RNA interaction networks can be extensively remodeled upon changes in cellular states (2). These changes were driven both by differences in transcript expression levels and by rewiring of interactions between existing transcripts, such as condition-specific binding of an sRNA to a sponge RNA. By identifying long-range and intermolecular interactions, future RNA proximity ligation studies will complement structural probing and help elucidate RNA interactomes across physiological conditions, subcellular compartments, tissues, organisms, and species.

Chromosome conformation also depends on genetic mutations: Depletion of the CTCF protein or removal of CTCF binding sites ablates boundaries between TADs, while deletions of specific transcription factors prevent associations between chromosomes in mouse olfactory neurons (44). Similarly, RNA structures and interaction networks are influenced by gene deletions (42), single-nucleotide mutations (73), and RNA modifications such as adenosine or cytosine methylation, adenosine deamination, or ribose 2′-*O*-methylation (33, 37, 51). To our knowledge, proximity ligation has yet to be applied to study the effects of genetic or epigenetic changes on the RNA interactome.

The combination of chromosome conformation capture with mathematical modeling has given insights into the three-dimensional organization of the genome. Analysis of correlations within chromosome interaction density maps revealed the partitioning of open and closed chromatin into two genome-wide compartments, and analysis of interaction probabilities as a function of linear distance along the chromosome suggested a mode of genome organization known as a fractal globule (36). A type of metagene plot known as aggregate peak analysis revealed the association of chromosome contacts with recurrent genomic features (44, 50). Although the principles of structural organization differ between DNA and RNA (for example, modular organization into TADs seems to be absent in RNA), the mathematical methods developed to study DNA are likely to

be applicable to RNA. An important challenge lies in the understanding of the functional consequences of RNA and DNA conformations. Unlike higher-order genome organization, many long-range RNA–RNA interactions are mediated by base-pairing, which makes it possible to validate these interactions by mutagenesis or analysis of sequence conservation. Indeed, such methods have already allowed the validation of structural elements detected in mammalian and viral genomes (28, 35, 39).

RNA molecules have been described as "screaming and kicking" (17, p. 474). As an RNA molecule is born, spliced, transported, translated, and degraded, it interacts with numerous different host proteins and RNAs that change its shape. As such, RNA structural populations inside cells are likely to exist as structural ensembles, and these dynamic changes could be observed in RNA–RNA interaction networks. Predicted shifts in structural ensembles due to mutations in mRNA untranslated regions have been associated with disease (23), and genome-wide experiments studying synonymous mutations show an association of RNA structure with cancer (54). Along with comparing RNA structures in normal and disease states, further experiments involving probing structure dynamics under different cellular conditions will provide a fuller picture of the complexity of structural dynamics in the cell. In addition, current proximity ligation sequencing strategies still rely on pairwise interactions, making it difficult to connect RNA structures along the full length of the transcript. With additional technology developments that couple long-read sequencing with shorter proximity ligation information, we could potentially stitch RNA structures together. The field of RNA proximity ligation is rapidly progressing, and new technologies and additional cellular systems could substantially advance our understanding of the full complexity of RNA-based regulation in the cell.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

G.K. was supported by Wellcome Trust fellowship 207507 and Medical Research Council grant MC_UU_00007/12. Y.W. acknowledges funding from an A*STAR Investigatorship grant and the National Research Foundation (NRF2019-NRF-ISF003-2970 and CRP21-2018-0101). Y.W. is a CIFAR Azrieli Global Scholar, and Y.W. and G.K. are EMBO Young Investigators.

LITERATURE CITED

1. Agarwal V, Bell GW, Nam JW, Bartel DP. 2015. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* 4:e05005
2. Aw JG, Shen Y, Wilm A, Sun M, Lim XN, et al. 2016. In vivo mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol. Cell* 62:603–17
3. Becker WR, Ober-Reynolds B, Jouravleva K, Jolly SM, Zamore PD, Greenleaf WJ. 2019. High-throughput analysis reveals rules for target RNA binding and cleavage by AGO2. *Mol. Cell* 75:741–55.e11
4. Bharathavikru R, Dudnakova T, Aitken S, Slight J, Artibani M, et al. 2017. Transcription factor Wilms' tumor 1 regulates developmental RNAs through 3′ UTR interaction. *Genes Dev.* 31:347–52
5. Bohnsack MT, Martin R, Granneman S, Ruprecht M, Schleiff E, Tollervey D. 2009. Prp43 bound at different sites on the pre-rRNA performs distinct functions in ribosome synthesis. *Mol. Cell* 36:583–92
6. Broughton JP, Lovci MT, Huang JL, Yeo GW, Pasquinelli AE. 2016. Pairing beyond the seed supports microRNA targeting specificity. *Mol. Cell* 64:320–33

7. Broughton JP, Pasquinelli AE. 2018. Detection of microRNA-target interactions by chimera PCR (ChimP). *Methods Mol. Biol.* 1823:153–65
8. Bullard WL, Kara M, Gay LA, Sethuraman S, Wang Y, et al. 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
9. Burger K, Muhl B, Kellner M, Rohrmoser M, Gruber-Eber A, et al. 2013. 4-Thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. *RNA Biol*. 10:1623–30
10. Dadonaite B, Gilbertson B, Knight ML, Trifkovic S, Rockman S, et al. 2019. The structure of the influenza A virus genome. *Nat. Microbiol.* 4:1781–89
11. Dekker J, Rippe K, Dekker M, Kleckner N. 2002. Capturing chromosome conformation. *Science* 295:1306–11
12. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, et al. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21
13. Engreitz JM, Sirokman K, McDonel P, Shishkin AA, Surka C, et al. 2014. RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent pre-mRNAs and chromatin sites. *Cell* 159:188–99
14. Flyamer IM, Gassler J, Imakaev M, Brandao HB, Ulianov SV, et al. 2017. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* 544:110–14
15. Friedländer MR, Lizano E, Houben AJ, Bezdan D, Báñez-Coronel M, et al. 2014. Evidence for the biogenesis of more than 1,000 novel human microRNAs. *Genome Biol*. 15:R57
16. Galej WP, Wilkinson ME, Fica SM, Oubridge C, Newman AJ, Nagai K. 2016. Cryo-EM structure of the spliceosome immediately after branching. *Nature* 537:197–201
17. Ganser LR, Kelly ML, Herschlag D, Al-Hashimi HM. 2019. The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell Biol.* 20:474–89
18. Garalde DR, Snell EA, Jachimowicz D, Sipos B, Lloyd JH, et al. 2018. Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* 15:201–6
19. Gibcus JH, Samejima K, Goloborodko A, Samejima I, Naumova N, et al. 2018. A pathway for mitotic chromosome formation. *Science* 359:eaao6135
20. Granneman S, Kudla G, Petfalski E, Tollervey D. 2009. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *PNAS* 106:9613–18
21. Grosswendt S, Filipchyk A, Manzano M, Klironomos F, Schilling M, et al. 2014. Unambiguous identification of miRNA:target site interactions by different types of ligation reactions. *Mol. Cell* 54:1042–54
22. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141:129–41
23. Halvorsen M, Martin JS, Broadaway S, Laederach A. 2010. Disease-associated mutations that alter the RNA structural ensemble. *PLOS Genet*. 6:e1001074
24. Hearst JE. 1981. Psoralen photochemistry and nucleic acid structure. *J. Investig. Dermatol.* 77:39–44
25. Helwak A, Kudla G, Dudnakova T, Tollervey D. 2013. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* 153:654–65
26. Helwak A, Tollervey D. 2014. Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH). *Nat. Protoc.* 9:711–28
27. Homan PJ, Favorov OV, Lavender CA, Kursun O, Ge X, et al. 2014. Single-molecule correlated chemical probing of RNA. *PNAS* 111:13858–63
28. Huber RG, Lim XN, Ng WC, Sim AYL, Poh HX, et al. 2019. Structure mapping of dengue and Zika viruses reveals functional long-range interactions. *Nat. Commun.* 10:1408
29. Jamonnak N, Creamer TJ, Darby MM, Schaughency P, Wheelan SJ, Corden JL. 2011. Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. *RNA* 17:2011–25
30. Jungkamp AC, Stoeckius M, Mecenas D, Grun D, Mastrobuoni G, et al. 2011. In vivo and transcriptome-wide identification of RNA binding protein target sites. *Mol. Cell* 44:828–40
31. Kim SH, Suddath FL, Quigley GJ, McPherson A, Sussman JL, et al. 1974. Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science* 185:435–40

32. Kirsch R, Olzog VJ, Bonin S, Weinberg CE, Betat H, et al. 2019. A streamlined protocol for the detection of mRNA–sRNA interactions using AMT-crosslinking in vitro. *Biotechniques* 67:178–83
33. Kiss T. 2001. Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.* 20:3617–22
34. Kudla G, Granneman S, Hahn D, Beggs JD, Tollervey D. 2011. Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *PNAS* 108:10010–15
35. Li P, Wei Y, Mei M, Tang L, Sun L, et al. 2018. Integrative analysis of Zika virus genome RNA structure reveals critical determinants of viral infectivity. *Cell Host Microbe* 24:875–86.e5
36. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326:289–93
37. Liu N, Zhou KI, Parisien M, Dai Q, Diatchenko L, Pan T. 2017. N^6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. *Nucleic Acids Res.* 45:6051–63
38. Liu T, Zhang K, Xu S, Wang Z, Fu H, et al. 2017. Detecting RNA-RNA interactions in *E. coli* using a modified CLASH method. *BMC Genom.* 3:343
39. Lu Z, Zhang QC, Lee B, Flynn RA, Smith MA, et al. 2016. RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* 165:1267–79
40. Maenner S, Blaud M, Fouillen L, Savoye A, Marchand V, et al. 2010. 2-D structure of the A region of Xist RNA and its implication for PRC2 association. *PLOS Biol.* 8:e1000276
41. Melamed S, Peer A, Faigenbaum-Romm R, Gatt YE, Reiss N, et al. 2016. Global mapping of small RNA-target interactions in bacteria. *Mol. Cell* 63:884–97
42. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, et al. 2013. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495:333–38
43. Metkar M, Ozadam H, Lajoie BR, Imakaev M, Mirny LA, et al. 2018. Higher-order organization principles of pre-translational mRNPs. *Mol. Cell* 72:715–26.e3
44. Monahan K, Horta A, Lomvardas S. 2019. LHX2- and LDB1-mediated *trans* interactions regulate olfactory receptor choice. *Nature* 565:448–53
45. Moore MJ, Scheel TK, Luna JM, Park CY, Fak JJ, et al. 2015. miRNA-target chimeras reveal miRNA 3′-end pairing as a major determinant of Argonaute target specificity. *Nat. Commun.* 6:8864
46. Morf J, Wingett SW, Farabella I, Cairns J, Furlan-Magaril M, et al. 2019. RNA proximity sequencing reveals the spatial organization of the transcriptome in the nucleus. *Nat. Biotechnol.* 37:793–802
47. Mustoe AM, Lama NN, Irving PS, Olson SW, Weeks KM. 2019. RNA base-pairing complexity in living cells visualized by correlated chemical probing. *PNAS* 116:24574–82
48. Nguyen TC, Cao X, Yu P, Xiao S, Lu J, et al. 2016. Mapping RNA-RNA interactome and RNA structure in vivo by MARIO. *Nat. Commun.* 7:12023
49. Ramani V, Qiu R, Shendure J. 2015. High-throughput determination of RNA structure by proximity ligation. *Nat. Biotechnol.* 33:980–84
50. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159:1665–80
51. Ries RJ, Zaccara S, Klein P, Olarerin-George A, Namkoong S, et al. 2019. m^6A enhances the phase separation potential of mRNA. *Nature* 571:424–28
52. Rivas E, Clements J, Eddy SR. 2017. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat. Methods* 14:45–48
53. Sharma E, Sterne-Weiler T, O'Hanlon D, Blencowe BJ. 2016. Global mapping of human RNA-RNA interactions. *Mol. Cell* 62:618–26
54. Sharma Y, Miladi M, Dukare S, Boulay K, Caudron-Herger M, et al. 2019. A pan-cancer analysis of synonymous mutations. *Nat. Commun.* 10:2569
55. Shchepachev V, Bresson S, Spanos C, Petfalski E, Fischer L, et al. 2019. Defining the RNA interactome by total RNA-associated protein purification. *Mol. Syst. Biol.* 15:e8689
56. Shen EZ, Chen H, Ozturk AR, Tu S, Shirayama M, et al. 2018. Identification of piRNA binding sites reveals the Argonaute regulatory landscape of the *C. elegans* germline. *Cell* 172:937–51.e18
57. Sheu-Gruttadauria J, Xiao Y, Gebert LF, MacRae IJ. 2019. Beyond the seed: structural basis for supplementary microRNA targeting by human Argonaute2. *EMBO J.* 38:e101153

58. Shi J, Deng Y, Huang S, Huang C, Wang J, et al. 2019. Suboptimal RNA-RNA interaction limits U1 snRNP inhibition of canonical mRNA 3′ processing. *RNA Biol.* 16:1448–60
59. Siegfried NA, Busan S, Rice GM, Nelson JA, Weeks KM. 2014. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* 11:959–65
60. Smith KC, Meun DH. 1968. Kinetics of the photochemical addition of cysteine-^{35}S to polynucleotides and nucleic acids. *Biochemistry* 7:1033–37
61. Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, et al. 2015. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* 519:486–90
62. Strobel EJ, Yu AM, Lucks JB. 2018. High-throughput determination of RNA structures. *Nat. Rev. Genet.* 19:615–34
63. Sugimoto Y, Vigilante A, Darbo E, Zirra A, Militti C, et al. 2015. hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. *Nature* 519:491–94
64. Sun L, Fazal FM, Li P, Broughton JP, Lee B, et al. 2019. RNA structure maps across mammalian cellular compartments. *Nat. Struct. Mol. Biol.* 26:322–30
65. Sussman JL, Holbrook SR, Warrant RW, Church GM, Kim SH. 1978. Crystal structure of yeast phenylalanine transfer RNA. I. Crystallographic refinement. *J. Mol. Biol.* 123:607–30
66. Tan L, Xing D, Daley N, Xie XS. 2019. Three-dimensional genome structures of single sensory neurons in mouse visual and olfactory systems. *Nat. Struct. Mol. Biol.* 26:297–307
67. Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25:1105–11
68. Travis AJ, Moody J, Helwak A, Tollervey D, Kudla G. 2014. Hyb: a bioinformatics pipeline for the analysis of CLASH (crosslinking, ligation and sequencing of hybrids) data. *Methods* 65:263–73
69. Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. 2003. CLIP identifies Nova-regulated RNA networks in the brain. *Science* 302:1212–15
70. Vara C, Paytuvi-Gallart A, Cuartero Y, Le Dily F, Garcia F, et al. 2019. Three-dimensional genomic structure and cohesin occupancy correlate with transcriptional activity during spermatogenesis. *Cell Rep.* 28:352–67.e9
71. Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY. 2011. Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.* 12:641–55
72. Wan Y, Qu K, Ouyang Z, Kertesz M, Li J, et al. 2012. Genome-wide measurement of RNA folding energies. *Mol. Cell* 48:169–81
73. Wan Y, Qu K, Zhang QC, Flynn RA, Manor O, et al. 2014. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505:706–9
74. Waters SA, McAteer SP, Kudla G, Pang I, Deshpande NP, et al. 2016. Small RNA interactome of pathogenic *E. coli* revealed through crosslinking of RNase E. *EMBO J.* 36:374–87
75. Yang A, Bofill-De Ros X, Shao TJ, Jiang M, Li K, et al. 2019. 3′ uridylation confers miRNAs with non-canonical target repertoires. *Mol. Cell* 75:511–22.e4
76. Zhong C, Zhang S. 2019. Accurate and efficient mapping of the cross-linked microRNA-mRNA duplex reads. *iScience* 18:11–19
77. Zisoulis DG, Lovci MT, Wilbert ML, Hutt KR, Liang TY, et al. 2010. Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*. *Nat. Struct. Mol. Biol.* 17:173–79
78. Ziv O, Gabryelska MM, Lun ATL, Gebert LFR, Sheu-Gruttadauria J, et al. 2018. COMRADES determines in vivo RNA structures and interactions. *Nat. Methods* 15:785–88