Genome **Biology**

'Oming in on RNA–protein interactions

John L Rinn[1,2,3*] and Jernej Ule[4*]

Welcome to the RBPome

RNA has consistently broken dogmas, owing to its multitude of unexpected functions. However, RNA does faithfully adhere to one rule: it always functions through interactions with proteins. The studies in this issue focus on the rapidly expanding repertoire of diverse RNA–protein interactions [1] and their functional roles and physiological consequences, both from the perspective of RNA-binding proteins (RBPs) and from the vantage point of RNAs - coding and noncoding - that interface with RBPs.

RNA-protein interactions are fascinating for many reasons, one being their role in evolution - from the earliest life forms to the most complex organisms (examples reviewed in [2-4]). For example, the interactions between pre-mRNA and proteins fine-tune alternative splicing in a manner that can gradually create new protein functionalities without the need to create additional genes and without affecting existing proteins [4-6]. Moreover, there has been an emergence of numerous noncanonical RBPs (that is, proteins not previously thought to function as RBPs) that are influenced by interactions with RNA transcripts coding and noncoding alike [7,8]. In fact, genome-wide footprinting articles in this issue [9-11] demonstrate a vast and diverse landscape of RNA–RBP complexes that play key regulatory roles.

With recent advances in technology, together with powerful combined experimental and computational developments, we have witnessed unprecedented new insights into the diverse and dynamic interactions that occur between RNA and proteins. These range from new functions of well-established RBPs to the molecular sequences and structures harbored in RNA that drive interactions with proteins. Despite this great progress, the present issue of *Genome Biology* demonstrates that there are still unresolved aspects of RBP biology.

Technology opening up new horizons of the RBPome

A major challenge in understanding RNA–protein interactions has been to remove non-specific abundant RNAs when mapping protein-binding sites in low-abundant RNAs. For instance, in order to understand how tissue specificity of splicing is determined, it was important to identify protein-binding sites on pre-mRNAs without contamination from mRNAs or ribosomal RNAs. This was achieved with the UV crosslinking and immunoprecipitation (CLIP) method, which employs stringent purification steps to identify specific binding sites to the exclusion of non-specific events [12].

As advances were made in CLIP technology [13,14], integrative computational approaches were developed to combine CLIP with genome-wide studies of alternative splicing to define regulatory maps for specific RBPs. These maps unraveled the position-dependent principles that were capable of predicting the functional binding sites of RBPs [15-17]. Moreover, the regulatory elements positioned around alternative exons were used to derive a splicing code with a capacity to predict tissue-specific splicing with a reasonable accuracy [18,19].

Technology developed to better understand RNA–protein interactions is not limited to CLIP, but is also maturing on several additional fronts [11,14]. Indeed, presented in this issue are a number of novel genome-wide experimental and analysis techniques that yield new insights into RBPology [9,10,20-25]. RNA pull-downs (also known as RIP-seq) and RNA-footprinting methods are revealing key RBP interactions that do not fit the mold of classic RBPs. Technologies will therefore need to advance further if we are to better understand the interplay of noncoding transcripts and RBPs. Why are long noncoding RNA (lncRNA) transcripts alternatively spliced by RBPs like their translated mRNA counterparts? Can lncRNAs serve as decoys, scaffolds or allosteric effectors of RBPs [26]?

Future studies will need to focus on additional technologies in order to identify the specific sequences and structures of RBP–RNA complexes on a larger scale. For example, new methods that determine the genome-wide

* Correspondence: john_rinn@harvard.edu; j.ule@ucl.ac.uk
[1]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA
[4]Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK
Full list of author information is available at the end of the article

structure of RNA molecules *in vivo* will be instrumental in achieving this goal [27,28]. The many other key aspects of the RBPome that still remain to be solved include understanding the combinatorial interactions of proteins on a given RNA substrate. Can multiple interactions provide combinatorial control? Methods to understand the structure and interactions of RBPs on full-length RNAs will be needed in the future.

New RNA species detected by high-throughput sequencing

The ability to detect low-abundant RNAs by using high-throughput sequencing has led to the discovery of thousands of noncoding RNAs (discussed in [29]). One facet of the noncoding transcriptome that has been of intensive research focus is lncRNAs. As perhaps expected, lncRNAs have been found to break the rules and form numerous noncanonical interactions with proteins such as chromatin regulatory complexes, cohesins and transcription factors (see, for example, [30-32]). Moreover, lncRNAs have been shown to influence the regulatory dynamics of small RNAs through sponging [33] and other mechanisms. However, it is not known whether lncRNA–protein interactions are generally required to mediate the functions of lncRNAs or, *vice versa*, whether lncRNAs modulate RBPs.

The advent of new techniques that can identify RNA–DNA and RNA–protein interactions are revealing a new regulatory layer of lncRNAs. These techniques, which include CHART [34], RAP [35] and ChOP [36], are conceptually similar to ChIP. Instead of mapping DNA–protein interactions, however, these methods determine the localization of RNA on DNA and, moreover, enable the study of proteins that interact with RNA at these sites. Experiments of this nature will provide missing clues into the regulatory principles of how lncRNAs arrive at their target sites, the proteins they interact with to get there and which sequences are specifically required. Collectively, lncRNA–protein interactions perhaps herald a new code for RNA localization around the genome and how it influences local and distal epigenetic aspects of nuclear architecture through these interactions. This is an area certain to be of intense development and research focus in the future.

New RNA modifications detected by high-throughput sequencing

Modified ribonucleic bases have been an area of active interest for over half a century. Yet recent genomic sequencing technologies have expanded the modified RNA space beyond classic tRNA and rRNA modifications to incorporate mRNA and lncRNA transcripts as well (reviewed in [37,38]). The field of epitranscriptomics is in its infancy, and many questions remain to be answered. Why are these modifications so prevalent?

How do they influence RNA-binding interactions and RNA structure? What is their function? Moreover, these same questions also apply to the regulatory layers formed by RNA editing [39]. Specifically, Levanon and colleagues demonstrate that only a very small fraction of sites edited by ADAR are conserved.

Understanding disease-causing mutations

Impact upon human disease is perhaps the most important new frontier in the study of the RBPome. We need to use classic genetic approaches and population studies to identify mutations in both RBPs and RNA substrates themselves. Great progress has been made in finding mutations in RBPs associated with disease risk, such as the RBPs FUS and TDP-43 in amyotrophic lateral sclerosis (reviewed in [40]). Yet more work will be needed to understand what changes occur to RNA substrates in disease and their effect on structure and function, although progress in this area is already underway (reviewed in this issue in [41]).

Several studies in this issue have taken on the challenge of examining the intersection between the RBPome and human disease. Mort, Mooney and colleagues identify single-base mutations that affect the alternative splicing of key regulatory mRNAs [25]. Kechavarzi and Janga explore the dysregulation of RBP-encoding transcripts in cancer and the resulting changes in protein interaction networks [42]. Finally, Tuschl, Wessels and colleagues use PAR-CLIP data to identify differential microRNA targeting that is correlated with breast cancer subtypes [43]. These types of studies are archetypical for numerous future studies to understand the influence of the RBPome on human health and disease.

The next challenge will be to integrate DNA and RNA biology to understand how the various transcriptional and post-transcriptional mechanisms cooperate to orchestrate gene expression. This knowledge will be crucial in order to benchmark the mutations that cause disease by changing gene expression.

Cometh the hour, cometh the special issue

The field's rapid diversification and growth, together with an increasing impact on human health, makes for perfect timing for a *Genome Biology* issue focused on the RBPome!

Author details
[1]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA. [2]Broad Institute, Cambridge, MA 02142, USA. [3]Beth Israel Deaconess Medical Center, Boston, MA 02215, USA. [4]Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK.

References

1. Attar N: The RBPome: where the brains meet the brawn. *Genome Biol* 2014, **15**:402.
2. Borsenberger V, Crowe MA, Lehbauer J, Raftery J, Helliwell M, Bhutia K, Cox T, Sutherland JD: Exploratory studies to investigate a linked prebiotic origin of RNA and coded peptides. *Chem Biodivers* 2004, **1**:203–246.
3. Bernhardt HS: The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others). *Biol Direct* 2012, **7**:23.
4. Keren H, Lev-Maor G, Ast G: Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet* 2010, **11**:345–355.
5. Zarnack K, König J, Tajnik M, Martincorena I, Eustermann S, Stévant I, Reyes A, Anders S, Luscombe NM, Ule J: Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* 2013, **152**:453–466.
6. Gal-Mark N, Schwartz S, Ram O, Eyras E, Ast G: The pivotal roles of TIA proteins in 5′ splice-site selection of Alu exons and across evolution. *PLOS Genet* 2009, **5**:e1000717.
7. Kaneko S, Son J, Shen SS, Reinberg D, Bonasio R: PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat Struct Mol Biol* 2013, **20**:1258–1264.
8. Davidovich C, Zheng L, Goodrich KJ, Cech TR: Promiscuous RNA binding by Polycomb repressive complex 2. *Nat Struct Mol Biol* 2013, **20**:1250–1257.
9. Silverman IM, Li F, Alexander A, Goff L, Trapnell C, Rinn JL, Gregory BD: RNase-mediated protein footprint sequencing reveals protein-binding sites throughout the human transcriptome. *Genome Biol* 2014, **15**:R3.
10. Schueler M, Munschauer M, Gregersen LH, Finzel A, Loewer A, Chen W, Landthaler M, Dieterich C: Differential protein occupancy profiling of the mRNA transcriptome. *Genome Biol* 2014, **15**:R15.
11. McHugh CA, Russell P, Guttman M: Methods for comprehensive experimental identification of RNA–protein interactions. *Genome Biol* 2014, **15**:203.
12. Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB: CLIP identifies Nova-regulated RNA networks in the brain. *Science* 2003, **302**:1212–1215.
13. Sugimoto Y, König J, Hussain S, Zupan B, Curk T, Frye M, Ule J: Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol* 2012, **13**:R67.
14. König J, Zarnack K, Luscombe NM, Ule J: Protein–RNA interactions: new genomic technologies and perspectives. *Nat Rev Genet* 2011, **13**:77–83.
15. Ule J, Stefani G, Mele A, Ruggiu M, Wang X, Taneri B, Gaasterland T, Blencowe BJ, Darnell RB: An RNA map predicting Nova-dependent splicing regulation. *Nature* 2006, **444**:580–586.
16. Zhang C, Lee KY, Swanson MS, Darnell RB: Prediction of clustered RNA-binding protein motif sites in the mammalian genome. *Nucleic Acids Res* 2013, **41**:6793–6807.
17. Zhang C, Frias MA, Mele A, Ruggiu M, Eom T, Marney CB, Wang H, Licatalosi DD, Fak JJ, Darnell RB: Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. *Science* 2010, **329**:439–443.
18. Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencowe BJ, Frey BJ: Deciphering the splicing code. *Nature* 2010, **465**:53–59.
19. Barash Y, Vaquero-Garcia J, González-Vallinas J, Xiong HY, Gao W, Lee LJ, Frey BJ: AVISPA: a web tool for the prediction and analysis of alternative splicing. *Genome Biol* 2013, **14**:R114.
20. Fukunaka T, Ozaki H, Terai G, Asai K, Iwasaki W, Kiryu H: CapR: revealing structural specificities of RNA-binding protein target recognition using CLIP-seq data. *Genome Biol* 2014, **15**:R16.
21. Chen B, Yun J, Kim MS, Mendell JT, Xie Y: PIPE-CLIP: a comprehensive online tool for CLIP-seq data analysis. *Genome Biol* 2014, **15**:R18.
22. Wang T, Xie Y, Xiao G: dCLIP: a computational approach for comparative CLIP-seq analyses. *Genome Biol* 2014, **15**:R11.
23. Maticzka D, Lange SJ, Costa F, Backofen R: GraphProt: modeling binding preferences of RNA-binding proteins. *Genome Biol* 2014, **15**:R17.
24. Cereda M, Pozzoli U, Rot G, Juvan P, Schweitzer A, Clark T, Ule J: RNAmotifs: prediction of multivalent RNA motifs that control alternative splicing. *Genome Biol* 2014, **15**:R20.
25. Mort M, Sterne-Weiler T, Li B, Ball EB, Cooper D, Radivojac P, Sanford J, Mooney SD: MutPred Splice: machine learning-based prediction of exonic variants that disrupt splicing. *Genome Biol* 2014, **15**:R19.
26. Ulitsky I, Bartel DP: lincRNAs: genomics, evolution, and mechanisms. *Cell* 2013, **154**:26–46.
27. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM: In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 2013, doi:10.1038/nature12756.
28. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS: Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 2013, doi:10.1038/nature12894.
29. Jarvis K, Robertson M: The noncoding universe. *BMC Biol* 2011, **9**:52.
30. Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C: Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* 2008, **32**:232–246.
31. Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP: Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci Signal* 2010, **3**:ra8.
32. Sun S, Del Rosario BC, Szanto A, Ogawa Y, Jeon Y, Lee JT: Jpx RNA activates Xist by evicting CTCF. *Cell* 2013, **153**:1537–1551.
33. Marques AC, Tan J, Ponting CP: Wrangling for microRNAs provokes much crosstalk. *Genome Biol* 2011, **12**:132.
34. Simon MD, Wang CI, Kharchenko PV, West JA, Chapman BA, Alekseyenko AA, Borowsky ML, Kuroda MI, Kingston RE: The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci U S A* 2011, **108**:20497–20502.
35. Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES, Plath K, Guttman M: The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 2013, **341**:1237973.
36. Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, Kugel JF, Goodrich JA: Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell* 2008, **29**:499–509.
37. Hussain S, Aleksic J, Blanco S, Dietmann S, Frye M: Characterizing 5-methylcytosine in the mammalian epitranscriptome. *Genome Biol* 2013, **14**:215.
38. Saletore Y, Meyer K, Korlach J, Vilfan ID, Jaffrey S, Mason CE: The birth of the Epitranscriptome: deciphering the function of RNA modifications. *Genome Biol* 2012, **13**:175.
39. Pinto Y, Cohen HY, Levanon EY: Mammalian conserved ADAR targets comprise only a small fragment of the human editosome. *Genome Biol* 2014, **15**:R5.
40. Lagier-Tourenne C, Cleveland DW: Rethinking ALS: the FUS about TDP-43. *Cell* 2009, **136**:1001–1004.
41. Sterne-Weiler T, Sanford JR: Exon identity crisis: disease-causing mutations that disrupt the splicing code. *Genome Biol* 2014, **15**:201.
42. Kechavarzi B, Janga SC: Dissecting the expression landscape of RNA-binding proteins in human cancers. *Genome Biol* 2014, **15**:R14.
43. Farazi TA, Hoeve JJ T, Brown M, Mihailovic A, Horlings HM, Vijver MJ Van D, Tuschl T, Wessels LF: Identification of distinct miRNA target regulation between breast cancer molecular subtypes using AGO2-PAR-CLIP and patient datasets. *Genome Biol* 2014, **15**:R9.