APPLICATIONS OF NEXT-GENERATION SEQUENCING

Protein–RNA interactions: new genomic technologies and perspectives

Julian König, Kathi Zarnack, Nicholas M. Luscombe and Jernej Ule

Abstract | RNA-binding proteins are key players in the regulation of gene expression. In this Progress article, we discuss state-of-the-art technologies that can be used to study individual RNA-binding proteins or large complexes such as the ribosome. We also describe how these approaches can be used to study interactions with different types of RNAs, including nascent transcripts, mRNAs, microRNAs and ribosomal RNAs, in order to investigate transcription, RNA processing and translation. Finally, we highlight current challenges in data analysis and the future steps that are needed to obtain a quantitative and high-resolution picture of protein–RNA interactions on a genome-wide scale.

During and after transcription, RNAs are subject to multiple processing and regulatory steps that are coordinated by RNA-binding proteins (RBPs)[1,2]. Therefore, to understand the fate and function of RNA molecules, a key task is to map protein–RNA interactions and to determine their effects on the transcriptome. In recent years, there has been great progress in the field of ribonomics, which uses genome-wide tools to study how the interactions of RNAs and proteins modulate co-transcriptional and post-transcriptional regulation of gene expression.

The first ribonomic approaches combined RNA immunoprecipitation with differential display or microarray analysis (RIP–chip) to identify RNAs that are bound by specific RBPs[3–5]. However, these methods were limited to stable ribonucleoprotein particles (RNPs) and were prone to detecting non-specific interactions[6]. Moreover, the resulting data were of low resolution, as the binding site in the co-purified RNA molecule remained unresolved. Therefore, defining precise RBP binding sites and reducing the number of false positives have been primary challenges for experimental ribonomics.

To identify the positions of protein–RNA interactions with a higher resolution and specificity, a method known as ultraviolet (UV) crosslinking and immunoprecipitation (CLIP) was developed[7,8]. CLIP combines UV crosslinking of RBPs to their cognate RNA molecules with rigorous purification schemes. Recently, CLIP has been coupled to high-throughput sequencing, which has allowed comprehensive genome-wide studies. In addition, CLIP-related techniques have been developed to determine RNA interactions with larger complexes such as the ribosome. With these developments at hand, we are now entering an exciting era of broad applications of ribonomic methods. In this Progress article, we describe these recent advances in ribonomic techniques. We also introduce approaches for data analysis, highlight the major challenges in this field and conclude with an outlook on future developments.

CLIP: landscapes of RNA binding

Most RBPs recognize short, degenerate RNA motifs, and therefore they often bind at several sites on most RNAs. Thus, it is not sufficient to determine whether a protein interacts with a particular RNA, but it is important to define the full landscape of interactions of the protein with the RNA. CLIP is a state-of-the-art technology that allows users to define these RNA landscapes[9]. Here, we describe the basic concepts of CLIP and introduce recent developments.

The beginnings of CLIP. CLIP relies on the principle that precise and stringent mapping of binding sites is achieved by preserving the *in vivo* protein–RNA interactions by irradiation of living cells or tissue with ultraviolet C (UVC) light[7,8]. The UVC light induces the formation of covalent crosslinks only at sites of direct contact between proteins and RNA. On cell lysis, the protein–RNA complex is immunoprecipitated with an antibody that is specific for the protein of interest (FIG. 1). If no antibody is available, the RBP can alternatively be fused to an epitope tag, which is then expressed as a transgene for affinity purification[10–12]. The co-purified RNA molecules are reverse-transcribed and amplified with the aid of 5′ and 3′ adaptors. In the original CLIP protocol, individual clones of the resulting cDNAs were subjected to Sanger sequencing. The resulting sequences were then mapped to the reference genome to reveal the sites of protein binding within the corresponding transcripts.

The accuracy of CLIP was demonstrated in a study of NOVA-dependent splicing regulation in the brain[7]. These first CLIP experiments revealed how NOVA binds at different positions to silence or to enhance exon inclusion. Other applications of CLIP have included uncovering a role for the heterogeneous nuclear ribonucleoprotein HNRNPA1 in microRNA (miRNA) processing[13] and identifying actively transported transcripts in fungal filaments[14].

CLIP goes high-throughput. To map protein–RNA binding sites more comprehensively, Licatalosi and co-workers[15] replaced the Sanger method with high-throughput sequencing, which enables millions of sequences to be determined in a single run. This approach, which is known as high-throughput sequencing of CLIP cDNA library (HITS-CLIP or CLIP–seq[16]), provides more comprehensive binding information (FIG. 1). The power of coupling CLIP and high-throughput sequencing was first demonstrated by an analysis of NOVA-dependent RNA processing in the brain[15]. The greater sequencing depth compared with Sanger sequencing provided new insights into the NOVA-dependent splicing regulation and also led to the discovery

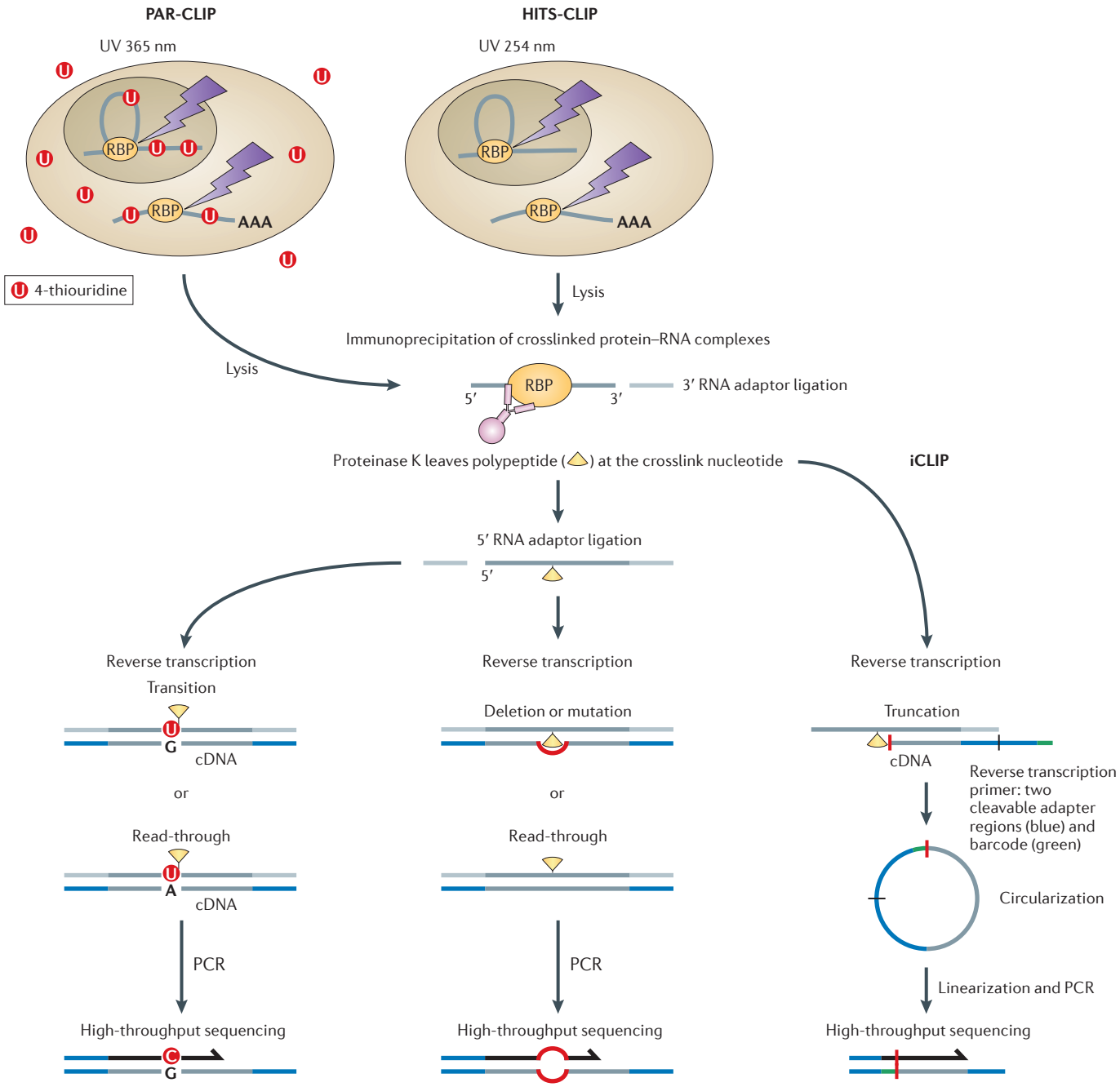

Figure 1 | **Comparison of HITS-CLIP and its latest variants, PAR-CLIP and iCLIP.** For high-throughput sequencing of RNA isolated by ultraviolet (UV) crosslinking and immunoprecipitation (HITS-CLIP), living cells or tissue samples are irradiated with UV light at a wavelength of 254 nm (shown in the centre of the figure). This induces the formation of covalent crosslinks between proteins and RNA, which are restricted to sites of direct contact. The cells are then lysed and RNA is partially digested to an approximate length of 30–50 nucleotides. Next, the protein–RNA complex is immunoprecipitated with an antibody that is specific for the protein of interest. After stringent washing, the RNA is radioactively labelled and an adaptor is ligated to the 3′ end of the RNA. Further purification is achieved through denaturing gel electrophoresis and transfer to a nitrocellulose membrane, which removes nonspecific RNAs. The radioactive label on the RNA is used to guide the excision of the protein–RNA complex from the membrane. The protein is then removed from the RNA by proteinase K digestion. An adaptor is ligated to the 5′ end, the RNA is reverse transcribed, and the resulting cDNAs are PCR amplified with primers that are complementary to the 5′ and 3′ adaptor regions. The resulting cDNA library is subjected to high-throughput sequencing. For photoactivatable ribonucleoside-enhanced CLIP (PAR-CLIP) (shown on the left of the figure), cells are fed with 4-thiouridine, which becomes incorporated into newly transcribed RNA. This allows crosslinking with UV light at 365 nm. During reverse transcription, the nucleoside analogue causes a base transition that can be used to pinpoint the crosslinked nucleotide. In the individual nucleotide resolution CLIP (iCLIP) protocol (shown on the right of the figure), crosslinking is carried out as in HITS-CLIP at 254 nm. However, in order to capture cDNAs that truncate at the peptide that remains at the crosslinked nucleotide after proteinase K digestion, the 5′ adaptor is added after reverse transcription. This is achieved through priming reverse transcription with an oligonucleotide that contains the 3′, as well as the 5′, adaptor region followed by circularization of the generated cDNAs.

of an unexpected role of NOVA in 3′ end processing. HITS-CLIP is also being used to study the tripartite complex between the Argonaute proteins, miRNAs and their target transcripts[17–19]. Although the direct pairing of an miRNA with its target mRNA cannot yet be deduced from these data, the detection of Argonaute binding sites in both miRNAs and mRNAs enabled the discovery of endogenous mRNA target sites. Recently, an important step was made towards direct monitoring of inter-RNA interactions within tripartite complexes: a method called crosslinking, ligation and sequencing of hybrids (CLASH) was developed, which exploits the formation of intermolecular RNA ligation events. As a proof of principle, this method was used to map *in vivo* RNA–RNA contact sites of small nucleolar RNAs (snoRNAs) with precursor ribosomal RNAs (pre-rRNAs) during ribosome biogenesis[20].

Advancing towards nucleotide resolution. In the traditional CLIP protocol, the resolution of binding site detection mostly corresponds to the length of the fragmented RNAs. However, Granneman and colleagues[11] showed that crosslink-induced point mutations and deletions can be used to identify the crosslink sites of RBPs within snoRNAs[11]. Recently, two approaches introduced new strategies that are based on modified crosslinking or library-preparation protocols to identify the crosslink sites on a genome-wide scale[12,21].

In the photoactivatable ribonucleoside-enhanced CLIP (PAR-CLIP) approach[12], photoactivatable nucleotide analogues such as 4-thiouridine (4-SU) or 6-thioguanosine (6-SG) are used (FIG. 1), which can be efficiently crosslinked with ultraviolet A (UVA) light (at a wavelength of 365 nm). The nucleotide analogues are readily taken up by cells and become incorporated into newly synthesized transcripts. Importantly, they lead to a base transition at the crosslink site during reverse transcription. Therefore, mutation analysis of the resulting cDNA sequences can be used to pinpoint crosslink sites at nucleotide resolution (discussed below). This method was successful in identifying crosslink sites of pumilio homologue 2 (PUM2), quaking (QKI), insulin-like growth factor 2 mRNA-binding protein 1 (IGF2BP1), IGF2BP2 and IGF2BP3, the Argonaute proteins and HUR (also known as ELAVL1) in HEK293 cells[12,22]. Similarly, analysis of point mutations and deletions was used for the genome-wide identification of crosslink sites from HITS-CLIP data[22,23].

An alternative method for achieving nucleotide resolution is known as individual nucleotide resolution CLIP (iCLIP)[21]. This method is based on the concept that reverse transcription can stop at nucleotides that are crosslinked to the peptides that remain after proteinase K digestion[24]. However, the truncated cDNAs that are produced would lack the 5′ adaptor region that is required for PCR amplification and would be lost during the standard CLIP library preparation. To capture these truncated cDNAs, iCLIP uses an alternative strategy for adaptor ligation and reverse transcription, replacing one of the intermolecular RNA ligation steps with an intramolecular cDNA circularization (FIG. 1). Importantly, sequencing the truncated cDNAs provides direct identification of the crosslink position, which is located one nucleotide upstream of the truncation site. As a demonstration of this method, iCLIP was used to resolve the footprint of adjacent HNRNPC binding sites within uridine tracts[21].

The goal: quantitative CLIP analysis. Owing to the small amount of starting material and the numerous steps at which material can be lost, the number of CLIP cDNAs generated from crosslinked RNA is an important limiting factor. As a consequence, the resulting cDNA libraries rarely contain the full range of RNA binding sites. An additional concern is that, owing to biases in the PCR amplification step, libraries can result in thousands of sequences that originated from a single cDNA. This can lead to data of limited complexity and informational content and can distort the quantitative analysis of protein–RNA interactions.

One way to avoid amplification artefacts is to count identical sequences only once; however, this approach reduces the dynamic range of the resulting data. A more sophisticated way to control for library complexity is to use a randomized sequence in the adaptor or reverse transcription primer. This sequence, which is referred to as a randomer, a degenerate or a random barcode, can be used to discriminate independent cDNAs from PCR duplicates[17,21,25]. For example, two sequences that map to the same genomic location and that share an identical randomer are treated as PCR duplicates, whereas they are identified as two unique cDNAs if they possess different randomers. It was recently shown that this approach can also improve the accuracy of other high-throughput sequencing methods[25], although random barcoding may also have its limitations.

Considerations and applications for CLIP. All CLIP variants — HITS-CLIP, PAR-CLIP and iCLIP — produce data of high quality and precision. However, the library preparation protocols for these techniques require a large number of enzymatic steps that potentially affect binding site detection. For example, it is important to optimize the conditions of partial RNase digestion, as overdigestion can decrease the number of identified sites[22]. Furthermore, it has been shown that the use of different RNA ligases can influence the cloning of short RNAs[26], and it remains to be seen how the choice of ligase influences the different CLIP protocols.

The crosslinking efficiency with UVC (HITS-CLIP and iCLIP) or UVA (PAR-CLIP) varies for different proteins, and the optimal protocol needs to be experimentally determined individually for the protein of interest[22]. However, the application of PAR-CLIP is currently limited to cultured cells that can efficiently incorporate nucleoside analogues. Conversely, the timing of nucleoside application provides the opportunity to restrict crosslinking to transcripts that were newly synthesized, promising new insights into RBP binding to nascent transcripts. CLIP has already been used to study RBP binding to diverse types of transcripts, including introns, mRNAs, miRNAs, snoRNAs and rRNAs[11,15,17]. Moreover, it should be possible to use CLIP technologies in any living organism, and they have already been used in yeast, fungi, worms and mammals[7,11,14,18].

Studying larger RNP complexes

CLIP technologies determine the direct contacts between individual RBPs and their cognate RNAs. In some cases, however, it is desirable to investigate the interactions of larger complexes with RNAs. Two recent approaches use the purification of intact ribosomes or RNA polymerase complexes to monitor translation and transcription on a genome-wide scale.

Footprinting the ribosome. Through an approach that is termed ribosome profiling, the Weissman laboratory provided high-resolution analysis of translation on a genome-wide scale[27] (FIG. 2a). This was achieved by stalling ribosomes on the transcripts using cycloheximide treatment, followed by cell lysis, RNase treatment and purification of the RNA fragments that were protected by the ribosome *in vivo*. The fragments were then subjected to circularization-based library preparation and high-throughput sequencing. The resulting

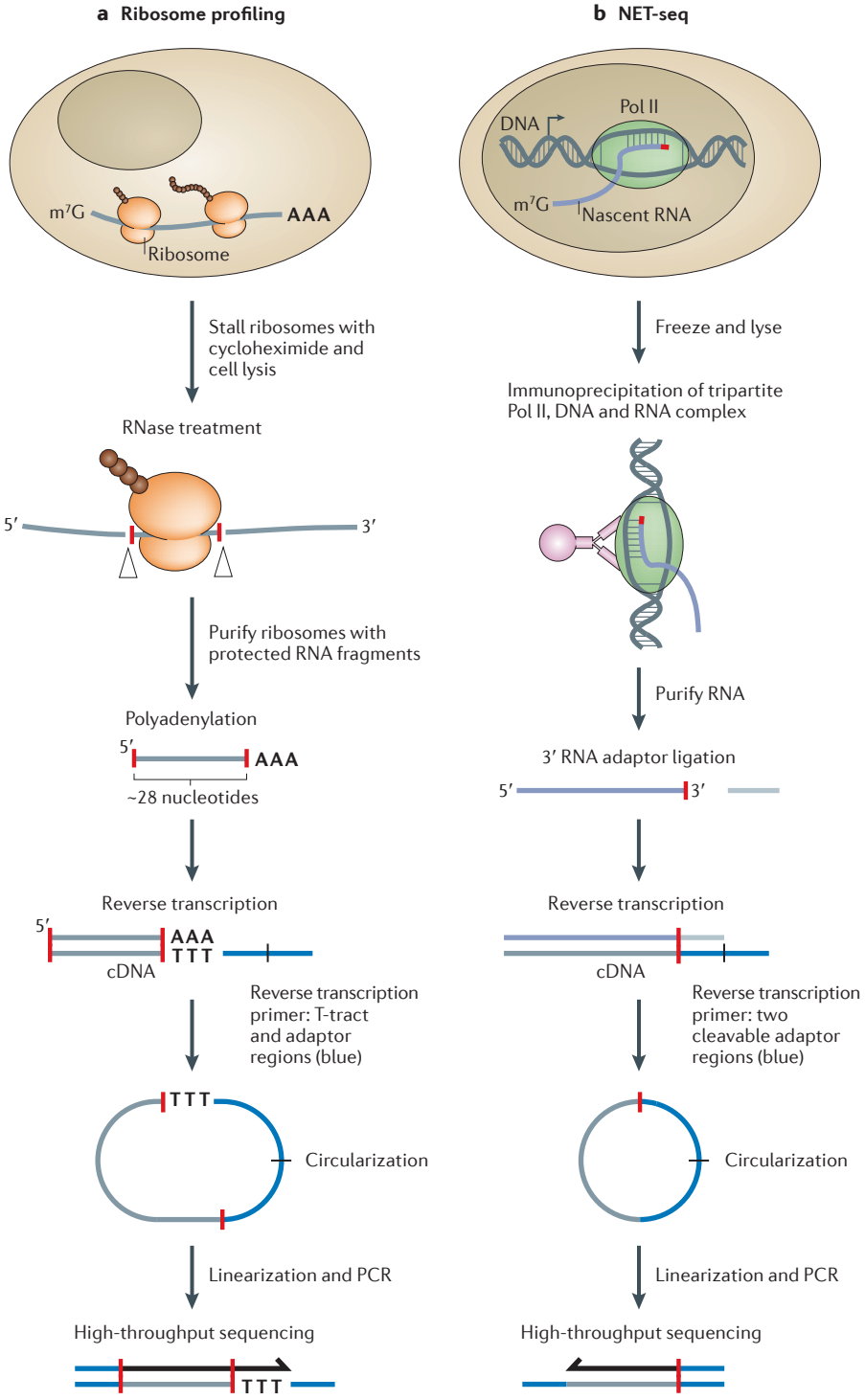

Figure 2 | **Ribonomic methods to study transcription and translation. a** | For ribosome profiling, ribosomes are stalled on the translated RNAs through cycloheximide treatment. After cell lysis, the RNA that is not covered by the ribosomes is degraded with RNase. The ribosomes are then purified together with the protected RNA fragments. The RNA fragments are then polyadenylated, which allows priming of reverse transcription and circularization-based cDNA library preparation. **b** | For native elongating transcript sequencing (NET-seq), cells are flash-frozen and lysed. The tripartite complex of RNA polymerase II (Pol II), DNA and nascent RNA is immunopurified. The nascent RNAs are separated and an adaptor is ligated to their 3′ ends. Upon reverse transcription, cDNA libraries are prepared using a circularization-based approach. High-throughput sequencing of these libraries provides information about the position of Pol II at nucleotide resolution.

ribosome footprints were used to obtain quantitative information about translation rates and ribosome density within transcripts. Notably, the resolution of the data is precise enough to gain information about the translated reading frame. However, it is important to note that this information was not obtained at the level of individual codons, but was inferred from an average signal over the complete open reading frame. Initially developed in yeast and used to study translational changes during the stress response[27], ribosome profiling was later adapted for use with mammalian cell lines[28]. In this system, research using ribosome profiling suggested that miRNAs predominantly function through the destabilization of target transcripts, rather than by silencing translation[28].

Chasing the RNA polymerase. A genome-wide view of transcription can be obtained from high-throughput sequencing of DNA fragments that are crosslinked to RNA polymerase II (Pol II) — a technique that is known as Pol II chromatin immunoprecipitation followed by sequencing (ChIP–seq) — or from approaches that are based on nuclear run-on, such as global run-on sequencing (GRO-seq)[29]. In order to monitor transcriptional states of unperturbed cells with a high resolution and strand specificity, the Weissman laboratory developed an approach to study Pol II binding to nascent RNAs, which is referred to as native elongating transcript sequencing (NET-seq)[30]. Without prior crosslinking, this technique combines Pol II affinity purification with sequencing of the 3′ ends of the co-purified RNAs, and so provides insights into transcription at single-nucleotide resolution (FIG. 2b). This approach is feasible owing to the high stability of the ternary complex that is formed between Pol II, the transcribed DNA and the nascent RNA. Churchman and Weissmann[30] exploited the strand information in the data to reveal a link between histone H4 acetylation and antisense transcription at promoters. In addition, the study investigated Pol II backtracking and nucleosome-induced pausing, which reflects the broad range of applications for NET-seq[30]. This technique promises to be a valuable tool for researchers who are interested in all aspects of transcription.

Data analysis and interpretation

The large amounts of data generated by ribonomic approaches require considerable computational efforts for biological interpretation. The first level of analysis is genomic mapping of the sequence reads, followed by

a second level of clustering and normalization to identify highly occupied binding sites. At the third level, the binding sites are integrated with functional information in order to deduce general regulatory principles. We discuss these different layers of data analysis and interpretation below.

Mapping the sequence reads. Fast and efficient alignment algorithms such as Bowtie[31] or Burrows–Wheeler alignment (BWA)[32] are standard tools for mapping high-throughput sequencing reads to the genome. However, if RBPs bind mature RNAs, the cDNA sequences often span exon–exon junctions. Therefore, mapping of sequence reads that are produced by CLIP approaches or ribosome profiling should ideally include either the use of splicing-aware algorithms such as TopHat[33] or direct alignment to processed transcripts. Another challenge is the mapping of sequences to genes that are present in multiple copies in the genome, such as small nuclear RNAs (snRNAs), rRNAs and snoRNAs. One solution is to use non-redundant databases that offer consensus sequences for multi-copy genes[11]. Another option is to allow mapping to multiple positions in the genome, but care needs to be taken when interpreting such data. To account for sequencing errors and crosslink-induced point mutations[11,12], it can be advantageous to allow one or more mismatches in the alignments. To capture crosslink-induced deletions, algorithms such as Novoalign, segemehl[34] or genomic short-read nucleotide alignment program (GSNAP)[35], which allow gapped alignments, should be used[11,23]. A valuable resource is the dedicated servers and databases that are available for the mapping and analyses of CLIP data that are generated by the different protocols[36–38].

Identification of binding sites. The high stringency of library preparation achieved with the different CLIP approaches is documented by the low number of nonspecific reads in control experiments, which use knockout tissue, omit the antibody or omit UV crosslinking. Thus, with proper purification of the protein–RNA complex, the vast majority of CLIP reads represent protein–RNA interaction sites. The occupancy of the RBP at these sites varies considerably; binding sites with low occupancy usually outnumber highly occupied binding sites. Importantly, highly occupied binding sites appear as clusters of reads when the CLIP library is of sufficient complexity. Approaches for identifying such read clusters involve the analysis of replicate experiments to ensure that binding at a given site is reproducible[15] or the calculation of significant enrichment over the background signal in surrounding areas on the same gene[16,21,22]. In this context, it is important to keep in mind that CLIP read counts are not necessarily a direct measure of RBP affinity, as they can be affected by other factors, such as the half-life of the bound RNA region or the crosslinking efficiency of a given sequence.

In addition to read-cluster identification, several different approaches have been implemented that directly identify the nucleotide that is crosslinked (FIG. 3a). It is important to note that the crosslinked nucleotide may not always reside within the binding site of the protein, regardless of which CLIP technology is used. For example, binding motifs of NOVA are mainly enriched in the sequences immediately surrounding, but not including, the crosslinked nucleotide[23]. The potential for such shifts has to be considered when using the position of crosslink nucleotides to investigate the sequence of the RNA motifs that are required for the high-affinity protein binding. Common tools for identifying binding motifs such as motif em for motif elicitation (MEME) and Phylogibbs are complemented by approaches that search for the enrichment of certain *k*-mers in the vicinity of read clusters or crosslink nucleotides[10,16,39,40].

Integrating functional information. Several CLIP studies indicate that many RBPs show high-affinity binding to thousands of different positions in the transcriptome. Therefore, it is likely that only a subset of the interactions is associated with specific functions. In order to identify functional interactions, the physical maps of protein–RNA interactions can be integrated with other genome-wide data sets that provide functional information about the RBP. For example, the integration of binding data with information from splice-junction microarrays or RNA-seq can be used to generate RNA maps, demonstrating position-dependent splicing regulation by RBPs[41,42]. So far, such maps have been successfully applied to determine the functional binding sites of several splicing regulators, including NOVA, FOX2 and HNRNPC[15,16,21,41]. Similarly, it will be interesting to study the concerted binding of different RBPs in more detail. For example, recent PAR-CLIP studies indicated that HUR binding sites in the 3′ UTR are enriched in the vicinity of Argonaute miRNA complexes in the same region[43,44]. Finally, a promising future direction will be to combine CLIP with other emerging ribonomic assays, such as ribosome profiling, which would allow the direct monitoring of the effect of RBP binding on translation.

Future directions

To date, CLIP studies have mainly been used for qualitative descriptions of RBP binding, and the generation of reliable quantitative information on RBP binding remains a major challenge. We expect to see further improvements in CLIP library preparation that will increase the complexity of cDNA libraries and allow better

Glossary

Argonaute proteins
Core components of the RNA-mediated silencing pathways. They provide the platform for target mRNA recognition by small non-coding RNAs and harbour the catalytic activity for mRNA cleavage.

Differential display
A PCR-based approach that was used to study differences in RNA populations. It has now been superseded by microarray and RNA sequencing approaches.

Global run-on sequencing
(GRO-seq). A technique that combines nuclear run-on assays with high-throughput sequencing to obtain genome-wide information about active transcription.

Heterogeneous nuclear ribonucleoprotein
(HNRNP). The core protein components of heterogeneous nuclear ribonucleoprotein particles that associate with all nascent transcripts. They are involved in diverse aspects of post-transcriptional regulation.

***k*-mers**
Nucleic acid sequences with a number of nucleotides of length *k*.

NOVA
A regulator of a biologically coherent set of RNAs important for synaptic function. It is involved in the neurological disorder paraneoplastic opsoclonus myoclonus ataxia.

Ribonomics
The genome-scale study of protein–RNA interactions and their functional consequences.

Ribonucleoprotein particles
(RNPs). Complexes consisting of protein and RNA components.

Small nuclear RNAs
(snRNAs). A class of non-coding RNAs that are found in the nucleus of eukaryotic cells and that constitute core components of all subunits of the spliceosome.

Small nucleolar RNAs
(snoRNAs). A class of small non-coding RNAs that are involved in guiding chemical modifications of other RNAs, such as ribosomal or transfer RNAs.

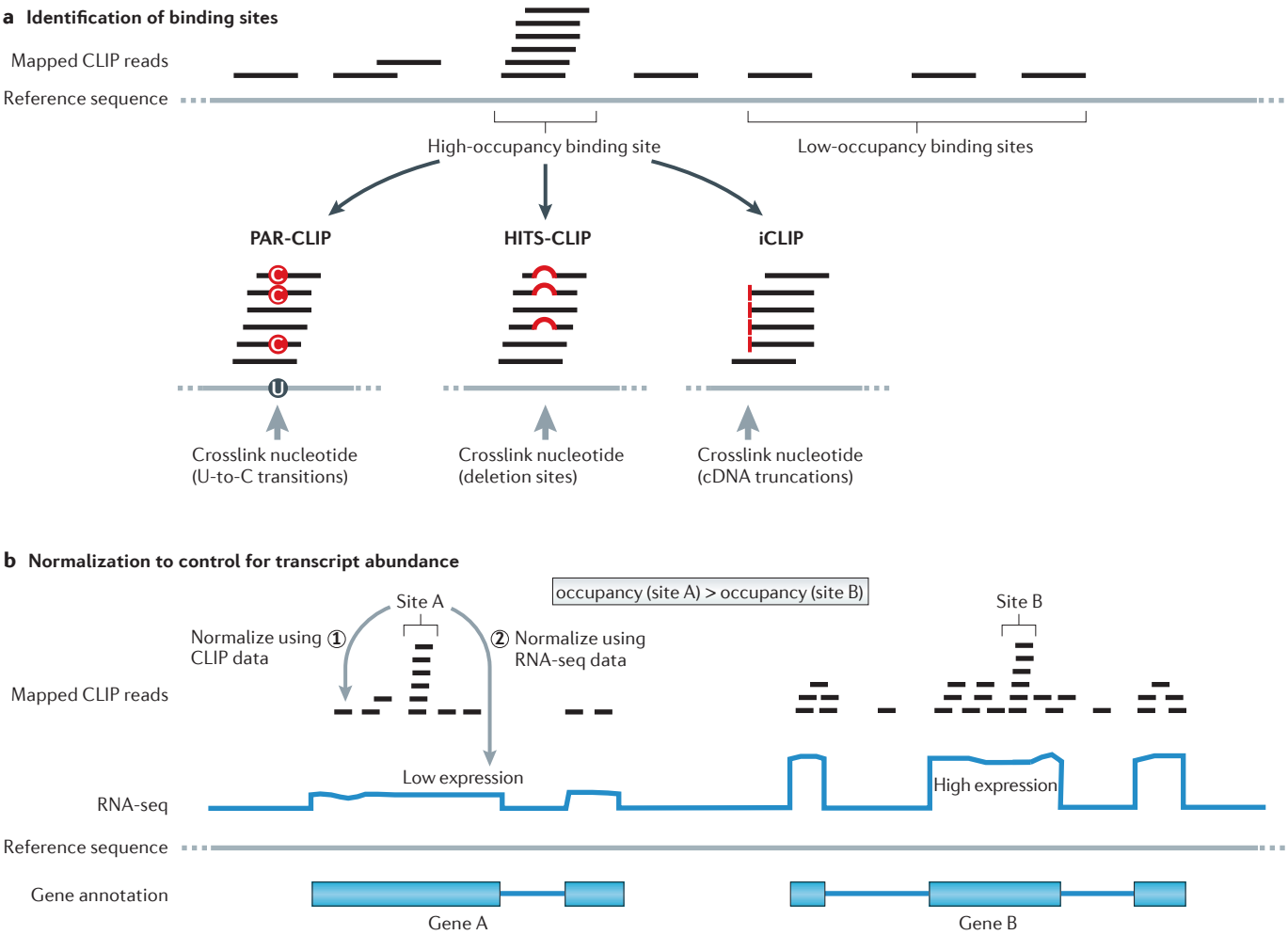

Figure 3 | **Identification of binding sites and normalization. a** | High-affinity binding sites appear as clusters of ultraviolet crosslinking and immunoprecipitation (CLIP) reads. In photoactivatable ribonucleoside-enhanced CLIP (PAR-CLIP), the crosslink nucleotide can be identified through U-to-C transitions, and in high-throughput sequencing of RNA isolated by CLIP (HITS-CLIP) through deletion sites. In individual nucleotide resolution CLIP (iCLIP), the crosslink nucleotide is located one nucleotide upstream of the truncation sites. **b** | A schematic description of different normalization strategies to correct for transcript abundance is shown. Normalization can be carried out (step 1) based on the overall protein binding within the transcript or (step 2) by incorporating external information on transcript abundance using methods such as RNA sequencing (RNA-seq).

quantification of RBP binding to individual RNA sites. It is also clear that read counts depend on the expression level of the corresponding transcript. Therefore, normalization of CLIP data will be required before binding sites can be compared across the full transcriptome (FIG. 3b). This could be achieved, for example, by normalizing to the average CLIP count within the transcript or by using expression information obtained from RNA-seq experiments. Using RNA-seq has proved to be useful for analyses of ribosome profiling data[28], and was also recommended by a recent study comparing several CLIP normalization strategies[22]. However, normalization to total CLIP counts within transcripts might be more applicable to nuclear RBPs that bind pre-mRNAs, because these are not efficiently quantified by standard RNA-seq techniques. In addition, bioinformatic approaches need to be developed to account for the effects of local sequence environment on the efficiency of protein–RNA crosslinking. First efforts in this direction have been made[22,23], but more analyses are needed to fully understand the sequence biases at the crosslink sites that have been identified by the different CLIP protocols.

In summary, the time is ripe to take CLIP from a qualitative assay to a quantitative tool. A potential advance in the near future would be the combination of CLIP with single-molecule RNA sequencing[45], which could monitor stalling at the crosslink nucleotide in real time. Finally, in parallel with the experimental advances, sophisticated computational analysis methods will need to be developed: for example, to model combinatorial RNA binding of multiple RBPs. These advances will take us closer to the goal of obtaining a complete picture of the diverse protein–RNA complexes in the cell.

Julian König and Jernej Ule are at the Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK.

Kathi Zarnack and Nicholas M. Luscombe are at the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL–EBI), Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK.

Nicholas M. Luscombe is also at the Okinawa Institute for Science and Technology Graduate University, 1919-1 Tancha, Onna-son, Kunigami-gun, Okinawa 904-0495, Japan.

Correspondence to J.U.
e-mail: jule@mrc-lmb.cam.ac.uk
doi:10.1038/nrg3141
Corrected online 31 January 2012

1. Moore, M. J. From birth to death: the complex lives of eukaryotic mRNAs. *Science* **309**, 1514–1518 (2005).
2. Keene, J. D. RNA regulons: coordination of post-transcriptional events. *Nature Rev. Genet.* **8**, 533–543 (2007).
3. Trifillis, P., Day, N. & Kiledjian, M. Finding the right RNA: identification of cellular mRNA substrates for RNA-binding proteins. *RNA* **5**, 1071–1082 (1999).
4. Brooks, S. A. & Rigby, W. F. Characterization of the mRNA ligands bound by the RNA binding protein hnRNP A2 utilizing a novel *in vivo* technique. *Nucleic Acids Res.* **28**, e49 (2000).
5. Tenenbaum, S. A., Carson, C. C., Lager, P. J. & Keene, J. D. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc. Natl Acad. Sci.* **97**, 14085–14090 (2000).
6. Mili, S. & Steitz, J. A. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* **10**, 1692–1694 (2004).
7. Ule, J. *et al.* CLIP identifies NOVA-regulated RNA networks in the brain. *Science* **302**, 1212–1215 (2003).
8. Ule, J., Jensen, K., Mele, A. & Darnell, R. B. CLIP: A method for identifying protein–RNA interaction sites in living cells. *Methods* **37**, 376–386 (2005).
9. Darnell, R. B. HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley Interdiscip. Rev. RNA* **1**, 266–286 (2010).
10. Wang, Z. *et al.* iCLIP predicts the dual splicing effects of TIA-RNA interactions. *PLoS Biol.* **8**, e1000530 (2010).
11. Granneman, S., Kudla, G., Petfalski, E. & Tollervey, D. 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. USA* **106**, 9613–9618 (2009).
12. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
13. Guil, S. & Caceres, J. F. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nature Struct. Mol. Biol.* **14**, 591–596 (2007).
14. König, J. *et al.* The fungal RNA-binding protein Rrm4 mediates long-distance transport of *ubi1* and *rho3* mRNAs. *EMBO J.* **28**, 1855–1866 (2009).
15. Licatalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469 (2008).
16. Yeo, G. W. *et al.* An RNA code for the FOX2 splicing regulator revealed by mapping RNA–protein interactions in stem cells. *Nature Struct. Mol. Biol.* **16**, 130–137 (2009).
17. Chi, S. W., Zang, J. B., Mele, A. & Darnell, R. B. Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps. *Nature* **460**, 479–486 (2009).
18. Zisoulis, D. G. *et al.* Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*. *Nature Struct. Mol. Biol.* **17**, 173–179 (2010).
19. Leung, A. K. *et al.* Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nature Struct. Mol. Biol.* **18**, 237–244 (2011).
20. Kudla, G., Granneman, S., Hahn, D., Beggs, J. D. & Tollervey, D. Cross-linking, ligation, and sequencing of hybrids reveals RNA–RNA interactions in yeast. *Proc. Natl Acad. Sci. USA* **108**, 10010–10015 (2011).
21. König, J. *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature Struct. Mol. Biol.* **17**, 909–915 (2010).
22. Kishore, S. *et al.* A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nature Methods* **8**, 559–564 (2011).
23. Zhang, C. & Darnell, R. B. Mapping *in vivo* protein–RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nature Biotechnol.* **29**, 607–614 (2011).
24. Urlaub, H., Hartmuth, K. & Lührmann, R. A two-tracked approach to analyze RNA-protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles. *Methods* **26**, 170–181 (2002).
25. Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nature Methods* 20 Nov 2011 (doi:10.1038/nmeth.1778).
26. Hafner, M. *et al.* RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA* **17**, 1697–1712 (2011).
27. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
28. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835–840 (2010).
29. Fuda, N. J., Ardehali, M. B. & Lis, J. T. Defining mechanisms that regulate RNA polymerase II transcription *in vivo*. *Nature* **461**, 186–192 (2009).
30. Churchman, L. S. & Weissman, J. S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368–373 (2011).
31. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
32. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
33. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics* **25**, 1105–1111 (2009).
34. Hoffmann, S. *et al.* Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput. Biol.* **5**, e1000502 (2009).
35. Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873–881 (2010).
36. Khorshid, M., Rodak, C. & Zavolan, M. CLIPZ: a database and analysis environment for experimentally determined binding sites of RNA-binding proteins. *Nucleic Acids Res.* **39**, D245–D252 (2011).
37. Corcoran, D. L. *et al.* PARalyzer: Definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol.* **12**, R79 (2011).
38. Yang, J. H. *et al.* starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP–seq and Degradome-seq data. *Nucleic Acids Res.* **39**, D202–D209 (2011).
39. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
40. Siddharthan, R., Siggia, E. D. & van Nimwegen, E. PhyloGibbs: a Gibbs sampling motif finder that incorporates phylogeny. *PLoS Comput. Biol.* **1**, e67 (2005).
41. Ule, J. *et al.* An RNA map predicting NOVA-dependent splicing regulation. *Nature* **444**, 580–586 (2006).
42. Witten, J. T. & Ule, J. Understanding splicing regulation through RNA splicing maps. *Trends Genet.* **27**, 89–97 (2011).
43. Lebedeva, S. *et al.* Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Mol. Cell* **43**, 340–352 (2011).
44. Mukherjee, N. *et al.* Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol. Cell* **43**, 327–339 (2011).
45. Schadt, E. E., Turner, S. & Kasarskis, A. A window into third-generation sequencing. *Hum. Mol. Genet.* **19**, R227–R240 (2010).

Acknowledgments
This work was supported by the Medical Research Council, the European Molecular Biology Laboratory (grant number U105185858), the European Research Council (206726-CLIP) and by a Human Frontiers Science Program Long-Term fellowship and an EMBL EIPOD fellowship to J.K. and K.Z., respectively.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Nicholas M. Luscombe's homepage: http://www.ebi.ac.uk/~luscombe
Jernej Ule's homepage: http://www2.mrc-lmb.cam.ac.uk/groups/jule
CLIP forum: http://megazord.rockefeller.edu/public/forum
CLIPZ: http://www.clipz.unibas.ch
GSNAP: http://share.gene.com/gmap
iCLIP questions and answers: http://goo.gl/4tSci
iCount pipeline: http://icount.biolab.si
Novoalign: http://www.novocraft.com/main/index.php
Segemehl: http://www.bioinf.uni-leipzig.de/Software/segemehl
starBase: http://starbase.sysu.edu.cn
Uwe Ohler's Research Group PARalyzer (PAR-CLIP data analyzer): http://www.genome.duke.edu/labs/ohler/research/PARalyzer

ALL LINKS ARE ACTIVE IN THE ONLINE PDF