

Review Article

Zooming in on protein–RNA interactions: a multi-level workflow to identify interaction partners

Alessio Colantoni^{1,*}, Jakob Rupert^{2,3,*}, Andrea Vandelli^{4,5,6,*},  Gian Gaetano Tartaglia^{1,2,3,4,5,7} and Elsa Zacco³

¹Center for Life Nanoscience, Istituto Italiano di Tecnologia, Viale Regina Elena 291, 00161 Rome, Italy; ²Department of Biology ‘Charles Darwin’, Sapienza University of Rome, P.le A. Moro 5, Rome 00185, Italy; ³Center for Human Technologies, Istituto Italiano di Tecnologia, Via Erico Meloni 83, 16152 Genoa, Italy; ⁴Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Dr. Aiguader 88, 08003 Barcelona, Spain; ⁵Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain; ⁶Systems Biology of Infection Lab, Department of Biochemistry and Molecular Biology, Biosciences Faculty, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Spain; ⁷Institució Catalana de Recerca i Estudis Avançats (ICREA), 23 Passeig Lluís Companys, 08010 Barcelona, Spain

Correspondence: Elsa Zacco (elsa.zacco@iit.it) or Gian Gaetano Tartaglia (gian.tartaglia@iit.it)



Interactions between proteins and RNA are at the base of numerous cellular regulatory and functional phenomena. The investigation of the biological relevance of non-coding RNAs has led to the identification of numerous novel RNA-binding proteins (RBPs). However, defining the RNA sequences and structures that are selectively recognised by an RBP remains challenging, since these interactions can be transient and highly dynamic, and may be mediated by unstructured regions in the protein, as in the case of many non-canonical RBPs. Numerous experimental and computational methodologies have been developed to predict, identify and verify the binding between a given RBP and potential RNA partners, but navigating across the vast ocean of data can be frustrating and misleading. In this mini-review, we propose a workflow for the identification of the RNA binding partners of putative, newly identified RBPs. The large pool of potential binders selected by in-cell experiments can be enriched by *in silico* tools such as catRAPID, which is able to predict the RNA sequences more likely to interact with specific RBP regions with high accuracy. The RNA candidates with the highest potential can then be analysed *in vitro* to determine the binding strength and to precisely identify the binding sites. The results thus obtained can furthermore validate the computational predictions, offering an all-round solution to the issue of finding the most likely RNA binding partners for a newly identified potential RBP.

Introduction

Since their discovery and until recently, RNA-binding proteins (RBPs) have been identified by the presence of one or more RNA-binding domains in their sequences [1]. However, concomitantly to a new appreciation for RNA as key biological macromolecule acting at post-transcriptional level [2–4], there has also been a re-evaluation of what constitutes an RBP. Since one of the principal ways by which RNA exerts its function is by the formation of ribonucleoprotein complexes, every protein capable of establishing even weak and extemporary interactions with an RNA molecule may be defined as RBP [5,6]. The interactions of proteins with RNA can be highly dynamic and heavily dependent on the cellular environment [7], which makes the goal of defining the range of affinities and specificities quite challenging [8]. In fact, indiscriminate binding of RNA by RBPs is a quite common phenomenon [9], and the assumption that stronger affinity translates into more relevant biological functions is not necessarily correct [10]. For the scientific community, the revelation of the dynamicity and malleability of the partnership between RBPs and RNA allows for the exploration of new possible interaction mechanisms, networks, genes and protein regulation systems to investigate. It becomes, therefore, increasingly important to complete the catalogue of eukaryotic RBPs, at present

*These authors contributed equally to this work.

Received: 12 June 2020

Revised: 17 July 2020

Accepted: 20 July 2020

Version of Record published:
21 August 2020

containing >315 000 elements (about 6000 orthologues from >150 species) of which 3500 are from *H. sapiens* and can be divided in conventionally-defined RBPs and several classes of non-canonical RBPs [5,11,12].

Large-scale identification of new potential RBPs can reveal unexpected biological and pathological functions and, when confronted with a novel RBP, the identification of its RNA binding partners is a critical step to define the protein's cellular and molecular roles. To achieve this goal, increasingly sophisticated high-throughput methodologies have been developed, spanning from methods that aim to preserve the native cellular RNA–RBP interactions [13–15] to finely-controlled *in vitro* techniques that allow to define kinetics and dynamic parameters for each binding pair [16,17]. However, the field that has probably seen the biggest evolution in the shortest time span is the one of computational prediction algorithms [18–22]. The wide range of computational tools available includes several data-driven methods based on learning models, in which the algorithms are trained using experimental outcomes and databases to identify RNA–RBP binding patterns and define the genome-wide profiling of RNA–protein interactions [22].

In this short review, we propose a work pipeline for the identification of the RNA binding partners for novel putative RBPs. Starting from in-cell data harvesting, we would like to guide the reader through the employment of the different tools offered by *catRAPID* [23,24], our in-house developed RNA–protein interaction prediction algorithm, and propose some indications on how to validate the outcome experimentally. Our wider goal is to support the scientific community in the identification of novel biologically relevant non-canonical RBPs.

Identification of RNA binding partners in cellular context

The validation and training of computational algorithms for the prediction of protein–RNA interactions is strongly supported by experimental data (Figure 1). Prediction software can be significantly enriched by the output of techniques able to identify a protein's RNA partners within the cellular environment; such procedures are key to defining native interaction pairs and to monitoring responses and variations upon physiological *stimuli* or under pathological stress.

RIP-based approaches

The main tool to obtain information about the RNA binding partners of a target protein in the cellular environment is immunoprecipitation (IP), a widespread technique to pull down the protein of interest together with its physiological RNA binding partners. RNA immunoprecipitation (RIP) requires incubation of cell lysates with an antibody raised against the target protein [25]. RNA molecules bound to the target protein can then be isolated and analysed to reconstruct physiological native complexes formed within the cell. RIP can be coupled to either microarrays (RIP-Chip) or high-throughput sequencing (RIP-Seq). In either case, RIP can only determine the identity of the RNA molecules associated to the target protein, unless digestion-optimized RIP (DO-RIP) is performed [26]. This variation of RIP introduces an RNase digestion step to preserve only the portion of RNA bound to the protein, allowing for binding-site mapping [27]. If, instead, the interest is focused on identifying multi-subunit ribonucleoproteins, the most appropriate RIP variant may be RIP in tandem (RIPiT), which employs two distinct IP steps performed either with antibodies against different proteins of the complex or with antibodies binding different regions of the same target protein [28]. Information about native protein–RNA complexes formed within the cell can also be obtained by employing affinity tags [29], without having to rely on the antibody's specificity and sensitivity.

CLIP-based approaches

To overcome RIP's limitations (enrichment of indirectly bound RNAs, detection of interactions not present in cell but formed after lysis, loss of weaker interactions due to the required stringent washing conditions), cross-linked RNA immunoprecipitation (CLIP) has been developed [30]. CLIP promotes the stabilisation of the bonds between a protein and its interacting RNA, generally by UV radiation [31]. A large amount of CLIP variants is available, and most of them can offer high-resolution results at the single nucleotide level. Pioneers among these are, for example, the high-throughput sequencing CLIP (HITS-CLIP), that enriches the RNA population for sequences corresponding to the RBP binding sites [32]; the individual nucleotide-resolution CLIP (iCLIP) [33,34] and enhanced CLIP (eCLIP) [35], that utilise different oligonucleotide adapter configurations to obtain RNAs of different lengths employed to build the interacting fragments at single-nucleotide resolution; and the photoactivatable ribonucleoside CLIP (PAR-CLIP), that relies on metabolic incorporation of labelled ribonucleoside analogues that yield photo-adducts when cross-linked at selected wavelengths [36]. Radiation-free CLIP variants, such as infrared-CLIP (irCLIP), have also been reported [37]. To overcome the



Figure 1. The workflow of discovering RNA partners for an RBP.

Protein and RNA sequence databases, structural information and results from RIP/CLIP experiments feed computational prediction tools such as *catRAPID*. The software utilises this information to define RNA sequences with high probability of interacting with a given RBP and rank them accordingly. Several *in vitro* techniques allow for the validation of predicted results, for the calculation of binding strength and the definition of the binding sites.

restrictions imposed by the use of irreversible UV cross-linking, formaldehyde-based techniques such as fCLIP, which are more efficient in capturing interactions with dsRNA [38], can also be employed [39]. One limit of CLIP-Seq approaches is that low crosslinking efficiency makes low abundant transcripts difficult to detect, while transcripts present at high amounts are usually over-represented in IP samples. This issue can be partially solved with a proper bioinformatic analysis.

Bioinformatic analysis

High-throughput sequencing of the RNAs isolated through RIP, CLIP and related protocols yields millions of short ‘reads’, which represent the sequenced portions of cDNA fragments obtained through RNase titration (a step omitted in standard RIP-Seq), followed by reverse transcription and PCR amplification. The RNase treatment allows to obtain fragments long enough to be uniquely mappable but short enough to identify the binding site with the highest possible resolution. Library preparation ends with the production of cDNA fragments flanked by adapters that allow amplification and sequencing, resulting in the generation of short reads that can undergo bioinformatic analysis aimed at identifying RNA targets for the RBP (Figure 2). Reads pre-processing steps, including demultiplexing and adapter trimming, are often required, especially in those cases in which Unique Molecule Identifiers (UMIs) are used [40]. UMIs are random barcodes which identify unique cDNA fragments, allowing to detect and remove PCR duplicates that are commonly produced during CLIP-Seq library preparation [34,35].

To identify the RNA molecules from which they derive, reads are aligned to a reference genome using splice-aware alignment programs commonly used to analyze standard RNA-Seq data, like TopHat2 [41] or Star [42]. If the RBP under investigation binds mature mRNAs, reads can be mapped directly to the transcriptome [43] using a splice-unaware mapper like Bowtie2 [44]. Reads coming from HITS-CLIP experiments have high mutation rates (usually deletions) at the cross-linking site (CIMS, standing for cross-linking induced mutation sites), which are due to residual amino acids hindering the reverse transcriptase [45–47]. Similarly, the use of 4-thiouridine (4-SU) or 6-thioguanosine (6-SG) in the PAR-CLIP protocol leads to a high number of transition events (T to C or G to A, respectively) at the cross-linking sites [48]. Reads mapping can be improved by taking into account the high rate of such mutations. For instance, the splice-unaware BWA aligner [49] has been modified in order to incorporate an error model that favours PAR-CLIP specific transitions [50,51]. Reads produced by iCLIP, and likely eCLIP, experiments do not require such special treatment, since such protocols enrich for cDNAs truncated at the cross-linking site (CITS, standing for cross-linking induced truncation sites), while only a minor proportion of fragments represent CIMS-containing read-through cDNAs [43].

Post-processing of aligned reads is a mandatory step in RIP-Seq and CLIP-Seq data analysis. Reads aligning to multiple genomic positions are usually removed [52]. However, such multi-mapped reads can be used to identify regulatory RNA sites localized within repetitive regions [53]. To filter out PCR duplicates, reads mapping to the same genomic position are collapsed; UMIs, when present, can be used to avoid removing natural read duplicates, common in case of high sequencing depth [54].

In RIP-Seq experiments, target RNAs can be identified either by transcript enrichment analysis of IP versus control samples [55–57], which can be performed using procedures commonly adopted in differential expression analysis, or with ad-hoc peak-calling tools [58,59]. Binding site identification in HITS-CLIP experiments is usually accomplished by means of peak-calling approaches, that identify regions which are enriched in reads with respect to their genomic context (gene, transcript, metagene region) [59], the background represented by control experiments (input, IgG, mock IP) [60], or baseline expression profiles [59]. Single-nucleotide

resolution in HITS-CLIP data analysis can be achieved by looking for statistically significant CIMS [59]. Such resolution in binding site identification can also be attained in PAR-CLIP, by searching for transition events that are not likely to be caused by sequencing noise, single nucleotide polymorphisms or contamination [61]. The first nucleotide of reads from iCLIP and eCLIP experiments most often marks the truncation site and it is therefore located one nucleotide after the cross-linking site. Statistically significant CITS can be identified by using tools like iCount [62] and PureCLIP [63].

***In silico* prediction of protein–RNA interactions**

The outcome of the data analysis obtained by means of CLIP-Seq experiments is of enormous value in approaching the study of a potential RBP with unknown RNA partners. However, there are two major drawbacks in limiting the investigation to CLIP-Seq approaches: the protein regions in direct contact with the target RNAs remain unknown, and there are not sufficient data to speculate on the strength of the interaction for each protein–RNA pair. The use of predictive algorithms such as *catRAPID* would integrate the results of a CLIP-Seq experiment with this information [24] (Figure 1).

catRAPID

catRAPID is an algorithm able to compute protein–RNA interaction propensities with strong predictive power (area under the Receiver Operating Characteristic curve of 0.78 on >1 000 000 interactions) [64], through the calculation of secondary structures, hydrogen bonding and van der Waals contributions. The algorithm was trained on PDB crystals [24] and was later adapted to predict CLIP-Seq interactions with long non-coding RNAs [65]. For large proteins (>750 amino acids) and RNAs (>1000 nucleotides), the algorithm fragments sequences into overlapping segments and computes the interaction propensity through the analysis of physical–chemical properties and secondary structures of the molecules. According to the chosen implementation (Table 1), the method can either reconstruct the overall interaction propensity score for each protein–RNA pair [66] or rank the fragments according to the predicted interaction strength [67]. The outcome of this analysis allows to map both protein and RNA binding sites and to estimate the overall strength of the interactions [68], overcoming the limitations of CLIP-Seq techniques mentioned above. If used in combination with CLIP-Seq data relative to the protein of interest, *catRAPID* can be employed to select the best targets based on calculated binding strength, but it could be also useful in predicting putative targets that are not expressed within the cellular system. If no information about the RNA targets is available, *catRAPID* can represent a promising tool for the investigation of the protein’s genome-wide RNA-binding potential against an RNA sequence library. Here we suggest a pipeline that could be followed in such circumstance:

1. If the RNA binding potential of a given protein is unknown, *catRAPID signature* [71] can be used to predict it, along with the putative RNA-binding regions. This approach is particularly recommended if a potential RBP needs to be selected from a panel of candidate proteins;
2. *catRAPID omics* [69] can then be used to predict the interactions between the protein and a precompiled or custom RNA library. The result is a ranked list of protein–RNA pairs;
3. If the protein of interest is human, its co-expression with the putative interactors in different tissues can be evaluated using *catRAPID express* [70];
4. Once the most promising binding partners have been identified, *catRAPID strength* can be employed to evaluate the strength of each interaction [67];
5. Finally, *catRAPID fragments* [66] can be run on the highest scoring protein–RNA pairs to predict the binding sites. Interactions with long RNAs can be analyzed using *Global Score* [65] or *omiXcore* [68].

By narrowing down the number of potential targets and by suggesting the most likely binding sites, such approach can be employed to guide further experimental and computational analyses. Being a predictive tool, there is always the chance that *catRAPID* may fail in identifying valid RNA targets. Prediction accuracy depends on the set provided to train the algorithm. As more and more data become available, retraining of the algorithm will be necessary to achieve better performances.

***catRAPID* alternatives**

catRAPID is only one of several possible computational methods developed for predicting protein–RNA interactions. We would like to mention here some valid alternatives:

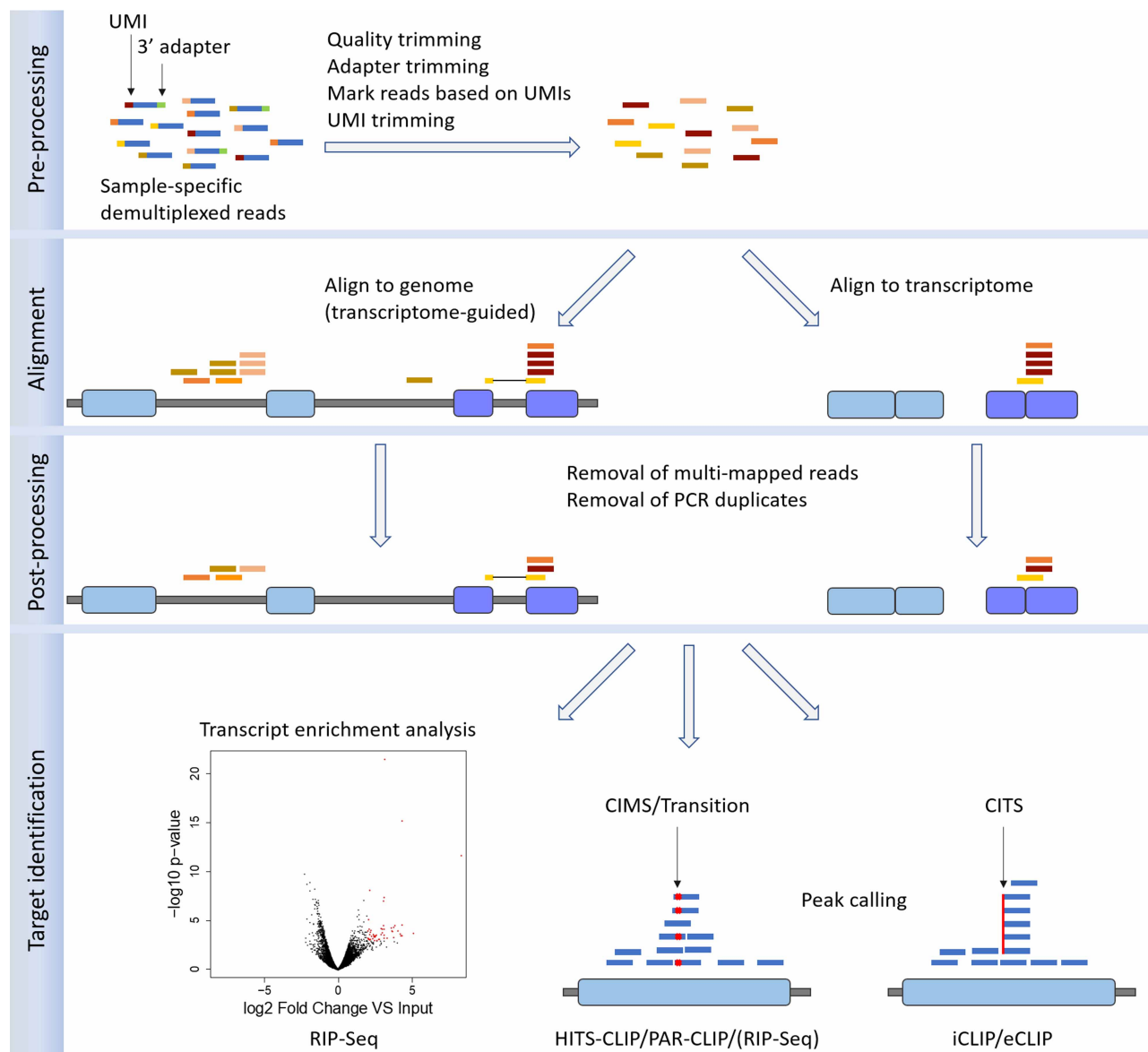


Figure 2. Bioinformatics analysis of CLIP-Seq and RIP-Seq data.

After being de-multiplexed based on sample-specific barcodes, reads undergo a **pre-processing** phase. UMIs are not always used, being more common in iCLIP and eCLIP protocols. When employed, they are sometimes used to remove PCR duplicates directly at this stage, but in most cases reads are simply marked based on UMI sequence, as shown by the colours assigned to trimmed reads. After the **alignment** of reads to the genome or to the transcriptome is performed, **post-processing** is needed to filter out multi-mapped reads and to collapse reads mapping at the same position, that are likely to represent PCR duplicates; if UMI-based read marking occurred, natural duplicates, which map at the same place but have different UMIs, can be retained, as shown here. **RNA target identification** and binding site detection strategy depend on the protocol. Roughly, such approaches can be divided into transcript enrichment analysis (RIP-Seq), which is analogous to differential expression analysis, and peak-calling (all protocols). Single-nucleotide resolution can be achieved using CIMSs in HITS-CLIP, transitions in PAR-CLIP and CITSs in iCLIP/eCLIP. HITS-CLIP experiments do not always produce clear and usable CIMS patterns.

- LncPro [72]: similar to *catRAPID* in the employment of RNA secondary structure, hydrogen-bonding and van der Waals interaction propensities. It is designed to predict whether a specific long non-coding RNA interacts with one or more protein sequences. Propensities are calculated for both protein and RNA and a probability score ranging from 0 to 100 is generated;

Table 1 A summary of the different *cat*RAPID implementations

Name of the algorithm	Description	Input	Output
<i>cat</i> RAPID fragments [66]	It divides inputted protein and RNA into fragments and computes the interaction propensity between each fragment.	<ul style="list-style-type: none"> • A protein sequence in FASTA format. • An RNA sequence in FASTA format. 	<ul style="list-style-type: none"> • Interaction profile plot^a. • Interaction matrix^b. • Table of interacting protein–RNA fragments.
<i>Global Score</i> [65] <i>omiXcore</i> [68]	A variant of <i>cat</i> RAPID fragments calibrated on CLIP-Seq data, it is able to predict interaction with >1000 nt long RNAs and to provide an overall interaction score.	<ul style="list-style-type: none"> • A protein sequence in FASTA format. • An RNA sequence in FASTA format. 	<ul style="list-style-type: none"> • Interaction profile plot^a • Interaction matrix^b. • Table of interacting protein–RNA fragments.
<i>cat</i> RAPID omics [69]	It computes the interactions between a molecule (protein/RNA) and the reference set (transcriptome/nucleotide-binding proteome) of a model organism.	<ul style="list-style-type: none"> • Protein/RNA sequence in FASTA format. • Reference set of protein/RNA sequences. 	<ul style="list-style-type: none"> • Graphical representation of protein sequence/ domains. • Pie chart with ranking distribution^c. • Table of interacting protein–RNA pairs.
<i>cat</i> RAPID express [70]	It allows the identification of co-expressed protein–RNA pairs in human tissues.	<ul style="list-style-type: none"> • A protein sequence in FASTA format • An RNA sequence in FASTA format. • (Only one protein sequence or one RNA sequence is required for the omics option). 	<ul style="list-style-type: none"> • Correlation coefficient representing the coexpression of the protein–RNA pair. • Interaction heatmap^d. • Table of tissue expression.
<i>cat</i> RAPID signature [71]	It scans a protein sequence for RNA-binding regions.	<ul style="list-style-type: none"> • One or more protein sequences in FASTA format. 	<ul style="list-style-type: none"> • Overall binding score. • Binding propensity plot^e.
<i>cat</i> RAPID library [69]	It allows the creation of a new reference set for <i>cat</i> RAPID omics.	<ul style="list-style-type: none"> • One or more protein or RNA sequences. 	<ul style="list-style-type: none"> • A library ID that can be used in <i>cat</i>RAPID omics.
<i>cat</i> RAPID strength [67]	It computes the interaction strength of a protein–RNA pair with respect to a reference set of sequences of similar length.	<ul style="list-style-type: none"> • A protein sequence in FASTA format • A RNA sequence in FASTA format. 	<ul style="list-style-type: none"> • Table of interaction strength (significance of interaction propensity). • Cumulative distribution function plots of protein–RNA interaction score^f.

^aThe interaction profile plot represents the interaction score (y-axis) of the protein along the RNA sequence (x-axis), giving information about the transcript regions that are most likely to be bound by the protein;

^bThe interaction matrix is an heatmap showing the interaction propensity between each possible fragment of the protein (y-axis) and the RNA (x-axis);

^cThe pie chart shows the proportion of targets having High, Moderate and Low star rating score. Star rating score weights the interaction based on the interaction propensity, the presence of RNA/DNA binding domains and the presence of known RNA motifs;

^dThe interaction heatmap shows the interaction score of the individual amino acid–nucleotide pairs;

^eThe binding propensity plot reports, for each amino acid (x-axis), the propensity to be part of a binding region;

^fThe Cumulative distribution function plots report the interaction score of the query protein–RNA pair within the distribution of the interaction scores from the reference set.

A more detailed explanation of the different algorithms is available on *cat*RAPID tutorial page (http://s.tartaglialab.com/static_files/shared/tutorial.html) and documentation page (http://s.tartaglialab.com/static_files/shared/documentation.html).

- RPISeq [73]: protein and RNA sequences are encoded into features that are then used to train Support Vector Machine (SVM) and Random Forests classifiers;
- RPI-Pred [74]: it combines RNA and protein sequences with predicted or actual 3D structures. The features are then used to train an SVM classifier;
- iDeepS [75]: a deep learning-based method that exploits convolutional neural networks (CNNs) trained on RNA sequences and predicted secondary structures. At the end of the pipeline, a classification layer is

Table 2 Methods for *in vitro* characterisation of protein–RNA binding

Method	Principle of detection	Sample requirements	Detection range	Sample capacity	Direct measurements
EMSA [90]	Detection of RNA–protein complex' electrophoretic mobility properties, typically different compared to free RNA.	<ul style="list-style-type: none"> • Labelled RNA. • nM concentrations of RNA and protein. 	$\geq 10^{-18}$ mol RNA.	0.5–500 μ l depending on electrophoresis setup.	K_d, n
Filter binding assay [88]	Quantification of 32 P-labelled RNA via imagine screen or scintillation counter.	<ul style="list-style-type: none"> • About 0.1 μM labelled RNA (usually with 32P). • Purified protein serial dilutions. 	$\geq 10^{-15}$ mol RNA.	Multi-well plate dot-blot setup.	k_{on}, k_{off}, K_d, n
Fluorescence anisotropy [91,92]	Changes in fluorescence anisotropy or polarisation of excitation light upon binding.	<ul style="list-style-type: none"> • Fluorescent labelling of one of the partners. • 1 nM RNA. 	nM ranges of fluorophores.	Multi-well plates.	K_d
FRET [93,94]	Energy transfer of between fluorophores detected as a change in fluorescence intensity.	<ul style="list-style-type: none"> • Two fluorophores, either one on each partner or strategically placed on one for structural studies. 	single-molecule experiments.	Single molecule to multi-well plates.	K_d, k_{on}, k_{off} , distance between fluorophores.
SPR [95]	Variations in the refractive index of polarised laser light upon molecular binding.	<ul style="list-style-type: none"> • About 200 μl 25 nM RNA/sensor. • Variable conc. of protein (ideally 100-times K_d), up to 4 ml of sample. • Immobilisation of one partner required. 	1 pM < K_d < 1 mM	Up to 16 channels with microfluidics.	K_d, k_{on}, k_{off}
BLI [96]	Detection of the variation of refracted white light upon the binding of the interaction partner to the immobilised ligand on the optical fibres.	<ul style="list-style-type: none"> • 1–50 μg/ml of ligand, immobilised on biosensor. • 1 nM–μM of receptor. • 5–250 μl of sample per measurement. 	1 nM < K_d < 10 mM	Single channel, 5 min per measurement (BLItz) or multi-well plate, 1–8 simultaneous channels.	K_d, k_{on}, k_{off}
MST [97]	Variations in temperature-induced fluorescence emission of a target as a function of the concentration of a non-fluorescent ligand.	<ul style="list-style-type: none"> • 1–20 μl, nM–μM concentrations. • Fluorescent labelling. 	pM < K_d < mM	Up to 96 samples per run in a multi-capillary system.	K_d
switchSENSE [98]	Voltage-dependent variations of the movement of short fluorescent DNA nanolevers attached to a gold surface upon binding of an analyte.	<ul style="list-style-type: none"> • Immobilisation of one binding partner. • 20 μl of 1 μM RNA for biochip saturation. • 250 μl of 0.2 μM protein. 	nM < K_d < μ M	Four flow channels with six microelectrodes for sampling per chip	$K_d, k_{on}, k_{off}, R_h$
ITC [99]	Measuring the heat consumed/released during titration of sample with the ligand in regard to reference cell.	<ul style="list-style-type: none"> • 200 μl–2 ml of 1–2 μM receptor. • 40–500 μl 10\times concentration ligand. 	<ul style="list-style-type: none"> • K_d > nM (direct measurements). • K_d > pM (competitive binding). 	single cell.	$K_d, \Delta H, n$

Kinetic constants are measured directly and are used as basis for equilibrium thermodynamic parameters calculations, apart from ITC where the reaction enthalpy can be obtained without relying on kinetic data. K_d : equilibrium dissociation constant; k_{on} : association rate constant; k_{off} : dissociation rate constant; n : stoichiometry of binding; R_h : hydrodynamic radius (radius of a theoretical sphere with the same translational diffusion coefficient); ΔH : reaction enthalpy.

responsible for RBP binding sites prediction. Deep learning models are generated individually for each RBP based on available CLIP-Seq data, allowing the formulation of predictions on a limited set of proteins.

Functional characterization of RNA targets and binding sites

Once the RNA targets of a protein have been determined, further analyses are necessary in order to verify the reliability of the results and to gain insights into the biological function of the RBP. Both tasks can be approached by looking at the function of target RNAs. A common way to do that is to perform an Over Representation Analysis (ORA) [76], which consists of identifying over-represented functional categories in a list of genes. Enrichment is evaluated against a background composed of all the expressed genes. A more sophisticated approach, called Gene Set Enrichment Analysis (GSEA) [76], involves ranking genes based on a certain score and evaluating if some categories are enriched at the top or the bottom of the ranked list. An implementation of this method that is specific for CLIP-Seq data is provided by the Seten tool [77]. This program requires as input a set of CLIP-Seq peaks, each with a score assigned by the peak-caller, but it could also work starting from predicted binding sites, as long as an interaction score is provided.

Another common analysis consists in scanning the identified binding sites in order to detect common patterns highlighting RBP binding preferences (motif analysis). Sequence motifs recurring in large sets of binding sites can be discovered using different tools, like MEME-ChIP [78] and SeAMotE [79]. Both tools start from a set of sub-sequences identified in the positive sequence set and evaluate their enrichment with respect to a control sequence set (unbound RNAs). A more recent tool, named mCross, exploits the single-nucleotide resolution offered by CLIP-Seq techniques to enhance the accuracy of *de novo* motif discovery [80].

Sequence alone may not be sufficient to fully explain the binding specificity of an RBP: a sequence motif could be accessible only when put in a proper secondary structure context. Tools like GraphProt [81], ssHMM [82] and BEAM [83] are able to detect motifs encoding both sequence and secondary structure information.

In vitro validation of predicted RBP–RNA interactions

To validate the prediction accuracy of the computational analysis proposed so far, it is ideal to evaluate each interacting pair within a controlled environment. A most accurate validation should start from the screening of potential binders, followed by the precise determination of binding sites and kinetic and thermodynamic parameters, and completed with structural insights into the drivers of the interaction (Figure 1). A comprehensive review of the methodology is beyond the scope of this article and for more details we refer to a number of recent reviews of the field [16,17,84–86].

Kinetics and thermodynamics of RBP–RNA interactions

As for other molecular interactions, the binding between a protein and an RNA molecule is kinetically characterised by the rate at which they associate (k_{on}) and dissociate (k_{off}). Conventionally, the dissociation constant (K_{d}), which is the ratio between k_{off} and k_{on} at the chemical equilibrium, is used to express the binding affinity: the lower the K_{d} , the greater the affinity. It is however important to note that binding pairs with the same K_{d} may have different k_{on} and k_{off} and therefore different binding mechanisms. To add another layer of complexity, many RBPs have multiple binding regions that may vastly differ in their affinity towards the same RNA [87].

There are several established methods for determining the kinetic and thermodynamic parameters of binding (Table 2). Techniques such as electrophoretic mobility shift assay (EMSA) and filter binding assay can be useful to estimate binding affinities with basic molecular biology tools [88,89]. Both these methods represent a viable first step analysis, especially because of short protocols and limited amounts of samples required. The latter criterion can be crucial in protein–RNA interaction studies, since some RBPs can be very difficult to isolate, tagged RNA synthesis can be expensive and advanced methods for more reliable and accurate determination of molecular binding characteristics generally have specific sample requirements and higher operation costs. Despite these advantages, since EMSA and filter binding experiments are performed within conditions very distinct from the *in vivo* ones (polyacrylamide gel and nitrocellulose filter, respectively), more reliable kinetic data may be obtained by other techniques. Examples of such methods are bio-layer interferometry (BLI), multi-channel surface plasmon resonance (SPR), microscale thermophoresis (MST), the employment of

fluorescence resonance energy transfer (FRET), and the most recent switchSENSE (Table 2, [86–89, 95]). Some of these require the labelling of one of the interactors with fluorescent dyes, as in the case of MST and FRET, or sample immobilisation on biosensors, as needed for SPR and BLI experiments. These requisites can impose certain structural restraints and thus compromise the binding and the results obtained. Comparing how the same RBP binds to RNA molecules that differ only in one nucleotide can help the identification of the RNA portion physically interacting with the protein. However, the determination of the exact nucleotides involved in

Table 3 A short overview of the major structural biology techniques with a comparison of their advantages and disadvantages for the study of protein–RNA interactions

Method	Principle of detection	Resolution	Sample requirements	Pros/Cons
<i>NMR</i> [111,114]	Detection of the electric current, induced by the magnetization of the non-equilibrium spins in a magnetic field. Upon Fourier transform, the results can be used to determine structural constraints and produce a molecular model.	atomic (<2 Å).	<ul style="list-style-type: none"> • Isotope labelling, side-chain deuteration essential for larger complexes to avoid lengthy relaxation times. • Protein concentration varies according to MW. 	<ul style="list-style-type: none"> • Solution-based, can observe time-resolved experiments and kinetics, most accessible on the list, possibilities of differential isotope labelling, saturation transfer experiments and more. • Poor signal-to-noise ratio, line broadening and complex spectra with higher molecular mass complexes.
<i>X-ray crystallography</i> [115,116]	Detection of diffracted X-ray photons, scattered by the crystal, from which an 3D electron density map is calculated, which is then used to build the molecular structure model.	atomic (<2 Å).	Crystals of the protein–RNA complex, frozen in liquid nitrogen.	<ul style="list-style-type: none"> • Highest resolution limit with free electron lasers. • Relies on quality crystals, often difficult to obtain.
<i>Cryogenic electron microscopy</i> [117]	Based on electron microscopy, the sample images are grouped into specific projections, with a 3D model calculated based on them.	high (<5 Å).	Monodisperse sample blotted onto grids and frozen under cryogenic conditions.	<ul style="list-style-type: none"> • Solution based, flexible buffer components, no need for crystals etc. • Maximum resolution limit around 3.5 Å for molecules with a MW ~50 kDa. • Very difficult to obtain sufficient quality data for determination of structures of elements with MW < 150 kDa under the resolution of 5 Å.
<i>Small angle scattering</i> [110,112,118]	Detection of diffracted X-ray photons (SAXS) or neutrons (SANS) on sample solutions under small angles (typically <10°), from which a scattering curve and a 3D shape can be calculated.	medium (>10 Å).	<ul style="list-style-type: none"> • Monodisperse sample, dilution series from 1 mg/ml to 20 mg/ml. • Possible deuteration or isotope labelling for SANS studies. 	<ul style="list-style-type: none"> • Investigation of molecule shape as well as other information, selective deuteration can provide valuable contrast (SANS). • Need for monodisperse sample, rather high protein concentrations for the dilution curve, no exact molecular structure.

the binding can be highly challenging without structural studies. Once the RNA targets have been precisely characterised, selective mutations on the protein region thought to be responsible for the binding may be useful to determine which amino acids are directly responsible for the interaction. The same approach could be extended to mutating selected nucleotides or regions of the RNA molecules for further characterisation of the binding spots without obtaining an atomic resolution structure.

Kinetic studies are essential for determining the binding propensity and validating different binding partners [10]. However, experiments conducted *in vitro* with isolated components often do not allow studies under physiologically relevant conditions. Isolated proteins can be sensitive to higher temperatures, especially under prolonged experiments that far exceed their in-cell lifespans. Since chemical equilibria are temperature-dependent, *in vivo* kinetics may thus significantly differ from those measured *in vitro*. The thermodynamics of binding can also reveal the energetic landscape of protein–RNA interactions [100–102]. It is therefore very important to measure the thermodynamics parameters directly, when experimentally possible, or to calculate them. Isothermal titration calorimetry (ITC) allows direct measurements of the equilibrium constant, stoichiometry and reaction enthalpy (ΔH) at a given temperature [99]. Using the van't Hoff equation, these can then be used to determine the temperature dependency of the equilibrium constant. SPR and MST also allow calculation of thermodynamic data based on the stable temperature of the measurement cell, while BLI is considered less reliable.

Structural approaches of studying RBP–RNA interactions

Kinetic and thermodynamic data obtained from interaction studies provide a good numerical description of the binding and, through the use of RNA-centric methods, a library of RNA sequences with high binding propensities [103]. However, protein and RNA sequences alone may not be sufficient to fully characterise their binding specificity. This is especially true in the case of non-canonical RBPs that do not contain a consensus RNA-recognition sequence [104]. RNA structure has been shown to be the driver behind most non-canonical RBP binding events, with highly structured RNAs having bigger protein interactomes [105,106]. The approaches to determine the structure of separate components have been extensively reviewed [107,108]. Among the most pertinent techniques to determine the macromolecular structure of a complex, there are X-ray crystallography, nuclear magnetic resonance (NMR) and cryogenic electron microscopy (cryo-EM) (Table 3). However, the definition of structural details by these methodologies can be challenging [107]. A large number of RBPs contain intrinsically disordered regions and tend to form macromolecular condensates, making their crystallisation impractical; NMR, that bypasses the need for crystals, is dependent on the molecular weight of the complexes that can make relaxation times slow and signal-to-noise ratio poor; obtaining results at atomic resolution with cryo-EM remains difficult [108]. These limitations make a strong case for the employment of complementary structural biology techniques [107]. A particularly promising development is the advance of small angle scattering and computational methods for data analysis, in particular small angle neutron scattering (SANS) [109]. Selective deuteration enables a higher degree of contrast between binding partners and can therefore provide a rough position for each component of the complex [110]. Data obtained from the above mentioned methods can be used as structural restraints for molecular dynamics simulations and data-driven docking [111,112] and can be integrated together in a hybrid multi-level approach for studying RNA-protein complexes, thus completing the full circle of integrated methodologies [75,113,119,120].

Perspectives

- Our understanding of many physiological and pathological phenomena cannot be exempted from an in-depth knowledge of protein–RNA interactions underlying them.
- Such comprehension, which goes from the identification of targets RNAs to binding modes characterization, requires a multidisciplinary approach involving biochemistry, molecular biology, bioinformatics and physics.

- As new and more powerful high-throughput methods and predictors are being developed, an integrated and productive usage of both approaches becomes more and more feasible. For instance, predictive tools such as *cat*RAPID, which is general enough to be applied to any protein–RNA pair, could also be employed to improve the specificity of omics studies.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

The research leading to this work has been supported by European Research Council (RIBOMYLOME_309545 and ASTRA_855923), by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754490 and under projects IASIS_727658 and INFORE_825080, by the Spanish Ministry of Economy and Competitiveness BFU2017-86970-P as well as the collaboration with Peter St. George-Hyslop financed by the Wellcome Trust.

Author Contributions

A.C., J.R. and A.V. drafted the sections of the review pertinent to their expertise. E.Z. and G.G.T. guided, supervised and coordinated the work. All authors contributed to writing and editing the manuscript and to the original production of all figures and tables provided.

Acknowledgements

The authors would like to thank all members of Tartaglia lab.

Abbreviations

BLI, biolayer interferometry; CIMS, crosslinking-induced mutation sites; CITS, crosslinking-induced truncation sites; CLIP, cross-linking followed by immunoprecipitation; CLIP-Seq, CLIP followed by high-throughput sequencing; eCLIP, enhanced CLIP; FRET, Förster resonance energy transfer; HITS-CLIP, high-throughput sequencing CLIP; iCLIP, individual nucleotide-resolution CLIP; IP, immunoprecipitation; MST, microscale thermophoresis; NMR, nuclear magnetic resonance; PAR-CLIP, photoactivatable ribonucleoside CLIP; RBP, RNA-binding protein; RIP, RNA immunoprecipitation; RIP-Seq, RIP followed by high-throughput sequencing; SPR, surface plasmon resonance; SVM, support vector machine; UMI, unique molecule identifier.

References

- 1 Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.M., Foehr, S. et al. (2016) Comprehensive identification of RNA-binding domains in human cells. *Mol. Cell* **63**, 696–710 <https://doi.org/10.1016/j.molcel.2016.06.029>
- 2 Blackinton, J.G. and Keene, J.D. (2014) Post-transcriptional RNA regulons affecting cell cycle and proliferation. *Semin. Cell Dev. Biol.* **34**, 44–54 <https://doi.org/10.1016/j.semcdb.2014.05.014>
- 3 Bernabò, P., Viero, G. and Lencioni, V. (2020) A long noncoding RNA acts as a posttranscriptional regulator of heat shock protein (HSP70) synthesis in the cold hardy *Diamesa tonsa* under heat shock. *PLoS One* **15**, e0227172 <https://doi.org/10.1371/journal.pone.0227172>
- 4 García-Mauriño, S.M., Rivero-Rodríguez, F., Velázquez-Cruz, A., Hernández-Vellisca, M., Díaz-Quintana, A., De la Rosa, M.A. et al. (2017) RNA binding protein regulation and cross-talk in the control of AU-rich mRNA fate. *Front. Mol. Biosci.* **4**, 71 <https://doi.org/10.3389/fmolb.2017.00071>
- 5 Hentze, M.W., Castello, A., Schwarzl, T. and Preiss, T. (2018) A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* **19**, 327–341 <https://doi.org/10.1038/nrm.2017.130>
- 6 Moore, S., Järvelin, A.I., Davis, I., Bond, G.L. and Castello, A. (2018) Expanding horizons: new roles for non-canonical RNA-binding proteins in cancer. *Curr. Opin. Genet. Dev.* **48**, 112–120 <https://doi.org/10.1016/j.gde.2017.11.006>
- 7 Backlund, M., Stein, F., Rettel, M., Schwarzl, T., Perez-Perri, J.I., Brosig, A. et al. (2020) Plasticity of nuclear and cytoplasmic stress responses of RNA-binding proteins. *Nucleic Acids Res.* **48**, 4740 <https://doi.org/10.1093/nar/gkaa256>
- 8 Gebauer, F., Preiss, T. and Hentze, M.W. (2012) From cis-regulatory elements to complex RNPs and back. *Cold Spring Harb. Perspect. Biol.* **4**, 1–14 <https://doi.org/10.1101/cshperspect.a012245>
- 9 Mukherjee, N., Wessels, H.H., Lebedeva, S., Sajek, M., Ghanbari, M., Garzia, A., et al. (2019) Deciphering human ribonucleoprotein regulatory networks. *Nucleic Acids Res.* **47**, 570–581 <https://doi.org/10.1093/nar/gky1185>
- 10 Jankowsky, E. and Harris, M.E. (2015) Specificity and nonspecificity in RNA-protein interactions. *Nat. Rev. Mol. Cell Biol.* **16**, 533–544 <https://doi.org/10.1038/nrm4032>

- 11 Liao, J.-Y.Y., Yang, B., Zhang, Y.-C.C.Y., Wang, X.-J.J., Ye, Y., Peng, J.-W.W., et al. (2020) EuRBPDB: a comprehensive resource for annotation, functional and oncological investigation of eukaryotic RNA binding proteins (RBPs). *Nucleic Acids Res.* **48**, 307–313 <https://doi.org/10.1093/nar/gkz823>
- 12 Gerstberger, S., Hafner, M. and Tuschl, T. (2014) A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829–845 <https://doi.org/10.1038/nrg3813>
- 13 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 https://doi.org/10.1007/978-1-60327-475-3_6
- 14 Ramanathan, M., Majzoub, K., Rao, D.S., Neela, P.H., Zarnegar, B.J., Mondal, S., et al. (2018) RNA-protein interaction detection in living cells. *Nat. Methods* **15**, 207–212 <https://doi.org/10.1038/nmeth.4601>
- 15 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 <https://doi.org/10.1016/j.ymeth.2005.07.018>
- 16 Dasti, A., Cid-Samper, F., Bechara, E. and Tartaglia, G.G. (2019) RNA-centric approaches to study RNA-protein interactions in vitro and in silico. *Methods* **178**, 11–18 <https://doi.org/10.1016/j.ymeth.2019.09.011>
- 17 Ye, X. and Jankowsky, E. (2019) High throughput approaches to study RNA-protein interactions in vitro. *Methods* **178**, 3–10 <https://doi.org/10.1016/j.ymeth.2019.09.006>
- 18 Lam, J.H., Li, Y., Zhu, L., Umarov, R., Jiang, H., Héliou, A., et al. (2019) A deep learning framework to predict binding preference of RNA constituents on protein surface. *Nat. Commun.* **10**, 1–13 <https://doi.org/10.1038/s41467-018-07882-8>
- 19 Zhou, Y.-K., Shen, Z.-A., Yu, H., Luo, T., Gao, Y. and Du, P.-F. (2020) Predicting lncRNA–protein interactions with miRNAs as mediators in a heterogeneous network model. *Front. Genet.* **10**, 1341 <https://doi.org/10.3389/fgene.2019.01341>
- 20 Qiu, J., Bernhofer, M., Heinzinger, M., Kemper, S., Norambuena, T., Melo, F. et al. (2020) ProNA2020 predicts protein–DNA, protein–RNA, and protein–protein binding proteins and residues from sequence. *J. Mol. Biol.* **432**, 2428–2443 <https://doi.org/10.1016/j.jmb.2020.02.026>
- 21 Adjero, D., Allaga, M., Tan, J., Lin, J., Jiang, Y., Abbasi, A. et al. (2018) Feature-based and string-based models for predicting RNA-protein interaction. *Molecules* **23**, 697 <https://doi.org/10.3390/molecules23030697>
- 22 Pan, X., Yang, Y., Xia, C.Q., Mirza, A.H. and Shen, H.B. (2019) Recent methodology progress of deep learning for RNA–protein interaction prediction. *Wiley Interdiscip. Rev. RNA* <https://doi.org/10.1002/wrna.1544>
- 23 catRAPID
- 24 Bellucci, M., Agostini, F., Masin, M. and Tartaglia, G.G. (2011) Predicting protein associations with long noncoding RNAs. *Nat. Methods* **8**, 444–445 <https://doi.org/10.1038/nmeth.1611>
- 25 Gagliardi, M. and Matarazzo, M.R. (2016) RIP: RNA immunoprecipitation. *Methods Mol. Biol.* **1480**, 73–86 https://doi.org/10.1007/978-1-4939-6380-5_7
- 26 Nicholson, C.O., Friedersdorf, M. and Keene, J.D. (2017) Quantifying RNA binding sites transcriptome-wide using DO-RIP-seq. *RNA* **23**, 32–46 <https://doi.org/10.1261/ma.058115.116>
- 27 Nicholson, C.O., Friedersdorf, M.B., Bisogno, L.S. and Keene, J.D. (2017) DO-RIP-seq to quantify RNA binding sites transcriptome-wide. *Methods* **118–119**, 16–23 <https://doi.org/10.1016/j.ymeth.2016.11.004>
- 28 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 <https://doi.org/10.1016/j.ymeth.2013.09.013>
- 29 Philippe, N., Salson, M., Combes, T. and Rivals, E. (2013) CRAC: an integrated approach to the analysis of RNA-seq reads. *Genome Biol.* **14**, R30 <https://doi.org/10.1186/gb-2013-14-3-r30>
- 30 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 <https://doi.org/10.1126/science.1090095>
- 31 Tiedge, H. (1991) The use of UV light as a cross-linking agent for cells and tissue sections in in situ hybridization. *DNA Cell Biol.* **10**, 143–147 <https://doi.org/10.1089/dna.1991.10.143>
- 32 Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., et al. (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469 <https://doi.org/10.1038/nature07488>
- 33 Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L.E., Sibley, C.R., Sugimoto, Y. et al. (2014) iCLIP: protein-RNA interactions at nucleotide resolution. *Methods* **65**, 274–287 <https://doi.org/10.1016/j.ymeth.2013.10.011>
- 34 König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B. et al. (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* **17**, 909–915 <https://doi.org/10.1038/nsmb.1838>
- 35 Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., et al. (2016) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514 <https://doi.org/10.1038/nmeth.3810>
- 36 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 <https://doi.org/10.1002/wrna.1103>
- 37 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 <https://doi.org/10.1038/nmeth.3840>
- 38 Kim, B., Jeong, K. and Kim, V.N. (2017) Genome-wide mapping of DROSHA cleavage sites on primary microRNAs and noncanonical substrates. *Mol. Cell* **66**, 258–269.e5 <https://doi.org/10.1016/j.molcel.2017.03.013>
- 39 Klockenbusch, C., O'Hara, J.E. and Kast, J. (2012) Advancing formaldehyde cross-linking towards quantitative proteomic applications. *Anal. Bioanal. Chem.* **404**, 1057–1067 <https://doi.org/10.1007/s00216-012-6065-9>
- 40 Kivioja, T., Vähärautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S. et al. (2012) Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74 <https://doi.org/10.1038/nmeth.1778>
- 41 Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) Tophat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 <https://doi.org/10.1186/gb-2013-14-4-r36>
- 42 Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S. et al. (2012) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 <https://doi.org/10.1093/bioinformatics/bts635>

- 43 Haberman, N., Huppertz, I., Attig, J., König, J., Wang, Z., Hauer, C., et al. (2017) Insights into the design and interpretation of iCLIP experiments. *Genome Biol.* **18**, 7 <https://doi.org/10.1186/s13059-016-1130-x>
- 44 Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 <https://doi.org/10.1038/nmeth.1923>
- 45 Granneman, S., Kudla, G., Petfalski, E. and Tollervey, D. (2009) Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc. Natl Acad. Sci. U.S.A.* **106**, 9613–9618 <https://doi.org/10.1073/pnas.0901997106>
- 46 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 <https://doi.org/10.1038/nbt.1873>
- 47 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**, 559–567 <https://doi.org/10.1038/nmeth.1608>
- 48 Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., et al. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 <https://doi.org/10.1016/j.cell.2010.03.009>
- 49 Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 <https://doi.org/10.1093/bioinformatics/btp324>
- 50 Kerpedjiev, P., Frellsen, J., Lindgreen, S. and Krogh, A. (2014) Adaptable probabilistic mapping of short reads using position specific scoring matrices. *BMC Bioinformatics* **15**, 100 <https://doi.org/10.1186/1471-2105-15-100>
- 51 Kloetgen, A., Borkhardt, A., Hoell, J.I. and McHardy, A.C. (2016) The PARA-suite: PAR-CLIP specific sequence read simulation and processing. *PeerJ* **4**, e2619 <https://doi.org/10.7717/peerj.2619>
- 52 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. *Nucleic Acids Res.* **43**, 5263–5274 <https://doi.org/10.1093/nar/gkv439>
- 53 Zhang, Z. and Xing, Y. (2017) CLIP-seq analysis of multi-mapped reads discovers novel functional RNA regulatory sites in the human transcriptome. *Nucleic Acids Res.* **45**, 9260–9271 <https://doi.org/10.1093/nar/gkx646>
- 54 Fu, Y., Wu, P.H., Beane, T., Zamore, P.D. and Weng, Z. (2018) Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. *BMC Genomics* **19**, 531 <https://doi.org/10.1186/s12864-018-4933-1>
- 55 Lu, Z., Guan, X., Schmidt, C.A. and Matera, A.G. (2014) RIP-seq analysis of eukaryotic Sm proteins identifies three major categories of Sm-containing ribonucleoproteins. *Genome Biol.* **15**, 1–23 <https://doi.org/10.1186/gb-2014-15-1-r1>
- 56 Moore, K.S. and t Hoen, P.A.C. (2019) Computational approaches for the analysis of RNA–protein interactions: a primer for biologists. *J. Biol. Chem.* **294**, 1–9 <https://doi.org/10.1074/jbc.REV118.004842>
- 57 Baquero-Perez, B., Antanaviciute, A., Yonchev, I.D., Carr, I.M., Wilson, S.A. and Whitehouse, A. (2019) The tudor SND1 protein is an m⁶A RNA reader essential for replication of kaposi's sarcoma-associated herpesvirus. *eLife* **8**, e47261 <https://doi.org/10.7554/eLife.47261>
- 58 Li, Y., Zhao, D.Y., Greenblatt, J.F. and Zhang, Z. (2013) RIPSeeker: a statistical package for identifying protein-associated transcripts from RIP-seq experiments. *Nucleic Acids Res.* **41**, e94 <https://doi.org/10.1093/nar/gkt142>
- 59 Kucukural, A., Özadam, H., Singh, G., Moore, M.J. and Cenik, C. (2013) ASPeak: an abundance sensitive peak detection algorithm for RIP-Seq. *Bioinformatics* **29**, 2485–2486 <https://doi.org/10.1093/bioinformatics/btt428>
- 60 Uren, P.J., Bahrami-Samani, E., Burns, S.C., Qiao, M., Karginov, F.V., Hodges, E. et al. (2012) Site identification in high-throughput RNA-protein interaction data. *Bioinformatics* **28**, 3013–3020 <https://doi.org/10.1093/bioinformatics/bts569>
- 61 Golumbeanu, M., Mohammadi, P. and Beerenwinkel, N. (2016) BMix: probabilistic modeling of occurring substitutions in PAR-CLIP data. *Bioinformatics* **32**, 976–983 <https://doi.org/10.1093/bioinformatics/btv520>
- 62 Curk, T., Rot, G., Gorup, Č., de los Mozos, I.R., König, J., Zmrzlikar, J. et al. (2019) iCount: protein-RNA interaction iCLIP data analysis (in preparation)
- 63 Krakau, S., Richard, H. and Marsico, A. (2017) PureCLIP: capturing target-specific protein-RNA interaction footprints from single-nucleotide CLIP-seq data. *Genome Biol.* **18**, 240 <https://doi.org/10.1186/s13059-017-1364-2>
- 64 Lang, B., Armaos, A. and Tartaglia, G.G. (2019) RNAct: protein-RNA interaction predictions for model organisms with supporting experimental data. *Nucleic Acids Res.* **47**, D601–D606 <https://doi.org/10.1093/nar/gky967>
- 65 Cirillo, D., Blanco, M., Armaos, A., Bunes, A., Avner, P., Guttman, M. et al. (2016) Quantitative predictions of protein interactions with long noncoding RNAs. *Nat. Methods* **14**, 5–6 <https://doi.org/10.1038/nmeth.4100>
- 66 Cirillo, D., Agostini, F., Klus, P., Marchese, D., Rodriguez, S., Bolognesi, B. et al. (2013) Neurodegenerative diseases: quantitative predictions of protein-RNA interactions. *RNA* **19**, 129–140 <https://doi.org/10.1261/ma.034777.112>
- 67 Agostini, F., Cirillo, D., Bolognesi, B. and Tartaglia, G.G. (2013) X-inactivation: quantitative predictions of protein interactions in the Xist network. *Nucleic Acids Res.* **41**, e31 <https://doi.org/10.1093/nar/gks968>
- 68 Armaos, A., Cirillo, D. and Gaetano Tartaglia, G. (2017) Omixcore: a web server for prediction of protein interactions with large RNA. *Bioinformatics* **33**, 3104–3106 <https://doi.org/10.1093/bioinformatics/btx361>
- 69 Agostini, F., Zanzoni, A., Klus, P., Marchese, D., Cirillo, D. and Tartaglia, G.G. (2013) CatRAPID omics: a web server for large-scale prediction of protein-RNA interactions. *Bioinformatics* **29**, 2928–2930 <https://doi.org/10.1093/bioinformatics/btt495>
- 70 Cirillo, D., Marchese, D., Agostini, F., Livi, C.M., Botta-Orfila, T. and Tartaglia, G.G. (2014) Constitutive patterns of gene expression regulated by RNA-binding proteins. *Genome Biol.* **15**, R13 <https://doi.org/10.1186/gb-2014-15-1-r13>
- 71 Livi, C.M., Klus, P., Delli Ponti, R. and Tartaglia, G.G. (2016) CatRAPID signature: identification of ribonucleoproteins and RNA-binding regions. *Bioinformatics* **32**, 773–775 <https://doi.org/10.1093/bioinformatics/btv629>
- 72 Lu, Q., Ren, S., Lu, M., Zhang, Y., Zhu, D., Zhang, X. et al. (2013) Computational prediction of associations between long non-coding RNAs and proteins. *BMC Genomics* **14**, 1 <https://doi.org/10.1186/1471-2164-14-1>
- 73 Muppilala, U.K., Honavar, V.G. and Dobbs, D. (2011) Predicting RNA-protein interactions using only sequence information. *BMC Bioinformatics* **12**, 489 <https://doi.org/10.1186/1471-2105-12-489>
- 74 Suresh, V., Liu, L., Adjeroh, D. and Zhou, X. (2015) RPI-Pred: predicting ncRNA-protein interaction using sequence and structural information. *Nucleic Acids Res.* **43**, 1370–1379 <https://doi.org/10.1093/nar/gkv020>
- 75 Pan, X., Rijnbeek, P., Yan, J. and Shen, H.B. (2018) Prediction of RNA-protein sequence and structure binding preferences using deep convolutional and recurrent neural networks. *BMC Genomics* **19**, 511 <https://doi.org/10.1186/s12864-018-4889-1>

- 76 Boyle, E.I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J.M. et al. (2004) GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes INTRODUCTION: MOTIVATION AND DESIGN. *Bioinformatics* **20**, 3710–3715 <https://doi.org/10.1093/bioinformatics/bth456>
- 77 Budak, G., Srivastava, R. and Janga, S.C. (2017) Seten: a tool for systematic identification and comparison of processes, phenotypes, and diseases associated with RNA-binding proteins from condition-specific CLIP-seq profiles. *RNA* **23**, 836–846 <https://doi.org/10.1261/ma.059089.116>
- 78 Machanick, P. and Bailey, T.L. (2011) MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 <https://doi.org/10.1093/bioinformatics/btr189>
- 79 Agostini, F., Cirillo, D., Delli Ponti, R. and Tartaglia, G.G. (2014) SeAMotE: a method for high-throughput motif discovery in nucleic acid sequences. *BMC Genomics* **15**, 925 <https://doi.org/10.1186/1471-2164-15-925>
- 80 Feng, H., Bao, S., Rahman, M.A., Weyn-Vanhenenryck, S.M., Khan, A., Wong, J. et al. (2019) Modeling RNA-binding protein specificity in vivo by precisely registering protein-RNA crosslink sites. *Mol. Cell* **74**, 1189–1204.e6 <https://doi.org/10.1016/j.molcel.2019.02.002>
- 81 Maticzka, D., Lange, S.J., Costa, F. and Backofen, R. (2014) Graphprot: modeling binding preferences of RNA-binding proteins. *Genome Biol.* **15**, R17 <https://doi.org/10.1186/gb-2014-15-1-r17>
- 82 Heller, D., Krestel, R., Ohler, U., Vingron, M. and Marsico, A. (2017) SSHMM: extracting intuitive sequence-structure motifs from high-Throughput RNA-binding protein data. *Nucleic Acids Res.* **45**, 11004–11018 <https://doi.org/10.1093/nar/gkx756>
- 83 Adinolfi, M., Pietrosanto, M., Parca, L., Ausiello, G., Ferrè, F., Ferrè, F. et al. (2019) Discovering sequence and structure landscapes in RNA interaction motifs. *Nucleic Acids Res.* **47**, 4958–4969 <https://doi.org/10.1093/nar/gkz250>
- 84 Licatalosi, D.D., Ye, X. and Jankowsky, E. (2020) Approaches for measuring the dynamics of RNA–protein interactions. *Wiley Interdiscip. Rev. RNA* **11**, 1–23 <https://doi.org/10.1002/wrna.1565>
- 85 Marchese, D., de Groot, N.S., Lorenzo Gotor, N., Livi, C.M. and Tartaglia, G.G. (2016) Advances in the characterization of RNA-binding proteins. *Wiley Interdiscip. Rev. RNA* **7**, 793–810 <https://doi.org/10.1002/wrna.1378>
- 86 Nechay, M. and Kleiner, R.E. (2020) High-throughput approaches to profile RNA-protein interactions. *Curr. Opin. Chem. Biol.* **54**, 37–44 <https://doi.org/10.1016/j.cbpa.2019.11.002>
- 87 Zacco, E., Martin, S.R., Thorogate, R. and Pastore, A. (2018) The RNA-recognition motifs of TAR DNA-binding protein 43 May play a role in the aberrant self-assembly of the protein. *Front. Mol. Neurosci.* **11**, 372 <https://doi.org/10.3389/fnmol.2018.00372>
- 88 Rio, D.C. (2012) Filter-binding assay for analysis of RNA-protein interactions. *Cold Spring Harb. Protoc.* **7**, 1078–1081 <https://doi.org/10.1101/pdb.prot071449>
- 89 Ryder, S.P., Recht, M.I. and Williamson, J.R. (2008) Quantitative analysis of protein-RNA interactions by gel mobility shift. *Methods Mol. Biol.* **488**, 99–115 https://doi.org/10.1007/978-1-60327-475-3_7
- 90 Hellman, L.M. and Fried, M.G. (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat. Protoc.* **2**, 1849–1861 <https://doi.org/10.1038/nprot.2007.249>
- 91 Lakowicz, J.R. (2006) Plasmonics in biology and plasmon-controlled fluorescence. *Plasmonics* **1**, 5–33 <https://doi.org/10.1007/s11468-005-9002-3>
- 92 Littler, D.R., Gully, B.S., Colson, R.N. and Rossjohn, J. (2020) Crystal structure of the SARS-CoV-2 non-structural protein 9, Nsp9. *iScience* **23**, 101258 <https://doi.org/10.1016/j.isci.2020.101258>
- 93 Beier, D.H., Carrocci, T.J., Van Der Feltz, C., Tretbar, U.S., Paulson, J.C., Grabowski, N. et al. (2019) Dynamics of the DEAD-box ATPase Prp5 RecA-like domains provide a conformational switch during spliceosome assembly. *Nucleic Acids Res.* **47**, 10842–10851 <https://doi.org/10.1093/nar/gkz765>
- 94 Meiser, N., Fuks, C. and Hengesbach, M. (2020) Cooperative analysis of structural dynamics in RNA-protein complexes by single-molecule Förster resonance energy transfer spectroscopy. *Molecules* **25**, 2057 <https://doi.org/10.3390/molecules25092057>
- 95 Vo, T., Paul, A., Kumar, A., Boykin, D.W. and Wilson, W.D. (2019) Biosensor-surface plasmon resonance: a strategy to help establish a new generation RNA-specific small molecules. *Methods* **167**, 15–27 <https://doi.org/10.1016/j.ymeth.2019.05.005>
- 96 Sultana, A. and Lee, J.E. (2015) Measuring protein-protein and protein-nucleic acid interactions by biolayer interferometry. *Curr. Protoc. Protein Sci.* **2015**, 19.25.1–19.25.26 <https://doi.org/10.1002/0471140864.ps1925s79>
- 97 Moon, M.H., Hillimire, T.A., Sanders, A.M. and Schneekloth, J.S. (2018) Measuring RNA-ligand interactions with microscale thermophoresis. *Biochemistry* **57**, 4638–4643 <https://doi.org/10.1021/acs.biochem.7b01141>
- 98 Cléry, A., Sohler, T.J.M., Welte, T., Langer, A. and Allain, F.H.T. (2017) switchSENSE: a new technology to study protein-RNA interactions. *Methods* **118–119**, 137–145 <https://doi.org/10.1016/j.ymeth.2017.03.004>
- 99 Feig, A.L. (2009) Studying RNA-RNA and RNA-protein interactions by isothermal titration calorimetry. *Methods Enzymol.* **468**, 409–422 [https://doi.org/10.1016/S0076-6879\(09\)68019-8](https://doi.org/10.1016/S0076-6879(09)68019-8)
- 100 Samatanga, B., Cléry, A., Barraud, P., Allain, F.H.-T. and Jelesarov, I. (2017) Comparative analyses of the thermodynamic RNA binding signatures of different types of RNA recognition motifs. *Nucleic Acids Res.* **45**, 6037–6050 <https://doi.org/10.1093/nar/gkx136>
- 101 Ruminski, D.J., Watson, P.Y., Mahen, E.M. and Fedor, M.J. (2016) A DEAD-box RNA helicase promotes thermodynamic equilibration of kinetically trapped RNA structures in vivo. *RNA* **22**, 416–427 <https://doi.org/10.1261/ma.055178.115>
- 102 Gu, S., Jin, L., Zhang, F., Huang, Y., Grimm, D., Rossi, J.J. et al. (2011) Thermodynamic stability of small hairpin RNAs highly influences the loading process of different mammalian Argonautes. *Proc. Natl Acad. Sci. U.S.A.* **108**, 9208–9213 <https://doi.org/10.1073/pnas.1018023108>
- 103 Amano, R., Takada, K., Tanaka, Y., Nakamura, Y., Kawai, G., Kozu, T. et al. (2016) Kinetic and thermodynamic analyses of interaction between a high-Affinity RNA aptamer and its target protein. *Biochemistry* **55**, 6221–6229 <https://doi.org/10.1021/acs.biochem.6b00748>
- 104 Ray, D., Ha, K.C.H., Nie, K., Zheng, H., Hughes, T.R. and Morris, Q.D. (2017) RNAcompete methodology and application to determine sequence preferences of unconventional RNA-binding proteins. *Methods* **118–119**, 3–15 <https://doi.org/10.1016/j.ymeth.2016.12.003>
- 105 Beckmann, B.M., Castello, A. and Medenbach, J. (2016) The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions. *Pflugers Arch. Eur. J. Physiol.* **468**, 1029–1040 <https://doi.org/10.1007/s00424-016-1819-4>
- 106 Sanchez de Groot, N., Armaos, A., Graña-Montes, R., Alriquet, M., Calloni, G., Vabulas, R.M. et al. (2019) RNA structure drives interaction with proteins. *Nat. Commun.* **10**, 1–13 <https://doi.org/10.1038/s41467-019-10923-5>
- 107 Schlundt, A., Tants, J.N. and Sattler, M. (2017) Integrated structural biology to unravel molecular mechanisms of protein-RNA recognition. *Methods* **118–119**, 119–136 <https://doi.org/10.1016/j.ymeth.2017.03.015>

- 108 Hennig, J. and Sattler, M. (2014) The dynamic duo: combining NMR and small angle scattering in structural biology. *Protein Sci.* **23**, 669–682 <https://doi.org/10.1002/pro.2467>
- 109 Mahieu, E. and Gabel, F. (2018) Biological small-angle neutron scattering: recent results and development. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 715–726 <https://doi.org/10.1107/S2059798318005016>
- 110 Lapinaite, A., Carlomagno, T. and Gabel, F. (2020) Small-angle neutron scattering of RNA–protein complexes. *Methods Mol. Biol.* **2113**, 165–188 https://doi.org/10.1007/978-1-0716-0278-2_13
- 111 Delhommel, F., Gabel, F. and Sattler, M. (2020) Current approaches for integrating solution NMR spectroscopy and small-angle scattering to study the structure and dynamics of biomolecular complexes. *J. Mol. Biol.* **432**, 2890–2912 <https://doi.org/10.1016/j.jmb.2020.03.014>
- 112 Gräwert, T.W. and Svergun, D.I. (2020) Structural modeling using solution small-angle X-ray scattering (SAXS). *J. Mol. Biol.* **432**, 3078–3092 <https://doi.org/10.1016/j.jmb.2020.01.030>
- 113 Patel, T.R., Chojnowski, G., Astha, Koul, A., McKenna, S.A. and Bujnicki, J.M. (2017) Structural studies of RNA–protein complexes: a hybrid approach involving hydrodynamics, scattering, and computational methods. *Methods* **118–119**, 146–162 <https://doi.org/10.1016/j.ymeth.2016.12.002>
- 114 Yadav, D.K. and Lukavsky, P.J. (2016) NMR solution structure determination of large RNA–protein complexes. *Prog. Nucl. Magn. Reson. Spectrosc.* **97**, 57–81 <https://doi.org/10.1016/j.pnmrs.2016.10.001>
- 115 Ke, A. and Doudna, J.A. (2004) Crystallization of RNA and RNA–protein complexes. *Methods* **34**, 408–414 <https://doi.org/10.1016/j.ymeth.2004.03.027>
- 116 Luo, X., Wang, X., Gao, Y., Zhu, J., Liu, S., Gao, G. et al. (2020) Molecular mechanism of RNA recognition by zinc-finger antiviral protein. *Cell Rep.* **30**, 46–52 <https://doi.org/10.1016/j.celrep.2019.11.116>
- 117 Partin, A.C., Zhang, K., Jeong, B.C., Herrell, E., Li, S., Chiu, W. et al. (2020) Cryo-EM structures of human drosha and DGCR8 in complex with primary microRNA. *Mol. Cell* **78**, 411–422 <https://doi.org/10.1016/j.molcel.2020.02.016>
- 118 Haahr Larsen, A., Wang, Y., Bottaro, S., Grudin, S., Arleth, L. and Lindorff-Larsen, K. (2020) Combining molecular dynamics simulations with small-angle X-ray and neutron scattering data to study multi-domain proteins in solution. *PLoS Comput. Biol.*, **16**, e1007870. <https://doi.org/10.1371/journal.pcbi.1007870>
- 119 Shapiro, B.A. and Le Grice, S.F.J. (2016) Advances in RNA structure determination. *Methods* **103**, 1–3 <https://doi.org/10.1016/j.ymeth.2016.06.006>
- 120 Bonomi, M., Heller, G.T., Camilloni, C. and Vendruscolo, M. (2017) Principles of protein structural ensemble determination. *Curr. Opin. Struct. Biol.* **42**, 106–116 <https://doi.org/10.1016/j.sbi.2016.12.004>