Published online 17 August 2021 *Nucleic Acids Research, 2022, Vol. 50, Database issue* ***D287–D294***
https://doi.org/10.1093/nar/gkab702

POSTAR3: an updated platform for exploring post-transcriptional regulation coordinated by RNA-binding proteins

Weihao Zhao[1,†], Shang Zhang[1,†], Yumin Zhu[1,2], Xiaochen Xi[1], Pengfei Bao[1], Ziyuan Ma[1,3], Thomas H. Kapral[4], Shuyuan Chen[1,5], Bojan Zagrovic[4], Yucheng T. Yang[6,7,8,*] and Zhi John Lu[1,*]

[1]MOE Key Laboratory of Bioinformatics, Center for Synthetic and Systems Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China, [2]Department of Maternal, Child and Adolescent Health, School of Public Health, Anhui Medical University, MOE Key Laboratory of Population Health Across Life Cycle, NHC Key Laboratory of Study on Abnormal Gametes and Reproductive Tract, Anhui Provincial Key Laboratory of Population Health and Aristogenics, No 81 Meishan Road, Hefei 230032, Anhui, China, [3]School of Pharmaceutical Sciences, Tsinghua University, Beijing 100084, China, [4]Department of Structural and Computational Biology, Max Perutz Labs, University of Vienna, Campus Vienna Biocenter 5, A-1030 Vienna, Austria, [5]Faculty of Science, University of Melbourne, Melbourne, Victoria 3010, Australia, [6]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai 200433, China, [7]MOE Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, Fudan University, Shanghai 200433, China and [8]MOE Frontiers Center for Brain Science, Fudan University, Shanghai 200433, China

Received June 07, 2021; Revised July 08, 2021; Editorial Decision July 28, 2021; Accepted August 14, 2021

ABSTRACT

RNA-binding proteins (RBPs) play key roles in post-transcriptional regulation. Accurate identification of RBP binding sites in multiple cell lines and tissue types from diverse species is a fundamental endeavor towards understanding the regulatory mechanisms of RBPs under both physiological and pathological conditions. Our POSTAR annotation processes make use of publicly available large-scale CLIP-seq datasets and external functional genomic annotations to generate a comprehensive map of RBP binding sites and their association with other regulatory events as well as functional variants. Here, we present POSTAR3, an updated database with improvements in data collection, annotation infrastructure, and analysis that support the annotation of post-transcriptional regulation in multiple species including: we made a comprehensive update on the CLIP-seq and Ribo-seq datasets which cover more biological conditions, technologies, and species; we added RNA secondary structure profiling for RBP binding sites; we provided miRNA-mediated degradation events validated by degradome-seq; we included RBP binding sites at circRNA junction regions; we expanded the annotation of RBP binding sites, particularly using updated genomic variants and mutations associated with diseases. POSTAR3 is freely available at http://postar.ncrnalab.org.

INTRODUCTION

RNA-binding proteins (RBPs) are essential regulators of RNA function in various biological processes (1,2) and are especially critical in post-transcriptional regulation (3–5). In recent years, several high-throughput sequencing technologies based on crosslinking and immunoprecipitation (CLIP) have been developed to detect genome-wide RBP binding sites (6,7). Moreover, we are able to investigate RNA secondary structure *in vivo* using secondary structure profiling (structure-seq) (8–10), and degradation of cellular RNAs caused by bound miRNAs using degradome sequencing (degradome-seq) (11–13). Together with these high-throughput sequencing technologies, RBP binding could be associated with RNA secondary structure and other types of post-transcriptional regulation events, which would be helpful to understand the post-transcriptional regulation networks that are coordinated by RBPs. Previous studies have revealed the relationship between RBP binding

*To whom correspondence should be addressed. Tel: +86 10 62789217; Fax: +86 10 62789217; Email: zhilu@tsinghua.edu.cn
Correspondence may also be addressed to Yucheng T. Yang. Email: yangyy@fudan.edu.cn
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

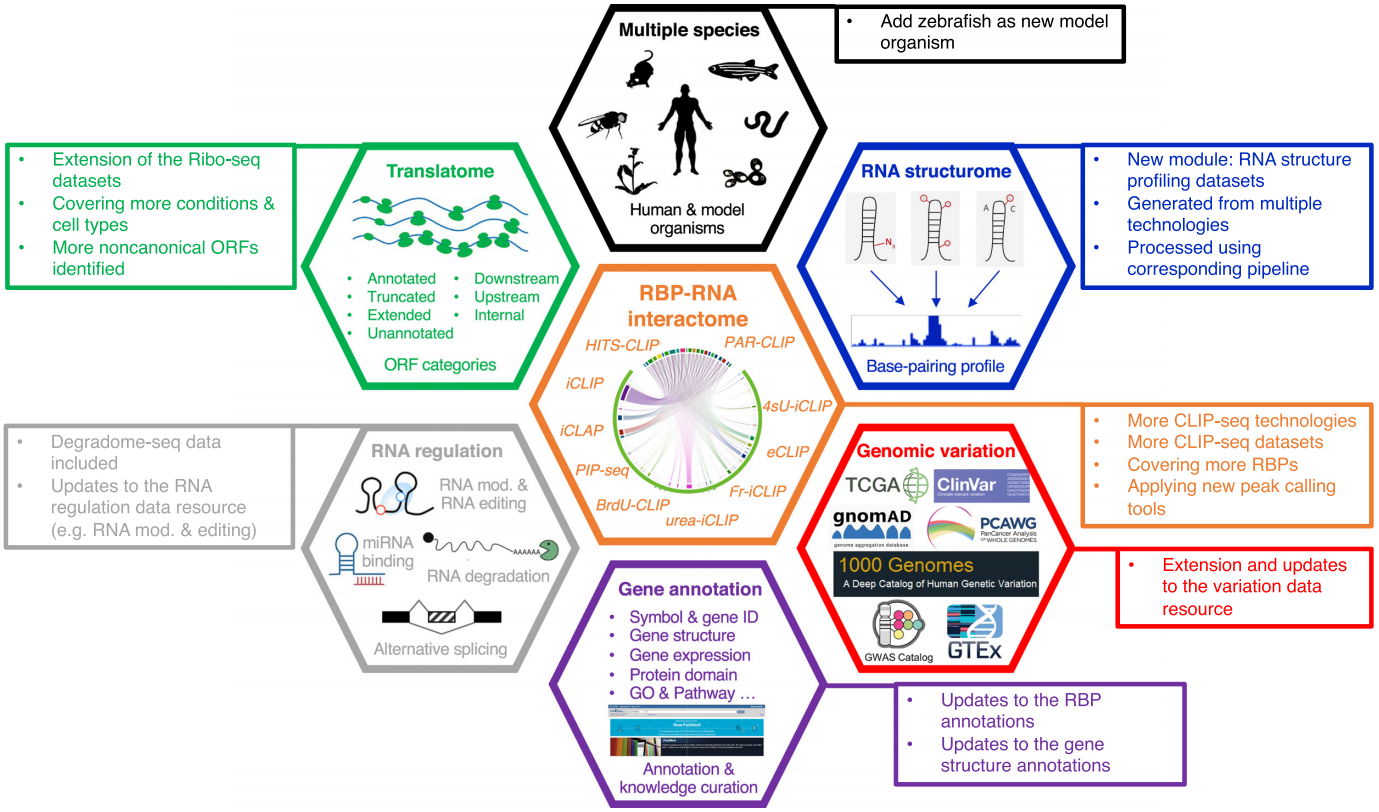

Figure 1. Overview of POSTAR3 database content. Our database is concentrated in RBP-RNA interaction network and reveals information related to RBP binding through CLIP-seq. Other types of post-transcriptional regulation events (RNA modification and editing, genomic variants, disease-associated mutations, secondary structure profile, miRNA-mediated decay, etc.) and translational dynamics from Ribo-seq is associated with RBP binding in order to give users novel insights to the relationship between these events.

and RNA secondary structure (14,15), as well as miRNA-mediated degradation (16). Furthermore, other studies have shown that RBP played an important role in circRNA formation and function (17,18). A platform summarizing RBP binding sites recovered by CLIP-seq and other post-transcriptional regulation events would definitely be helpful for the study in the field.

We have developed a series of CLIPdb/POSTAR databases that focus on the functional annotations of RBP binding sites, as well as their association to other types of post-transcriptional regulation events (19–21). As both the types and volume of these high-throughput dataset have dramatically increased in recent years, it is imperative to update the database to a new version, curating more comprehensive information for RBP binding and post-transcriptional regulation. Here, we present POSTAR3, an update to our existing database of RBP binding records and RNA post-transcriptional regulation (19–21). POSTAR3 curated 339 new CLIP-seq datasets, which spanned nine CLIP-seq technologies from human and other six model species, as well as 300 Ribo-seq datasets covering ∼100 tissue types, cell lines, developmental stages, and experimental conditions from six species, 82 secondary structure profiling datasets, and 83 degradome-seq datasets paired with small RNA sequencing (sRNA-seq) data. We also included RBP binding sites on circRNA junction regions. We associated the RBP binding sites identified from CLIP-seq datasets with other levels of information, including RNA post-transcriptional regulation, genomic variants, disease-associated mutations, secondary structure profile and model, and miRNA-mediated decay from various sources. We also re-designed and modified our database interface to provide an informative display of different types of data and a valuable platform to explore their relationship. We expect that POSTAR3 would be a valuable resource and platform for researchers to investigate post-transcriptional regulation, RNA secondary structure dynamics, miRNA-mediated decay, and their relationship with RBP binding.

DATA COLLECTION UPDATES AND DATA PROCESSING

Updates on the CLIP-seq dataset collection

To expand the spectrum of RBP binding events in our database, we manually collected 339 new publicly available CLIP-seq datasets that used CLIP-seq technologies from Gene Expression Omnibus (GEO) (22), Sequence Read Archive (SRA) (23), ArrayExpress (24), and DDBJ Sequence Read Archive (DRA) (25) (Supplementary Table S1 and Supplementary Table S2). We also updated ENCODE eCLIP to the latest release (26,27), which contains 225 eCLIP datasets from 150 RBPs (Supplementary Table S3). By combining the binding sites from our new datasets with our previous records (21), POSTAR3 contains 1499

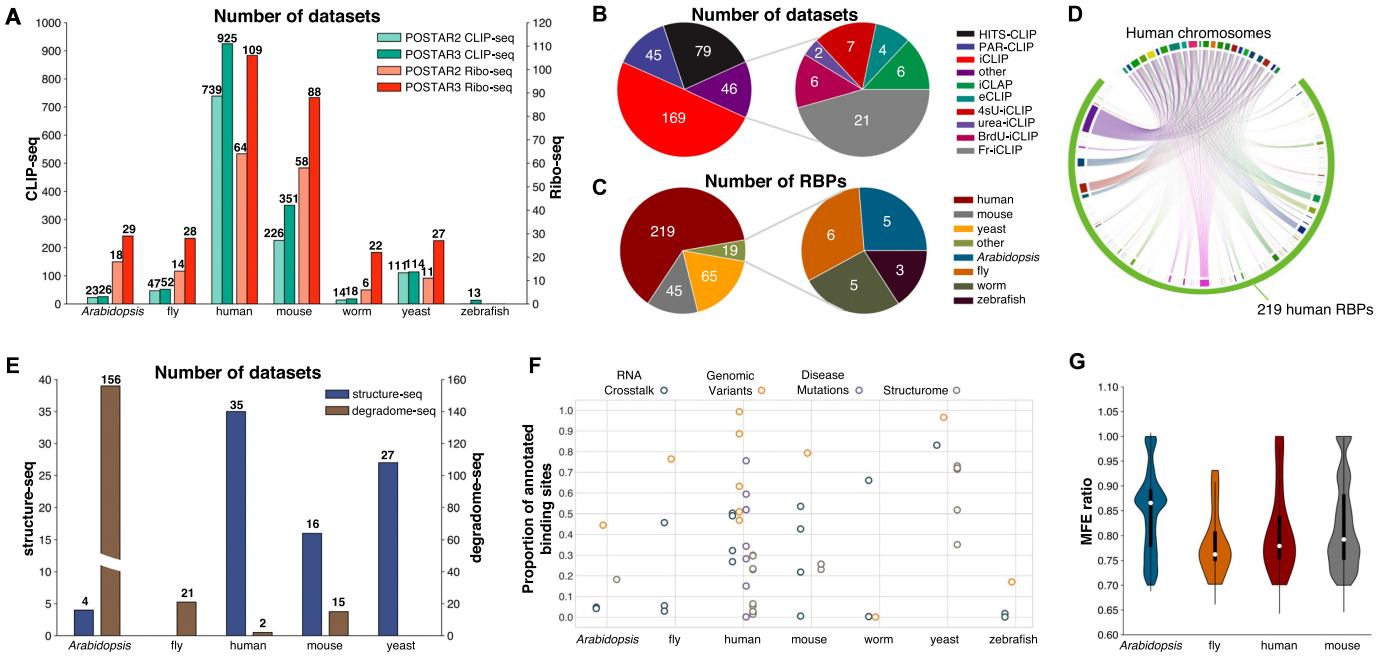

Figure 2. Statistics of data curated in POSTAR3 database. (**A**) Number of CLIP-seq and Ribo-seq datasets in seven species, compared with our previous version POSTAR2. (**B**) Number of newly curated CLIP-seq datasets using different technologies. (**C**) Number of curated RBPs in seven species. (**D**) RBP-RNA interactome network of human in POSTAR3. Arcs on the top represents chromosomes in human, and bottom ones represents RBPs. (**E**) Number of structure-seq and degradome-seq datasets curated in POSTAR3. (**F**) Annotation status of RBP binding sites in different modules. Each dot indicates a specific set of data. (**G**) MFE ratio distribution in all degradome duplex across 4 species.

CLIP-seq datasets from 348 RBPs in total (Figure 1 and Supplementary Table S1), which is a significant improvement in terms of the number of CLIP-seq datasets as well as the RBPs covered (Figure 2A). In summary, comparing to the four CLIP-seq technologies in POSTAR2, POSTAR3 has covered 10 various CLIP-seq technologies (i.e. HITS-CLIP, PAR-CLIP, iCLIP, eCLIP, iCLAP, urea-iCLIP, 4sU-iCLIP, BrdU-CLIP, Fr-iCLIP and PIP-seq). In total, it includes 348 RBPs from seven species (i.e. human, mouse, zebrafish, fly, worm, *Arabidopsis* and yeast) (Figure 2B and C). To our knowledge, POSTAR3 provides the largest collection of RBP binding sites from diverse CLIP-seq technologies and multiple species.

Identification of RBP binding sites from CLIP-seq datasets

For each newly collected CLIP-seq dataset, we followed the same analysis procedure as we developed in POSTAR2 (21) with some modifications. To improve the read mapping quality, we removed unique molecular identifier (UMI) in the raw sequencing file using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit). The actual number of nucleotides that needed to be removed was determined according to the description in the original publications. We also updated the technology-specific peak callers: we used CLIPper (28) (human)/CTK (29) (other species) for HITS-CLIP related technology (HITS-CLIP, BrdU-CLIP), Mi-Clip (30) for PAR-CLIP, and PureCLIP (31) for iCLIP related technology (iCLIP, eCLIP, iCLAP, urea-iCLIP, 4sU-iCLIP, Fr-iCLIP) with default parameters (Supplementary Table S4). For ENCODE eCLIP datasets, we obtained the binding sites from the ENCODE data portal (https://www.encodeproject.org/, May 2020). We also downloaded human RBP binding sites on circRNA junction regions from several recent studies (32,33) and converted the region coordinates to hg38 using liftOver (34). The binding records curated in our database enabled us to construct an RBP-RNA interactome network (Figure 2D).

Adding structure-seq datasets

In POSTAR3, we added a novel 'Structurome' module, where we collected 66 structure-seq datasets (Supplementary Table S5) from GEO (22) and SRA (23) database (Figure 2E), and processed the data as in the original publications. We also collected six processed icSHAPE datasets (Supplementary Table S6) from ENCODE (35). After we obtained the base-pairing information from these datasets, we tried to predict the secondary structure model around RBP binding sites. We extended the RBP binding sites to 150nt flanking the midpoint, and extracted the genomic sequences from the genomes of their respective species as well as the matched structural profiles. Notably, we did this calculation only for the binding sites on long RNAs. We then predicted the RNA secondary structure using Fold from RNAstructure (36) and RNAfold from ViennaRNA (37) with default parameters, in which the structural profile was used as soft constraint. Together with other annotations, POSTAR3 provides users with enough resources to investigate the relationship between RBP binding and other types of post-transcriptional regulatory events (Figure 2F).

Updates of Ribo-seq datasets

We have collected 129 new Ribo-seq datasets (Supplementary Table S7), as well as their matched RNA-seq datasets (Supplementary Table S8) from GEO (22) and SRA (23) database (Figure 2A). We followed the processing procedure from our previous paper (21), with modifications as follows. We used RiboCode (38) to process Ribo-seq mapped reads and identify all types of putative open reading frames (ORFs). We then used Ribotaper (39), ORFscore (40) and RibORF (41) to identify and evaluate translated ORFs in the newly collected datasets. The translation efficiency of the ORF was defined as the RPKM ratio of Ribo-seq to the paired RNA-seq. We obtained the RPKM values of the ORFs based on the raw read density from Ribo-seq datasets, as well as the processed read density from RiboCode (38).

Adding Degradome-seq datasets

In POSTAR3, we also added a Degradome module, where we collected 83 degradome-seq datasets (Supplementary Table S9) and 111 matched small RNA-seq (sRNA-seq) datasets (Supplementary Table S10) from public resource (Figure 2E). To avoid false discovery and unnecessary bias, we excluded datasets without raw fastq files or matched sRNA-seq datasets. Briefly, we removed the adapter sequence using Cutadapt (42) and filtered low quality samples based on the trimming results using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The cleaned fastq files of sRNA-seq datasets were subsequently aligned to annotated miRNA sequences using bowtie2 (43) with the following parameters: -p 12 -n 0 -m 5 –best –strata. We then identified miRNA-mediated degradation events with fastq files converted from sRNA bam files and cleaned degradome-seq fastq files using PAREsnip2 (44) with the stringent mode, Carrinton rule, and the corresponding transcriptome annotations. In addition, we found that the Minimum Free Energy (MFE) ratio (actual binding MFE versus theoretical MFE) of the duplex regions are relatively high in the four species (i.e. human, mouse, fly and *Arabidopsis*) (Figure 2G).

Updates on the annotations of RBP and RBP binding sites

Other than the RBP binding sites itself, we also made significant efforts to update the annotation of RBPs and RBP binding sites. We added annotation information for newly-added RBP and binding sites from zebrafish. We also retrieved information on circRNA from circBase (45) and miRNA from miRbase (46) to annotate respective RNAs. We included overexpression information of the RBP in respective CLIP-seq experiments. We added ∼78 million SNV from 1000 genomes (47), ∼679 million SNV from gnomAD (48), ∼40 million eQTLs, and ∼16 million sQTLs from GTEx (49,50) to annotate RBP binding sites with genomic variants, as well as ∼906k CCLE (51) variants, ∼406k denovo-db (52) variants and ∼7k HmtDB (53) variants as disease-associated mutations. Detailed annotation process for RBP and RBP binding sites is described in Supplemental Methods.

DATABASE FEATURES AND APPLICATIONS

Database and website architecture

All data in POSTAR3 were processed and stored in a MySQL Database (version 5.6.50). We implemented the client-side user interface by the HTML5 and JavaScript libraries, including jQuery (http://jquery.com) and Bootstrap (http://getbootstrap.com), and the server-side using PHP scripts (version 5.6) and JavaScript. Plots of query results in POSTAR3 were generated by plotly.js library (https://plot.ly) and Highcharts (https://www.highcharts.com). Tables of query results were produced by the DataTables JavaScript library (https://www.datatables.net) that allows users to search and sort results. We generated RNA secondary structure visualization by forna (54). We used UCSC Genome Browser (34) to visualize genome in our website. We have tested the web page in several popular browsers including Google Chrome, Safari, Microsoft Edge and Firefox. Users could get access to the website link either on a computer or mobile device.

Overview of the web interface

In POSTAR3, we have updated the website design, which provides a user-friendly web interface for searching, browsing, and downloading data from seven species and eight modules. Here, we briefly describe the implementation of each module.

The ‘CLIPdb’ module provides the annotation of RBPs with their binding sites identified from CLIP-seq datasets. In POSTAR3, we have updated the annotation for the query RBP such as RNA recognition domains, RBP ontology, sequence motifs, and structural preferences in this module. We also provided the overexpression status of the RBP in the original experiment when searching for RBP binding sites. The ‘RBP Binding Sites’ module displays all the RBP binding sites identified with different CLIP-seq technologies and peak calling methods when searching the target gene. The table and network view present the interaction between RBPs and target genes. We also collected genomic location, associated diseases, and expression patterns across different cell lines, tissue types, developmental stages, or conditions for annotation of the target gene. Notably, we generate an overview of the high-occupancy target regions by defining the ‘RBP binding hotspots’ according to the number of RNA binding sites of each 20nt bin on the RNA’s precursor. The ‘RNA Crosstalk’ module provides the interactions between RBP binding sites and other post-transcriptional regulation events, including miRNA targets, RNA modification, and RNA editing. The ‘Genomic Variants’ module and the ‘Disease Mutations’ module integrate SNVs and disease-associated mutations with RBP binding sites to provide insight into the causal variants and the underlying regulatory mechanisms of human diseases. The ‘Translatome’ module characterizes the translation landscape of RNAs with one summary frame and three tables for seven categories of ORFs, respectively. For each data table in POSTAR3, we provide ‘Export data to CSV file’ option for users to download the results of the whole table. Moreover, to provide users with a convenient view of different modules in our database, we have also constructed

a 'POSTAR3 Central' page. At the bottom of each RNA-centric module, there is a link to this 'POSTAR3 Central' page. Users could click the link to enter this page and transfer to other modules by clicking the respective link.

We would like to highlight another two new modules that are included in POSTAR3. The new 'Structurome' module is constructed for characterizing the secondary structure landscape of RNAs. Users can choose a species (e.g. human, mouse, zebrafish, fly, worm, *Arabidopsis* or yeast) and input the desired gene name. POSTAR3 then returns a genome browser, a network and a table: the genome browser contains regions for predicted secondary structure and RBP binding sites corresponding to the table; the network represents interacting RBPs with the queried RNA; the table presents structure information of RBP binding sites for the searched gene. Reactivity score and RNA secondary structure are plotted at each row in the table. The 'Degradome' module provides binding information between miRNA and other types of RNA which leads to the degradation of the other RNA validated by degradome-seq data. Users can obtain detailed information about every validated sRNA-fragment pair by selecting a species and input a target RNA name or small RNA name.

Example applications

POSTAR3 provides users with a friendly and informative platform for exploring the relationship between RBP binding and various types of post-transcriptional regulation events, genomic variants, and translational dynamics. Here, we present two example applications using our database, particularly the two new modules, to demonstrate how to decipher potential regulatory mechanisms related to human disease and response to external stimuli in plants.

In the first example, Ireb2 (also known as Irp2) encodes an essential iron responsive element binding protein in mouse, and its homologous gene has been reported to be related to iron homeostasis in human cells (55). Further studies in mice revealed that Ireb2 could regulate insulin production by influencing iron levels and triggering downstream biochemical reactions (56). However, little effort has been made to demonstrate the relationship between RBP binding and RNA post-transcriptional regulation, especially the secondary structural change during response to iron and production of insulin. When we queried 'Ireb2' in 'Structurome' module in our database, the website returned a genome browser showing the position of RBP binding sites, a network view of interacting RBP of this RNA, and a table displaying all the binding sites and its secondary structure model enhanced by structure profiling data (Figure 3A). In one of the SRSF3 binding sites on Ireb2, we could observe that the binding site was placed at a stem-loop structure (Figure 3B). At the same time, if we query 'Ireb2' in 'Genomic Variants' module, we could retrieve genomic variation information coordinated with RBP binding sites, including one SNV event from dbSNP in this binding site, while the score for the RBP binding site was relatively high (Figure 3C). This variant caused a G changing to an A, thus affecting the secondary structure of this local binding site. These results suggest that this variation could have putative association with the secondary structure change of Ireb2 mRNA, thus influencing the binding of SRSF3, and further affect insulin production and development of diabetes in mouse and human.

Another example is AT2G33830 (also known as DRM2) in *Arabidopsis*. Recent studies have revealed that the expression of AT2G33830 could be related to plants' response to stress and external stimuli, including response to light (57). However, the mechanism of controlled AT2G33830 expression has not been fully understood. When we searched 'AT2G33830' in the new 'Degradome' module, the database returned a table containing peaks of miRNA binding and degradation in degradome-seq data (Figure 3D). All these peaks were identified from a study that investigate the response to excessive light in plants (58), with a relatively high MFE ratio, suggesting stable degradation pairs were formed between the miRNA and the target RNA. Meanwhile, if we search AT2G33830 in the 'Genomic Variants' module, one SNV was found in the base pairing region bound by miRNA, where multiple RBP binding sites with high binding score resided around this region (Figure 3E). Taking all these results together, we could propose a possible mechanism of light response in *Arabidopsis* that the expression of AT2G33830 can be regulated by miRNA binding and degradation, and also affected by SNPs and RBP binding in this local region.

DISCUSSION AND FUTURE DIRECTIONS

We systematically updated our database to the new version, POSTAR3, to enable users to make discoveries and decipher regulatory mechanisms underlying post-transcriptional regulation events related to RBPs. POSTAR3 records ∼50 million RBP binding sites from seven species (human, mouse, zebrafish, fly, worm, *Arabidopsis*, and yeast) and diverse CLIP-seq technologies (HITS-CLIP, PAR-CLIP, iCLIP, PIP-seq, eCLIP, iCLAP, urea-iCLIP, 4sU-iCLIP, BrdU-CLIP, Fr-iCLIP). To our knowledge, POSTAR3 provides the largest collection of RBP binding sites that are identified from CLIP-seq datasets. We annotated the binding sites by incorporating other high-throughput sequencing data, including Ribo-seq, RNA secondary structure profiling, and degradome-seq, as well as other types of post-transcriptional regulation events and genomic variants, shedding light on the relationship between RBP binding and regulatory mechanism at the post-transcriptional and translational level.

Compared with our previous release of POSTAR2, POSTAR3 has made the following updates and improvements: (i) POSTAR3 provides more RBP binding sites that are identified from CLIP-seq datasets and ORFs recovered from Ribo-seq datasets, covering more species and experimental technologies; (ii) POSTAR3 contains two new modules: 'Structurome' and 'Degradome', which provide secondary structure profiling data and model of RBP binding sites, and sRNA-fragment binding records leading to degradation of other RNAs validated by degradome-seq; (iii) POSTAR3 curates RBP binding sites on circRNA junction regions that were recovered from CLIP-seq datasets; (iv) POSTAR3 added annotation information for RBPs, especially the overexpression status information in each CLIP-seq experiment; (v) POSTAR3 updates the annotation for

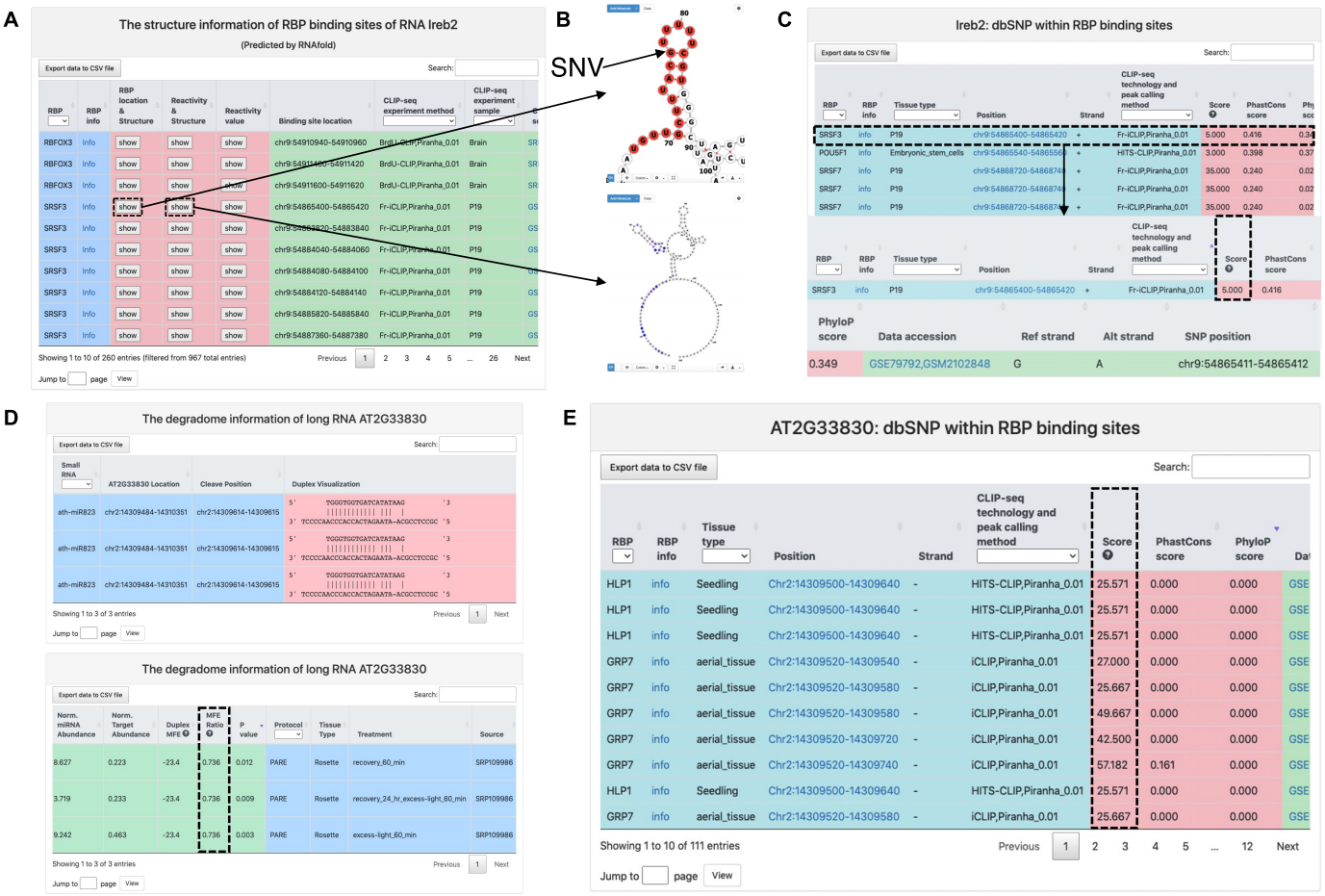

Figure 3. Example applications of POSTAR3: studying Ireb2 in mouse and AT2G33830 in *Arabidopsis*. (**A**) Search of mouse Ireb2 gene in 'Structurome'. In the 'Structurome' module, users could observe the secondary structure model predicted by algorithms enhanced by secondary structure profiling data. (**B**) They could also click the 'RBP location & Structure' or 'Reactivity & structure' button to visualize secondary structure using forna, along with other layers of information. (**C**) Search of mouse Ireb2 gene in 'Disease Mutations' module. 'Disease Mutations' module provides users with information of disease-associated mutations associated with RBP binding in human. Notice that the score for this binding site was relatively high. (**D**) Search of *Arabidopsis* AT2G33830 gene in 'Degradome' module. Search in 'Degradome' module returns a table containing knowledge of miRNA–mRNA binding and degradation peaks, with statistical scores indicating the reliability of the degradation pair. (**E**) Search of *Arabidopsis* AT2G33830 gene in 'Genomic Variants' module. 'Genomic Variants' module gives us information on genomic variants resided within the RBP binding sites.

RBP binding sites, including post-transcriptional regulation events, genomic variants, and disease-associated mutations; (vi) POSTAR3 re-designed and modified our website to build a user-friendly interface for scientists. Since mobile devices are now used more and more widely, we also invested efforts to ensure a compatible web interface on these devices.

It is noticed that sometimes, there is discrepancy between established motifs and motifs discovered from CLIP-seq data in our database. Nevertheless, in our opinion, this should not be a problem. Most experimental motif discovery methods were *in vitro*, such as SELEX or RNAcompete. However, CLIP-seq experiments were conducted *in vivo*, and it is sometimes difficult to identify motifs from CLIP-seq experiments due to protein cofactor interactions or non-specific background (59). As a result, it is possible that our motif discovery process might not be able to find those established motifs from the in vitro experiments. We followed the process pipeline in previous versions of our database to ensure reliable motif detection.

With the development of novel high-throughput sequencing technologies designed to decode the post-transcriptional regulation and release of high-quality data for all kinds of regulatory events, datasets that cover more species and biological conditions will become available to the public in the near future. We would like to continue to incorporate new high-throughput data and improve website for better navigation and exploration of curated data. We will continue to maintain and update our POSTAR3 database to make sure it remains a useful resource for researchers in this area.

DATA AVAILABILITY

POSTAR3 is freely available at http://postar.ncrnalab.org (also at http://lulab.life.tsinghua.edu.cn/postar). Data in POSTAR3 can be downloaded and used in accordance with the GNU Public License and the license of their primary data sources.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Natural Science Foundation of China [31771461, 81972798]; Shanghai Municipal Science and Technology Major Project [2018SHZDZX01]; Natural Science Foundation of Shanghai [21ZR1408100]; 111 Project [B18015]; Tsinghua-Foshan Innovation Special Fund; Fok Ying-Tong Education Foundation; Beijing Advanced Innovation Center for Structural Biology; Open Research Fund Program of Beijing National Research Center for Information Science and Technology; Bio-Computing Platform of Tsinghua University Branch of China National Center for Protein Sciences; Austrian Science Fund FWF Standalone Grants [P 30550, P 30680-B21]; Bioinformatics Platform of National Center for Protein Sciences (Beijing) [2021-NCPSB-005]. Funding for open access charge: National Natural Science Foundation of China [31771461].
Conflict of interest statement. None declared.

REFERENCES

1. Pereira,B., Billaud,M. and Almeida,R. (2017) RNA-binding proteins in cancer: old players and new actors. *Trends Cancer*, **3**, 506–528.
2. Taylor,J.P., Brown,R.H. Jr and Cleveland,D.W. (2016) Decoding ALS: from genes to mechanism. *Nature*, **539**, 197–206.
3. Apicco,D.J., Zhang,C., Maziuk,B., Jiang,L., Ballance,H.I., Boudeau,S., Ung,C., Li,H. and Wolozin,B. (2019) Dysregulation of RNA Splicing in Tauopathies. *Cell Rep.*, **29**, 4377–4388.
4. Blanc,V., Navaratnam,N., Henderson,J.O., Anant,S., Kennedy,S., Jarmuz,A., Scott,J. and Davidson,N.O. (2001) Identification of GRY-RBP as an apolipoprotein B RNA binding protein that interacts with both apobec-1 and with apobec-1 complementation factor (ACF) to modulate C-to-U editing. *Gastroenterology*, **120**, A306–A306.
5. McCloskey,A., Taniguchi,I., Shinmyozu,K. and Ohno,M. (2012) hnRNP C tetramer measures RNA length to classify RNA polymerase II transcripts for export. *Science*, **335**, 1643–1646.
6. Lee,F.C.Y. and Ule,J. (2018) Advances in CLIP technologies for studies of protein-RNA interactions. *Mol. Cell*, **69**, 354–369.
7. Hafner,M., Katsantoni,M., Köster,T., Marks,J., Mukherjee,J., Staiger,D., Ule,J. and Zavolan,M. (2021) CLIP and complementary methods. *Nat Rev Methods Primers*, **1**, 20.
8. Kertesz,M., Wan,Y., Mazor,E., Rinn,J.L., Nutter,R.C., Chang,H.Y. and Segal,E. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature*, **467**, 103–107.
9. Rouskin,S., Zubradt,M., Washietl,S., Kellis,M. and Weissman,J.S. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, **505**, 701–705.
10. Spitale,R.C., Flynn,R.A., Zhang,Q.C., Crisalli,P., Lee,B., Jung,J.W., Kuchelmeister,H.Y., Batista,P.J., Torre,E.A., Kool,E.T. *et al.* (2015) Structural imprints in vivo decode RNA regulatory mechanisms. *Nature*, **519**, 486–490.
11. German,M.A., Luo,S., Schroth,G., Meyers,B.C. and Green,P.J. (2009) Construction of parallel analysis of RNA ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. *Nat. Protoc.*, **4**, 356–362.
12. Addo-Quaye,C., Eshoo,T.W., Bartel,D.P. and Axtell,M.J. (2008) Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. *Curr. Biol.*, **18**, 758–762.
13. Gregory,B.D., O'Malley,R.C., Lister,R., Urich,M.A., Tonti-Filippini,J., Chen,H., Millar,A.H. and Ecker,J.R. (2008) A link between RNA metabolism and silencing affecting Arabidopsis development. *Dev. Cell*, **14**, 854–866.
14. Taliaferro,J.M., Lambert,N.J., Sudmant,P.H., Dominguez,D., Merkin,J.J., Alexis,M.S., Bazile,C. and Burge,C.B. (2016) RNA sequence context effects measured in vitro predict in vivo protein binding and regulation. *Mol. Cell*, **64**, 294–306.
15. Dominguez,D., Freese,P., Alexis,M.S., Su,A., Hochman,M., Palden,T., Bazile,C., Lambert,N.J., Van Nostrand,E.L., Pratt,G.A. *et al.* (2018) Sequence, structure, and context preferences of human RNA binding proteins. *Mol. Cell*, **70**, 854–867.
16. Hou,C.Y., Wu,M.T., Lu,S.H., Hsing,Y.I. and Chen,H.M. (2014) Beyond cleaved small RNA targets: unraveling the complexity of plant RNA degradome data. *BMC Genomics*, **15**, 15.
17. Du,W.W., Zhang,C., Yang,W.N., Yong,T.Q., Awan,F.M. and Yang,B.B. (2017) Identifying and characterizing circRNA-protein interaction. *Theranostics*, **7**, 4183–4191.
18. Zang,J., Lu,D. and Xu,A. (2020) The interaction of circRNAs and RNA binding proteins: An important part of circRNA maintenance and function. *J. Neurosci. Res.*, **98**, 87–97.
19. 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.
20. Hu,B., Yang,Y.T., Huang,Y., Zhu,Y. and Lu,Z.J. (2017) POSTAR: a platform for exploring post-transcriptional regulation coordinated by RNA-binding proteins. *Nucleic Acids Res.*, **45**, D104–D114.
21. Zhu,Y., Xu,G., Yang,Y.T., Xu,Z., Chen,X., Shi,B., Xie,D., Lu,Z.J. and Wang,P. (2019) POSTAR2: deciphering the post-transcriptional regulatory logics. *Nucleic Acids Res.*, **47**, D203–D211.
22. Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M., Holko,M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets-update. *Nucleic Acids Res.*, **41**, D991–D995.
23. Sayers,E.W., Beck,J., Bolton,E.E., Bourexis,D., Brister,J.R., Canese,K., Comeau,D.C., Funk,K., Kim,S., Klimke,W. *et al.* (2021) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **49**, D10–D17.
24. Athar,A., Fullgrabe,A., George,N., Iqbal,H., Huerta,L., Ali,A., Snow,C., Fonseca,N.A., Petryszak,R., Papatheodorou,I. *et al.* (2019) ArrayExpress update - from bulk to single-cell expression data. *Nucleic Acids Res.*, **47**, D711–D715.
25. Kodama,Y., Shumway,M., Leinonen,R. and International Nucleotide Sequence Database, C. (2012) The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Res.*, **40**, D54–D56.
26. Van Nostrand,E.L., Pratt,G.A., Yee,B.A., Wheeler,E.C., Blue,S.M., Mueller,J., Park,S.S., Garcia,K.E., Gelboin-Burkhart,C., Nguyen,T.B. *et al.* (2020) Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. *Genome Biol.*, **21**, 90.
27. Van Nostrand,E.L., Freese,P., Pratt,G.A., Wang,X., Wei,X., Xiao,R., Blue,S.M., Chen,J.Y., Cody,N.A.L., Dominguez,D. *et al.* (2020) A large-scale binding and functional map of human RNA-binding proteins. *Nature*, **583**, 711–719.
28. Lovci,M.T., Ghanem,D., Marr,H., Arnold,J., Gee,S., Parra,M., Liang,T.Y., Stark,T.J., Gehman,L.T., Hoon,S. *et al.* (2013) Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat. Struct. Mol. Biol.*, **20**, 1434–1442.
29. Shah,A., Qian,Y., Weyn-Vanhentenryck,S.M. and Zhang,C. (2017) CLIP Tool Kit (CTK): a flexible and robust pipeline to analyze CLIP sequencing data. *Bioinformatics*, **33**, 566–567.
30. Wang,T., Chen,B., Kim,M., Xie,Y. and Xiao,G. (2014) A model-based approach to identify binding sites in CLIP-Seq data. *PLoS One*, **9**, e93248.
31. 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.
32. Okholm,T.L.H., Sathe,S., Park,S.S., Kamstrup,A.B., Rasmussen,A.M., Shankar,A., Chua,Z.M., Fristrup,N., Nielsen,M.M., Vang,S. *et al.* (2020) Transcriptome-wide profiles of circular RNA and RNA-binding protein interactions reveal effects on circular RNA biogenesis and cancer pathway expression. *Genome Med*, **12**, 112.
33. Dudekulay,D.B., Panda,A.C., Grammatikakis,I., De,S., Abdelmohsen,K. and Gorospe,M. (2016) CircInteractome: a web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biol*, **13**, 34–42.
34. Navarro Gonzalez,J., Zweig,A.S., Speir,M.L., Schmelter,D., Rosenbloom,K.R., Raney,B.J., Powell,C.C., Nassar,L.R.,

Maulding,N.D., Lee,C.M. *et al.* (2021) The UCSC Genome Browser database: 2021 update. *Nucleic Acids Res.*, **49**, D1046–D1057.
35. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
36. Reuter,J.S. and Mathews,D.H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, **11**, 129.
37. Lorenz,R., Bernhart,S.H., Honer Zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
38. Xiao,Z., Huang,R., Xing,X., Chen,Y., Deng,H. and Yang,X. (2018) De novo annotation and characterization of the translatome with ribosome profiling data. *Nucleic Acids Res.*, **46**, e61.
39. Calviello,L., Mukherjee,N., Wyler,E., Zauber,H., Hirsekorn,A., Selbach,M., Landthaler,M., Obermayer,B. and Ohler,U. (2016) Detecting actively translated open reading frames in ribosome profiling data. *Nat. Methods*, **13**, 165–170.
40. Bazzini,A.A., Johnstone,T.G., Christiano,R., Mackowiak,S.D., Obermayer,B., Fleming,E.S., Vejnar,C.E., Lee,M.T., Rajewsky,N., Walther,T.C. *et al.* (2014) Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J.*, **33**, 981–993.
41. Ji,Z., Song,R.S., Regev,A. and Struhl,K. (2015) Many lncRNAs, 5 ' UTRs, and pseudogenes are translated and some are likely to express functional proteins. *Elife*, **4**, e08890.
42. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10–12.
43. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
44. Thody,J., Folkes,L., Medina-Calzada,Z., Xu,P., Dalmay,T. and Moulton,V. (2018) PAREsnip2: a tool for high-throughput prediction of small RNA targets from degradome sequencing data using configurable targeting rules. *Nucleic Acids Res.*, **46**, 8730–8739.
45. Glazar,P., Papavasileiou,P. and Rajewsky,N. (2014) circBase: a database for circular RNAs. *RNA*, **20**, 1666–1670.
46. Griffiths-Jones,S., Grocock,R.J., van Dongen,S., Bateman,A. and Enright,A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.
47. 1000 Genomes Project Consortium, Auton,A., Brooks,L.D., Durbin,R.M., Garrison,E.P., Kang,H.M., Korbel,J.O., Marchini,J.L., McCarthy,S., McVean,G.A. *et al.* (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.
48. Karczewski,K.J., Francioli,L.C., Tiao,G., Cummings,B.B., Alfoldi,J., Wang,Q.B., Collins,R.L., Laricchia,K.M., Ganna,A., Birnbaum,D.P. *et al.* (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, **581**, 434–443.
49. Aguet,F., Brown,A.A., Castel,S.E., Davis,J.R., He,Y., Jo,B., Mohammadi,P., Park,Y., Parsana,P., Segre,A.V. *et al.* (2017) Genetic effects on gene expression across human tissues. *Nature*, **550**, 204–213.
50. Aguet,F., Barbeira,A.N., Bonazzola,R., Brown,A., Castel,S.E., Jo,B., Kasela,S., Kim-Hellmuth,S., Liang,Y.Y., Parsana,P. *et al.* (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, **369**, 1318–1330.
51. Ghandi,M., Huang,F.W., Jane-Valbuena,J., Kryukov,G.V., Lo,C.C., McDonald,E.R., Barretina,J., Gelfand,E.T., Bielski,C.M., Li,H. *et al.* (2019) Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*, **569**, 503–508.
52. Turner,T.N., Yi,Q., Krumm,N., Huddleston,J., Hoekzema,K., HA,F.S., Doebley,A.L., Bernier,R.A., Nickerson,D.A. and Eichler,E.E. (2017) denovo-db: a compendium of human de novo variants. *Nucleic Acids Res.*, **45**, D804–D811.
53. Clima,R., Preste,R., Calabrese,C., Diroma,M.A., Santorsola,M., Scioscia,G., Simone,D., Shen,L., Gasparre,G. and Attimonelli,M. (2017) HmtDB 2016: data update, a better performing query system and human mitochondrial DNA haplogroup predictor. *Nucleic Acids Res.*, **45**, D698–D706.
54. Kerpedjiev,P., Hammer,S. and Hofacker,I.L. (2015) Forna (force-directed RNA): simple and effective online RNA secondary structure diagrams. *Bioinformatics*, **31**, 3377–3379.
55. Wang,H., Shi,H., Rajan,M., Canarie,E.R., Hong,S., Simoneschi,D., Pagano,M., Bush,M.F., Stoll,S., Leibold,E.A. *et al.* (2020) FBXL5 regulates IRP2 stability in iron homeostasis via an oxygen-responsive [2Fe2S] cluster. *Mol. Cell*, **78**, 31–41.
56. dos Santos,M.C.F., Anderson,C.P., Neschen,S., Zumbrennen-Bullough,K.B., Romney,S.J., Kahle-Stephan,M., Rathkolb,B., Gailus-Durner,V., Fuchs,H., Wolf,E. *et al.* (2020) Irp2 regulates insulin production through iron-mediated Cdkal1-catalyzed tRNA modification. *Nat. Commun.*, **11**, 296.
57. Rae,G.M., Uversky,V.N., David,K. and Wood,M. (2014) DRM1 and DRM2 expression regulation: potential role of splice variants in response to stress and environmental factors in Arabidopsis. *Mol. Genet. Genomics*, **289**, 317–332.
58. Crisp,P.A., Ganguly,D.R., Smith,A.B., Murray,K.D., Estavillo,G.M., Searle,I., Ford,E., Bogdanovic,O., Lister,R., Borevitz,J.O. *et al.* (2017) Rapid recovery gene downregulation during excess-light stress and recovery in Arabidopsis. *Plant Cell*, **29**, 1836–1863.
59. Friedersdorf,M.B. and Keene,J.D. (2014) Advancing the functional utility of PAR-CLIP by quantifying background binding to mRNAs and lncRNAs. *Genome Biol.*, **15**, R2.