Title

rMAPS: RNA map analysis and plotting server for alternative exon regulation

Journal

Nucleic Acids Research, 44(W1)

ISSN

0305-1048

Authors

Park, Juw Won
Jung, Sungbo
Rouchka, Eric C
et al.

Publication Date

2016-07-08

DOI

10.1093/nar/gkw410

Published online 12 May 2016 *Nucleic Acids Research, 2016, Vol. 44, Web Server issue* ***W333–W338***
doi: 10.1093/nar/gkw410

rMAPS: RNA map analysis and plotting server for alternative exon regulation

Juw Won Park[1,2,3,*,†], Sungbo Jung[1,†], Eric C. Rouchka[1,2], Yu-Ting Tseng[3] and Yi Xing[3,*]

[1]Department of Computer Engineering and Computer Science, University of Louisville, Louisville, KY 40292, USA, [2]KBRIN Bioinformatics Core, University of Louisville, Louisville, KY 40202, USA and [3]Department of Microbiology, Immunology & Molecular Genetics, University of California, Los Angeles, CA 90095, USA

Received March 7, 2016; Revised May 3, 2016; Accepted May 3, 2016

ABSTRACT

RNA-binding proteins (RBPs) play a critical role in the regulation of alternative splicing (AS), a prevalent mechanism for generating transcriptomic and proteomic diversity in eukaryotic cells. Studies have shown that AS can be regulated by RBPs in a binding-site-position dependent manner. Depending on where RBPs bind, splicing of an alternative exon can be enhanced or suppressed. Therefore, spatial analyses of RBP motifs and binding sites around alternative exons will help elucidate splicing regulation by RBPs. The development of high-throughput sequencing technologies has allowed transcriptome-wide analyses of AS and RBP–RNA interactions. Given a set of differentially regulated alternative exons obtained from RNA sequencing (RNA-seq) experiments, the rMAPS web server (http://rmaps.cecsresearch.org) performs motif analyses of RBPs in the vicinity of alternatively spliced exons and creates RNA maps that depict the spatial patterns of RBP motifs. Similarly, rMAPS can also perform spatial analyses of RBP–RNA binding sites identified by cross-linking immunoprecipitation sequencing (CLIP-seq) experiments. We anticipate rMAPS will be a useful tool for elucidating RBP regulation of alternative exon splicing using high-throughput sequencing data.

INTRODUCTION

RNA-binding proteins (RBPs) are proteins that bind to mRNAs or non-coding RNAs and play diverse roles in post-transcriptional processing and regulation of RNA (1). One of the important post-transcriptional regulatory roles of RBPs is the regulation of alternative splicing (AS). AS is a prevalent mechanism for generating transcriptomic and proteomic complexity in eukaryotic cells (2). Many studies have shown that RBP regulation of AS is dependent on the binding site positions of RBP with respect to the regulated alternative exons (3–7). For example, the Epithelial Splicing Regulatory Proteins 1 and 2 (ESRP1/2) are master RBP regulators of AS in epithelial cells (4,8). In general, splicing is enhanced when ESRP binds to pre-mRNA downstream of the exon, while splicing is suppressed when ESRP binds to pre-mRNA upstream of or within the exon (Figure 1A) (3,4). Therefore, spatial analyses of RBP motif or binding sites around regulated alternative exons may reveal whether and how AS events in specific biological processes are controlled by specific RBPs. Figure 1B shows a schematic RNA map depicting the spatial distribution of the ESRP motif around ESRP-regulated alternative exons, which indicates a higher motif/binding site density downstream of ESRP-enhanced exons (red peaks) and upstream of or within ESRP-silenced exons (blue peaks).

The significant advances in high-throughput sequencing technologies have enabled transcriptome-wide analyses of AS and RBP–RNA interactions. RNA-seq has become a widely used approach to identify differentially regulated alternative exons in specific biological processes (9). In parallel, cross-linking immunoprecipitation (CLIP)-seq can identify transcriptome-wide binding sites of a given RBP within endogenous RNA transcripts (10). To date, specialized bioinformatics tools and pipelines have been developed for AS analysis of RNA-seq data (9,11), or for RBP binding site analysis of CLIP-seq data (12–14), but these two types of analyses are carried out separately. To our knowledge, there is no user-friendly software or web-based tool that integrates RNA-seq analysis of AS, with information of RBP consensus motifs or CLIP-seq RBP binding sites. Such a tool will be useful for identifying candidate RBPs that control regulated AS events in specific biological processes, and reveal their potential position-dependent effects on exon splicing.

In this paper, we introduce a new web server rMAPS (RNA Map Analysis and Plotting Server, http://rmaps.cecsresearch.org), to perform RBP motif or binding site

*To whom correspondence should be addressed. Tel: +1 310 825 6806; Fax: +1 310 206 3663; Email: yxing@ucla.edu
Correspondence may also be addressed to Juw Won Park. Tel: +1 502 852 6307; Fax: +1 502 852 4713; Email: juw.park@louisville.edu
† These authors contributed equally to the paper as first authors.

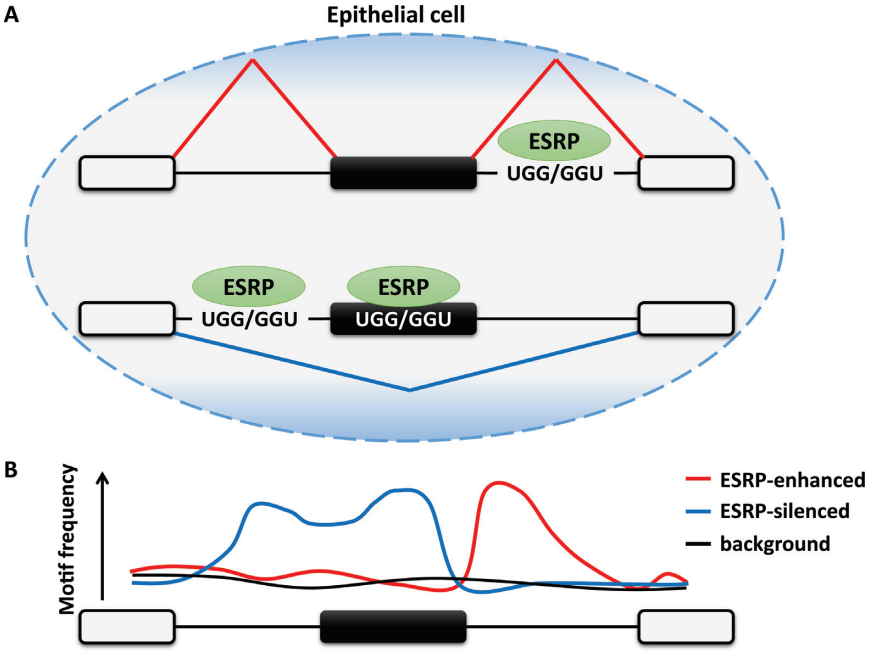

Figure 1. Binding-site-position dependent regulation of epithelial specific AS by ESRP proteins. The solid black exon in the middle is the ESRP regulated exon. (**A**) ESRP generally enhances splicing when it binds to the downstream intron while it suppresses splicing when it binds to the upstream intron or within the regulated exon. (**B**) A schematic RNA map for the spatial distribution of the ESRP consensus motif around different categories of alternative exons. The red line represents the enriched ESRP motif in the downstream intron of ESRP-enhanced exons, while the blue line represents the enriched ESRP motif in the upstream intron and within the exon of ESRP-silenced exons. The black line represents the motif density for background (non-regulated) exons.

analyses for differentially regulated alternative exons obtained from RNA-seq experiments. Specifically, from a list of alternatively spliced exons, rMAPS performs motif enrichment analyses of RBPs within the exonic and flanking intronic regions and plots the spatial distributions of RBP motif density (RNA maps) for a set of known RBP motifs collected from the literature (15–17) (see Supplementary file 1 for the list of RBPs currently supported). Furthermore, rMAPS can examine RBP binding sites within endogenous RNAs identified by CLIP-seq and create an RNA map to visualize the spatial distribution of binding sites for the RBP of interest. Collectively, the rMAPS web server provides an efficient and user-friendly interface for generating RNA maps using known RBP motifs as well as CLIP-seq data of any specific RBP. rMAPS is free and open to all users and there is no login requirement.

rMAPS WEB SERVER

rMAPS provides two major types of analyses: RBP motif analysis and CLIP-seq RBP binding site analysis. In the context of this work, we define RBP motif as the consensus binding sequence of a given RBP, represented in the form of a regular expression. Consensus motif patterns of numerous RBPs have been defined by *in vitro* protein–RNA binding assays, such as RNAcompete (17) and RNA Bind-n-Seq (18). We define RBP binding sites as the locations where the RBP binds to endogenous RNA transcripts, which can be identified across the transcriptome using CLIP-seq. After introducing the overall workflow of the rMAPS web server below, we will explain both analyses in detail and describe the required inputs, the resulting outputs, and details of the actual implementation.

Overall workflow

The overall workflow of rMAPS is shown in Figure 2. During the pre-processing step (outside of the dashed red box), RNA-seq data is processed offline using existing AS analysis tools (such as rMATS (11) and MISO (9)) to compile the sets of differentially regulated and non-regulated control exons. The coordinates of these exons along with known RBP motifs are used in the motif analysis that generates a series of RNA maps depicting the spatial distributions of RBP motifs around the regulated exons. Users can examine the resulting maps to identify candidate RBPs as well as their position-dependent effects on exon inclusion or skipping.

Similarly, for the CLIP-seq binding site analysis, the coordinates of regulated alternative exons are coupled with CLIP-seq binding sites identified by CLIP-seq peak callers such as Piranha (12), PIPE-CLIP (13), and PARalyzer (14) to examine the spatial distributions of RBP binding sites within the alternative exons and adjacent intronic regions.

RBP motif analysis

In the RBP motif analysis, rMAPS seeks to identify the known consensus motifs of RBPs that are significantly enriched in differentially regulated AS events between two sample groups as compared to control (non-regulated background) AS events. Additionally, rMAPS generates plots

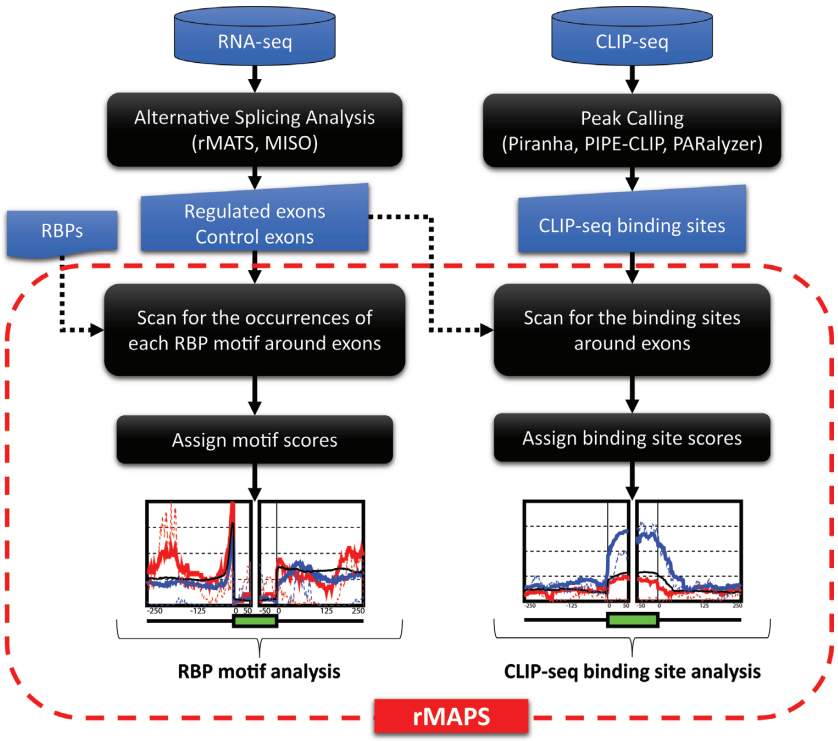

Figure 2. Overall workflow of rMAPS. The RBP motif analysis takes the AS analysis results from RNA-seq data, then scans for the occurrences of the known RBP motifs around the differentially regulated alternative exons to generate an RNA map for each RBP. Similarly the CLIP-seq binding site analysis generates an RNA map using the signals of CLIP-seq binding sites around the differentially regulated alternative exons.

to visualize the spatial distributions of RBP motif density within the exons and flanking intronic regions. Motif analysis requires the following inputs:

(i) *AS events from RNA-seq data*: these events can be automatically processed and retrieved from rMATS (11) or MISO (9) outputs for exon skipping events or provided by users following the required input format. When rMATS or MISO output files are uploaded to the web server, exon skipping events with at least 5% increase in exon inclusion level at rMATS FDR 5% (or MISO Bayes factor > 10) between treatment (case) versus control sample groups are compiled as upregulated exons. Exon skipping events with at least 5% decrease in exon inclusion level at rMATS FDR 5% (or MISO Bayes factor > 10) are treated as downregulated exons. Alternative exons without splicing changes (rMATS FDR > 50% or MISO Bayes factor < 1, maximum exon inclusion level > 15%, minimum exon inclusion level < 85%) are compiled and used as control (non-regulated background) exons. When the user provides the events, each event is required to have eight tab-delimited columns. The eight columns are chromosome, strand and six columns for the start and end coordinates of the target exon as well as its upstream and downstream flanking exons.

(ii) *RBPs with known consensus motifs*: rMAPS has collected known consensus motifs of 115 RBPs including numerous well-characterized splicing factors from the literature (15–17). These motifs are included in the analysis by default. Users can also input additional consensus motifs as shown in Figure 3A.

(iii) *Optional parameters to specify flanking intron size, exon size, sliding window size, step-size (or interval) and additional motifs*: For each motif, rMAPS scans for motif occurrences separately in exons or their upstream and downstream introns. By default, rMAPS examines 250bp upstream or downstream intronic sequences and the first 50bp of the 5′ or 3′ end of exonic sequences. To calculate motif density, rMAPS examines a 50bp sliding window to count the number of nucleotides covered by a given motif. rMAPS slides this window by 1bp at a time. Users can customize these values as shown in Figure 3A to set the size of the intronic or exonic sequence included in analysis, the size of the sliding window, the step-size (or interval), as well as additional user-defined motifs provided in the form of regular expression.

Once all required inputs are provided, rMAPS scans for the occurrences of each RBP motif in the upstream exon, upstream flanking intron, target exon, downstream flanking intron and downstream exon separately. For intronic sequences, rMAPS excludes the 20bp sequence within the 3′ splice site and the 6bp sequence within the 5′ splice site which are strongly constrained by the consensus splice site signals. For each motif, rMAPS calculates its motif score (density) within a given 50bp window as the overall percentage of nucleotides covered by the motif. After computing motif scores separately for the upregulated, down-

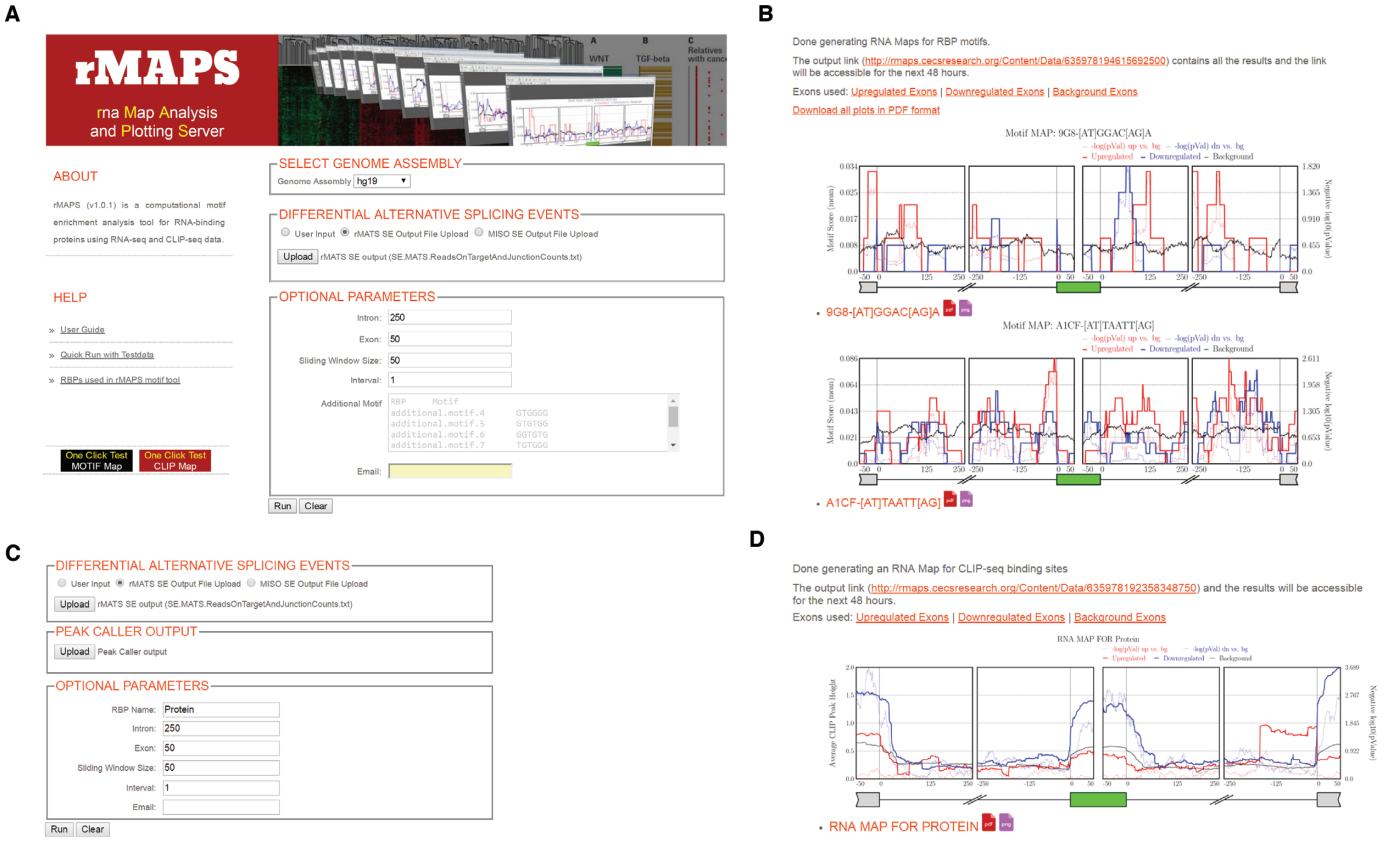

Figure 3. A view of rMAPS input and output pages. (**A**) Input page for the motif analysis tool. After providing all required inputs, clicking on 'Run' button will start the analysis. (**B**) An example output page of the motif analysis tool. It reports RNA maps for known RBP motifs along with a URL link for the entire results for future use. (**C**) Input page for the CLIP-seq binding site analysis tool. It requires AS events from RNA-seq data and peaks from CLIP-seq data. (**D**) An example output page of the CLIP-seq binding site analysis tool. It reports an RNA map for the RBP of interest and a URL link for the entire results for future use.

regulated and control (non-regulated background) exons, rMAPS plots motif scores for these three sets of exons in red, blue and black respectively. rMAPS then generates an RNA map for each motif to visualize the spatial distributions of motif score (density) for each RBP (Figure 3B). In addition, to identify specific positions with significant difference in motif score between different sets of exons, rMAPS performs Wilcoxon's rank sum test for each sliding window between upregulated versus control or downregulated versus control exons, and plots *P*-values (Figure 3B). rMAPS also provides a URL link for the entire results for future use (Figure 3B).

CLIP-seq binding site analysis

In the CLIP-seq binding site analysis, rMAPS combines differentially regulated AS events from RNA-seq with RBP binding sites from CLIP-seq. rMAPS then generates an RNA map of binding sites near the target alternative exons. The following inputs are required for the CLIP-seq binding site analysis (Figure 3C):

(i) *AS events from RNA-seq data*: the same input used for the RBP motif analysis (described above) can be used for the CLIP-seq binding site analysis.

(ii) *RBP binding sites from CLIP-seq data*: RBP binding sites called from CLIP-seq peak callers such as Piranha (12), PIPE-CLIP (13), and PARalyzer (14). Each peak is required to have at least seven tab-delimited columns which are chromosome, peak start coordinate, peak end coordinate, peak ID, peak height (or read count), strand, and *P*-value.

(iii) *RBP name:* name of the RBP used in the CLIP-seq experiment. This name is displayed in the RNA map.

(iv) *Optional parameters to specify flanking intron size, exon size, sliding window size, and step-size (or interval):* the same optional parameters used for the RBP motif analysis (described above) can be used for the CLIP-seq binding site analysis.

Similar to the RBP motif analysis described previously, after compiling differential AS events and control AS events from RNA-seq data, rMAPS evaluates the signals of CLIP-seq binding sites in the upstream exon, upstream flanking intron, target exon, downstream flanking intron, and downstream exon separately. rMAPS then generates an RNA map by plotting the spatial distributions of average CLIP-seq signals around upregulated, downregulated, and control (non-regulated) exons in different colors (Figure 3D). rMAPS also provides a URL link for the entire results for future use (Figure 3D).

rMAPS web server implementation

The server is implemented in HTML, PHP and Python. The front page runs on Windows Server 2012 R2 with Intel Xeon E5345 2.33GHz 4-core CPUs. To ensure a reasonable run time for each analysis, the actual analysis is performed on a cluster that runs on CentOS7 with Intel Xeon X5670 2.93GHz 6-core CPUs equipped with 144GB DDR3-10600R ECC Registered Memory. The typical run time of each analysis is less than 10 min.

CONCLUSIONS

We have developed rMAPS, a web server to integrate RNA-seq results of differential alternative exon regulation, with information of RBP motif occurrence or CLIP-seq binding sites across the transcriptome. RNA-seq has become a popular approach for global analyses of AS. It is typical for an RNA-seq study to identify hundreds or even thousands of differential AS events between different biological states. However, RNA-seq data alone does not provide information on how these differential AS events are regulated. RBPs are master regulators of post-transcriptional RNA processing including AS (1). Numerous studies have demonstrated that global changes in AS in specific biological processes are usually mediated by differential expression or activity of a small number of RBPs that interact with the pre-mRNAs (19) through their consensus motifs. Moreover, a number of RBPs have been found to display binding-site-position dependent activity on exon splicing (3–7). Based on these observations and assumptions, it is customary for RNA-seq studies of AS regulation to search for enriched RBP motifs, or explore the spatial distributions of consensus RBP motifs or binding sites around the differentially spliced exons. rMAPS provides an easy-to-use web interface to perform these important analyses, with a primary goal on identifying RBPs with significant motif/binding site enrichment and/or potential position-dependent regulatory roles. We expect that rMAPS will be a useful tool for elucidating RBP regulation of AS using high-throughput sequencing data.

The implementation of the rMAPS web server still has limitations and room for improvement. Currently, rMAPS only works with the exon skipping type of AS events. This is the most prevalent form of AS in animals and RNA map analyses have been traditionally performed on exon skipping events (3–7). In the future, we will expand the plotting functions of rMAPS to generate RNA maps for other types of AS, such as alternative 5′/3′ splice sites, mutually exclusive exon usage, and intron retention. Also, the current rMAPS server requires users to provide CLIP-seq data/peak calling results as the input for CLIP-seq binding site analyses of AS. This is not a major issue because many RBP studies carry out RNA-seq and CLIP-seq experiments in parallel. Nonetheless, we should note that CLIP-seq datasets on a large number of RBPs across a variety of cell types and tissues have been generated and released into the public domain (20). In the future, we plan to incorporate public CLIP-seq datasets into the rMAPS server, which will enhance its ability to explore RBP binding events in any RNA-seq dataset.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Ron Lile for extensive testing and helpful comments. We would like to thank Julia Chariker and Ben Harrison for insightful discussions.

FUNDING

National Institutes of Health (NIH) [P20GM103436 to J.W.P., E.C.R.; R01GM088342 to Y.X.]; Alfred Sloan Research Fellowship [BR2013-117 to Y.X.]. Funding for open access charge: NIH [P20GM103436].
Conflict of interest statement. None declared.

REFERENCES

1. Li,X., Kazan,H., Lipshitz,H.D. and Morris,Q.D. (2014) Finding the target sites of RNA-binding proteins. *Wiley Interdiscip. Rev. RNA*, **5**, 111–130.
2. Chen,M. and Manley,J.L. (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.*, **10**, 741–754.
3. Bebee,T.W., Park,J.W., Sheridan,K.I., Warzecha,C.C., Cieply,B.W., Rohacek,A.M., Xing,Y. and Carstens,R.P. (2015) The splicing regulators Esrp1 and Esrp2 direct an epithelial splicing program essential for mammalian development. *Elife*, **4**, e08954.
4. Warzecha,C.C., Jiang,P., Amirikian,K., Dittmar,K.A., Lu,H., Shen,S., Guo,W., Xing,Y. and Carstens,R.P. (2010) An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. *EMBO J.*, **29**, 3286–3300.
5. Martinez-Contreras,R., Fisette,J.F., Nasim,F.U., Madden,R., Cordeau,M. and Chabot,B. (2006) Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. *PLoS Biol.*, **4**, e21.
6. Ule,J., Stefani,G., Mele,A., Ruggiu,M., Wang,X., Taneri,B., Gaasterland,T., Blencowe,B.J. and Darnell,R.B. (2006) An RNA map predicting Nova-dependent splicing regulation. *Nature*, **444**, 580–586.
7. Yeo,G.W., Coufal,N.G., Liang,T.Y., Peng,G.E., Fu,X.D. and Gage,F.H. (2009) An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat. Struct. Mol. Biol.*, **16**, 130–137.
8. Warzecha,C.C., Sato,T.K., Nabet,B., Hogenesch,J.B. and Carstens,R.P. (2009) ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. *Mol. Cell*, **33**, 591–601.
9. Katz,Y., Wang,E.T., Airoldi,E.M. and Burge,C.B. (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods*, **7**, 1009–1015.
10. 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.
11. Shen,S., Park,J.W., Lu,Z.X., Lin,L., Henry,M.D., Wu,Y.N., Zhou,Q. and Xing,Y. (2014) rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E5593–E5601.
12. Uren,P.J., Bahrami-Samani,E., Burns,S.C., Qiao,M., Karginov,F.V., Hodges,E., Hannon,G.J., Sanford,J.R., Penalva,L.O. and Smith,A.D. (2012) Site identification in high-throughput RNA-protein interaction data. *Bioinformatics*, **28**, 3013–3020.
13. Chen,B., Yun,J., Kim,M.S., Mendell,J.T. and Xie,Y. (2014) PIPE-CLIP: a comprehensive online tool for CLIP-seq data analysis. *Genome Biol.*, **15**, R18.
14. Corcoran,D.L., Georgiev,S., Mukherjee,N., Gottwein,E., Skalsky,R.L., Keene,J.D. and Ohler,U. (2011) PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol.*, **12**, R79.

15. Anderson,E.S., Lin,C.H., Xiao,X., Stoilov,P., Burge,C.B. and Black,D.L. (2012) The cardiotonic steroid digitoxin regulates alternative splicing through depletion of the splicing factors SRSF3 and TRA2B. *RNA*, **18**, 1041–1049.
16. Dittmar,K.A., Jiang,P., Park,J.W., Amirikian,K., Wan,J., Shen,S., Xing,Y. and Carstens,R.P. (2012) Genome-wide determination of a broad ESRP-regulated posttranscriptional network by high-throughput sequencing. *Mol. Cell. Biol.*, **32**, 1468–1482.
17. Ray,D., Kazan,H., Cook,K.B., Weirauch,M.T., Najafabadi,H.S., Li,X., Gueroussov,S., Albu,M., Zheng,H., Yang,A. *et al.* (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, **499**, 172–177.
18. Lambert,N., Robertson,A., Jangi,M., McGeary,S., Sharp,P.A. and Burge,C.B. (2014) RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol. Cell*, **54**, 887–900.
19. Fu,X.D. and Ares,M. Jr (2014) Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.*, **15**, 689–701.
20. Yang,Y.C., Di,C., Hu,B., Zhou,M., Liu,Y., Song,N., Li,Y., Umetsu,J. and Lu,Z.J. (2015) CLIPdb: a CLIP-seq database for protein-RNA interactions. *BMC Genomics*, **16**, 51.