Published online 2 November 2022 *Nucleic Acids Research, 2023, Vol. 51, Database issue* ***D1549–D1557***
https://doi.org/10.1093/nar/gkac971

RBP Image Database: A resource for the systematic characterization of the subcellular distribution properties of human RNA binding proteins

Louis Philip Benoit Bouvrette[1,2,†], Xiaofeng Wang[1,†], Jonathan Boulais[1], Jian Kong[1], Easin Uddin Syed[1,2], Steven M. Blue[3], Lijun Zhan[4], Sara Olson[4], Rebecca Stanton[3], Xintao Wei[4], Brian Yee[3], Eric L. Van Nostrand[3,5], Xiang-Dong Fu[3], Christopher B. Burge[6], Brenton R. Graveley[4], Gene W. Yeo[3] and Eric Lécuyer[1,2,7,*]

[1]Institut de Recherches Cliniques de Montréal (IRCM) Montréal, Québec, Canada, [2]Département de Biochimie et Médecine Moléculaire, Université de Montréal, Montréal, Québec, Canada, [3]Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA, USA, [4]Department of Genetics and Genome Sciences, UConn Health Center, Farmington, CT, USA, [5]Present Address: Verna & Marrs McLean Department of Biochemistry & Molecular Biology and Therapeutic Innovation Center, Baylor College of Medicine, Houston, TX, USA, [6]Program of Computational and Systems Biology, Department of Biology, MIT, Cambridge, MA, USA and [7]Division of Experimental Medicine, McGill University, Montréal, Québec, Canada

Received August 15, 2022; Revised October 04, 2022; Editorial Decision October 07, 2022; Accepted October 31, 2022

ABSTRACT

RNA binding proteins (RBPs) are central regulators of gene expression implicated in all facets of RNA metabolism. As such, they play key roles in cellular physiology and disease etiology. Since different steps of post-transcriptional gene expression tend to occur in specific regions of the cell, including nuclear or cytoplasmic locations, defining the subcellular distribution properties of RBPs is an important step in assessing their potential functions. Here, we present the RBP Image Database, a resource that details the subcellular localization features of 301 RBPs in the human HepG2 and HeLa cell lines, based on the results of systematic immuno-fluorescence studies conducted using a highly validated collection of RBP antibodies and a panel of 12 markers for specific organelles and subcellular structures. The unique features of the RBP Image Database include: (i) hosting of comprehensive representative images for each RBP-marker pair, with ∼250,000 microscopy images; (ii) a manually curated controlled vocabulary of annotation terms detailing the localization features of each factor; and (iii) a user-friendly interface allowing the rapid querying of the data by target or annotation. The RBP Image Database is freely available at https://rnabiology.ircm.qc.ca/RBPImage/.

INTRODUCTION

From the moment they are synthesized, RNA molecules are subjected to an array of co- and post-transcriptional regulatory mechanisms that dictate their maturation, function, and fate in the cell. These events are in large part controlled by RNA binding proteins (RBPs), which assemble within dynamic ribonucleoprotein modules. RBPs represent a major class of cellular regulatory factors and, from an evolutionary point of view, include some of the most deeply conserved cellular machineries (1–3). As a group, they control all facets of RNA metabolism, including RNA synthesis, splicing, cleavage and polyadenylation, epitranscriptomic modifications, intracellular transport, translation and degradation (2). These factors are classified according to the type of RNA binding domain (RBD) they contain, which dictates their capacity to recognize specific sequence or structural features within target RNAs (2,4–7). In the last decade, a number of surveys have aimed to define the human RBP repertoire, either using literature curation (7), sequence-oriented and interaction network-informed computational predictions (2,8,9), affinity-based proteomic profiling studies to identify proteins associated with polyadenylated (10,11) and non-polyadenylated RNA in cells (12–14), or methods using RNase digestion to define RNA-dependent protein interactions (15,16). Combined, these efforts indicate that the human genome encodes ∼2,500 RBPs (10–14), painting a staggeringly complex portrait of post-transcriptional gene regulation.

*To whom correspondence should be addressed. Email: eric.lecuyer@ircm.qc.ca
†Co-first authors.

While the properties of most of RBPs remain poorly defined, several systematic experimental pipelines have been deployed to begin characterizing their functions in RNA regulation. This includes methods to determine their RNA binding specificities, either using *in vitro* selection strategies with purified RBPs and randomized RNA pools (17,18) or crosslinking and immunoprecipitation (CLIP) approaches to map their transcriptomic binding profiles (19–21), to assess the impact of their functional perturbations on transcriptome integrity or to evaluate their subcellular localization properties (21). Moreover, a number of databases have been established to compile the results of proteome-wide RBP isolation studies [e.g. RBP2GO (22), EuRBPDP (23)], CLIP-seq datasets detailing their transcriptome-binding features [e.g. CLIPdb (24), doRINA (25)], or to map putative RBP binding sites across human and model organism transcriptomes [e.g. CISBP-RNA (17), RBPmap (26), ATtRACT (27), oRNAment (28)].

The various steps of post-transcriptional gene regulation tend to be carried out in different subregions of the cell, ranging from subnuclear domains to cytoplasmic organelles and non-membranous phase-separated bodies. As such, defining the intracellular localization properties of RBPs is an important feature in helping to dissect their properties in RNA regulation within subcellular space. Herein, we describe the RBP Image Database, a resource that catalogues and characterizes the subcellular localization patterns of 301 RBPs in relation to a dozen markers for various organelles and non-membrane delimited structures in two human cell lines, HepG2 and HeLa, which have been widely used in genomics studies. The resource houses ∼250,000 microscopy images, enhanced by manually curated controlled vocabulary annotations, which detail the subcellular distribution features of a significant fraction of human RBPs. Both images and pattern annotations can be interrogated through web-based 'protein centric' or 'general overview' search functionalities or downloaded for external usage. The RBP Image Database is freely accessible at https://rnabiology.ircm.qc.ca/RBPImage/.

MATERIALS AND METHODS

Integration of Systematic Immuno-Fluorescence (IF) Data

The RBP Image Database houses imaging data from systematic immuno-fluorescence (IF) assays conducted using the human hepatocellular carcinoma cell line HepG2 and the HeLa cervical cancer cell line (Figure 1A). These assays were conducted using a collection of highly-validated RBP antibodies, for which the specificities were validated through work in the ENCODE consortium (see: www.encodeproject.org), using immuno-precipitation (IP), Western blotting and shRNA/CRISPR-Cas9 loss-of-function methodologies (8,21). Collectively, these RBPs have been implicated in diverse post-transcriptional regulatory steps (e.g. RNA splicing, stability, translation), include a variety of RBD classes, with a number of them having yet poorly defined functions in RNA metabolism (21). IF data for a total of 301 RBPs were interrogated in these assays, 292 of which were commonly used across both cell lines. Each RBP was imaged in conjunction with a DNA counter stain (i.e. DAPI) and with marker antibodies labeling a dozen distinct organelles and subcellular structures (Figure 1B).

Antibodies and IF Procedure

Rabbit polyclonal RBP antibodies were obtained from several commercial vendors (i.e. Bethyl/Fortis Life Sciences, GeneTex and MBL) and subjected to validations steps as detailed previously (8). Subcellular marker antibodies were as follows: mouse anti-CD63 (ab8219, Abcam); mouse anti-Coilin (GTX11822, GeneTex); mouse anti-DCP1 (ENZSPA827D, Enzo Life Sciences); mouse anti-Fibrillarin (ab4566, Abcam); mouse anti-GM130 (610822, BD); mouse anti-KDEL (SC-100706, Santa Cruz Biotech); mouse anti-Phospho Tyrosine (9411A, NEB); mouse anti-PML (sc-966, Santa Cruz Biotech); mouse anti-SC35 (GTX11826, GeneTex); rat anti-aTubulin (MCA78G, Serotec). For staining with the Mitotracker (Molecular Probes, M22426), cells were incubated with 100nM of Mitotracker in tissue culture media for 45 min at 37°C prior to fixation. For staining with Phalloidin (Sigma, P5282), cells were incubated with 50ug/ml of Phalloidin for 20 min prior DAPI staining.

HepG2 and HeLa cells were cultured in Dulbecco's modified Eagle's medium (DMEM) (SH30022.01, Hyclone) supplemented with 10% FBS and 1% penicillin/streptavidin (450–201-EL, Wisent) at 37°C and 5% CO_2. For IFs, cells were seeded in Poly-L-Lysine coated 96-well clear bottom half-area microplates (Corning, #3882), at a concentration of 2,000 cells per well in DMEM + 10% FBS. After 72h in standard growth conditions (i.e. 37°C and 5% CO2), cells were fixed with 4% formaldehyde, permeabilized in PBS + 0.5% Triton X-100 and blocked in PBS + 0.2% Tween-20 + 2% BSA (PBTB), all conducted for 20 min at room temperature. Primary antibodies directed against specific RBPs (all rabbit antibodies) and marker proteins were subsequently applied to the cells at a final concentration of 2 μg/mL in PBTB and incubated overnight at 4°C. The cells were next washed 3 × in PBST (10min per wash) and incubated with secondary antibodies (Alexa647 donkey anti-rabbit and Alexa488 donkey anti-mouse, both diluted 1:500 in PBTB) for 90 min at room temperature. After 3 × PBTB washes, the cells were counter stained with DAPI for 5 min, washed 3 × in PBS and stored in PBS at 4°C prior to imaging.

Microscopy and Image Processing

Images were acquired on an ImageXpress Micro high content screening system (Molecular Devices Inc). For each RBP-marker combination, 10–20 high resolution images were acquired in three channels, DAPI, FITC and Cy5 by using a 40 × objective. An auto-exposure function was used to achieve optimal imaging intensity. The RBP channel exposure time ranged from 250–3000 ms. The DAPI channel ranged from 50–100 ms. The marker channel ranged from 100–500 ms. Laser based auto-focusing was used to have the best focus in a fast automated fashion. Altogether, >388,000 images were captured across all RBP/marker co-labeling conditions, which were then subjected to quality

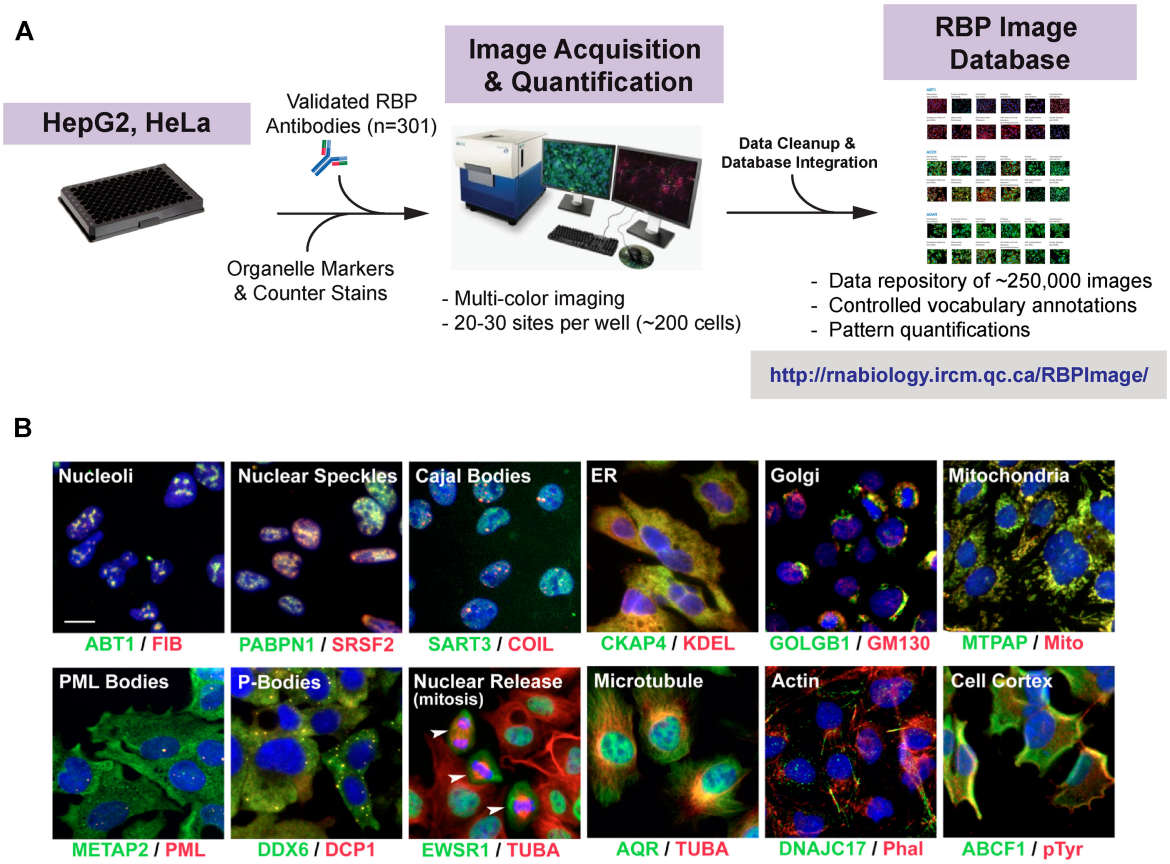

Figure 1. Overview of microscopy imaging data integrated within the RBP Image Database. (A) Immuno-fluorescence pipeline implemented to systematically document the subcellular localization features of 301 RBPs in HepG2 and HeLa cells, as detailed previously in Van Nostrand et al (21). (B) Examples of RBPs localized to the various subcellular structures profiled in the RBP Image Database (white text). The labeled RBPs (green) and co-labeled structures (red) are indicated below each figure. DAPI = Blue. The arrow heads indicate mitotic cells.

control processing steps to eliminate images bearing staining artifacts or with insufficient numbers of cells, and 10–20 images were selected for each RBP-marker pair. An in house Matlab script was run on all images to adjust their intensities (imadjust) and transform gray scale images to color images (ind2grb), in order to have the most accurate image possible.

To complement the imaging data, a controlled vocabulary of annotation terms was implemented to categorize the subcellular localization features of each RBP following manual data inspection by a team of expert annotators. Therefore, all images were manually inspected and annotated using this controlled vocabulary of localization descriptors devised to document RBP subcellular localization features.

Database Implementation

The RBP Image Database is built upon the relational database management system (RDBMS) MySQL. The server side back end of the web application is implemented with the scripting language PHP. The client interface is made in HTML with the inclusion, for a greater interactive experience, of the JavaScript libraries jquery, fancybox, tree, and datatables. The layout styling was created with Bootstrap and Bootstrap-material-design.

Bioinformatics Analyses

Pearson correlations were calculated in R (https://www.r-project.org/) and illustrated with the corrplot package. The relative information (RI) of RNA biotypes was extracted from previously published RBP eCLIP experiments conducted in HepG2 cells (20) and merged to our RBP localization dataset. For each RNA biotype, the RI distribution of RBPs observed at a particular localization was compared against the RI distribution of unlocalized RBPs by a one-sided Wilcoxon rank-sum test in R with the stats package. After testing separately both alternative hypotheses (greater and lesser), we adjusted the Log10(p-values) sign accordingly to the alternative hypothesis showing a p-value < 0.05 (positive for the greater hypothesis and negative for the lesser hypothesis). From these adjusted p-values, a heatmap was generated by using the pheatmap R package.

RESULTS

Web Interface and Functionality

The RBP Image Database harbours a collection of 247,008 representative microscopy images (i.e. 140,725 for HeLa and 106,283 for HepG2) detailing the subcellular localization features of 301 RBPs in relation to a dozen subcellular markers, accompanied by localization pattern de-

scriptors tabulated through visual analysis by expert curators, utilizing a controlled vocabulary of annotation terms. The database interface allows users to access the imaging data, pattern annotations and quality control assessments for specific RBP candidates across interrogated cell lines and subcellular compartments, with integrated links to external metadata information (e.g. Ensemble gene descriptions, antibody validation data, antibody vendor details, ENCODE/ENCORE consortium links). The web portal also contains a tutorial page to help the user navigate the database, including background information on the underlying experimental pipeline, available image/annotation data formats and data retrieval options. The primary functionalities include:

Search by RBP

This functionality allows the user to query the database for a specific target RBP of interest (Figure 2A). The search first returns a set of single representative images depicting the co-staining of the candidate RBP in conjunction with each subcellular marker. Each row of images, corresponding to a specific RBP/marker co-labeling, is divided by column showing 1) the RBP alone, 2) the subcellular marker alone, 3) the merged RBP/marker co-labeling, and 4) the merged image of the RBP and marker with DAPI staining (nuclear counter-staining). Mouse clicking on the image thumbnails triggers the appearance of a larger high-resolution version of the image and keyboard arrow keys can be used to navigate from one image to the next. The complete list of manually-curated annotation terms for that RBP is also displayed next to the image set. Finally, by mouse clicking on the name of the markers at the left of each column, the user is guided to a page containing a minimum of 5 additional representative images of the particular RBP/marker co-labeling. By selecting the 'Download set' tab, the user can download a csv file containing the annotation matrix for the selected cell line, as well as all the images pertaining to the selected RBP, with specific RBP/marker co-labeling images being grouped in the same subfolder.

Search by cellular location

This functionality allows the user to query the database for a specific cell line and annotation. The user can query the database for all images annotated with all selected labels (e.g. Cytoplasm AND Endoplasmic reticulum) or any of the selected labels (e.g. Cajal bodies OR P bodies). This search returns a single page displaying representative triple-labeled images (i.e. RBP, marker and DAPI overlays) of each RBP annotated with the queried labels (Figure 2B). Mouse clicking on the RBP names leads to the same RBP page, as the user would see in gene symbol-based searches. Also, by clicking on the 'show all RBPs' button, a list of all RBPs that pertain to the annotation query will appear, which can be retrieved by the user.

Browse all images

This functionality allows the user to query the database for all images acquired for a given cell line. Similar to the 'search by annotation' function, this search option returns a single page displaying representative triple-labeled images (i.e. RBP, marker and DAPI overlays) of all RBPs labeled in the specified cell line, which appear in alphabetic order. By selecting 'Download, this search box also allows the user to download, instead of browsing, all images included in the database for the selected cell type, along with a csv file of the binary annotation matrix summarizing the results of all RBP-marker co-labeling experiments.

Annotation Table

By selecting 'View Annotation Table' on the main page, the user is taken to a page containing a summarizing table that lists all of the RBPs that were imaged, with active links towards the imaging data pages for each cell line, information about quality control assessments, subcellular localization characteristics, as well as link outs to ENSEMBL gene description pages, RBP antibody validation data generated by the Encyclopedia of RNA Elements/ENCORE group initiative and antibody vendor pages. Finally, by clicking on the blue tab on this page, all of the information in this summarizing table can be downloaded by the user in a common csv file.

Characteristics of RBP Subcellular Distribution

The RBP Image Database was created to help characterize the subcellular distribution properties of human RBPs in the broadly utilized human cell lines HepG2 and HeLa, with a particular focus on subcellular structures known to be important for RNA regulation. Altogether, the assembled data reveal that RBPs exhibit a broad diversity of subcellular localization patterns across all of the organelles and subcellular structures demarcated by the tested marker antibodies (Figure 3A). Indeed, prevalent RBP targeting is observed to specific locations of the cytoplasm and nucleus, such as endosomal networks (∼75–80%), nuclear speckles (∼30%), mitochondria (∼25%) and nucleoli (15–20%). These distribution patterns are broadly consistent between HeLa and HepG2 cells, with most localization classes displaying > 75% similarity in RPB repertoires between both cell lines (Figure 3A). For some structures, such as focal adhesions labeled with a phospho-tyrosine antibody (29), co-localizing RBPs were almost exclusively observed in HeLa cell data, reflecting the enhanced adhesive properties of these cells compared to HepG2. While the utilized antibodies have been thoroughly characterized for specificity (8,21), to validate these localization patterns, we performed a detailed comparison to imaging studies conducted with independent antibodies generated in the context of the Human Protein Atlas project (30). While this comparison reveals a general agreement between the studies, we found our annotation data to be richer with regards to the number of patterns described (Figure 3B). This likely reflects the enhanced annotation precision resulting from systematic co-labeling experiments of each RBP in relation to all subcellular markers, providing unambiguous evidence of co-localization to specific structures.

We next performed correlative analyses of the different subgroups of localized RBPs in order to assess their

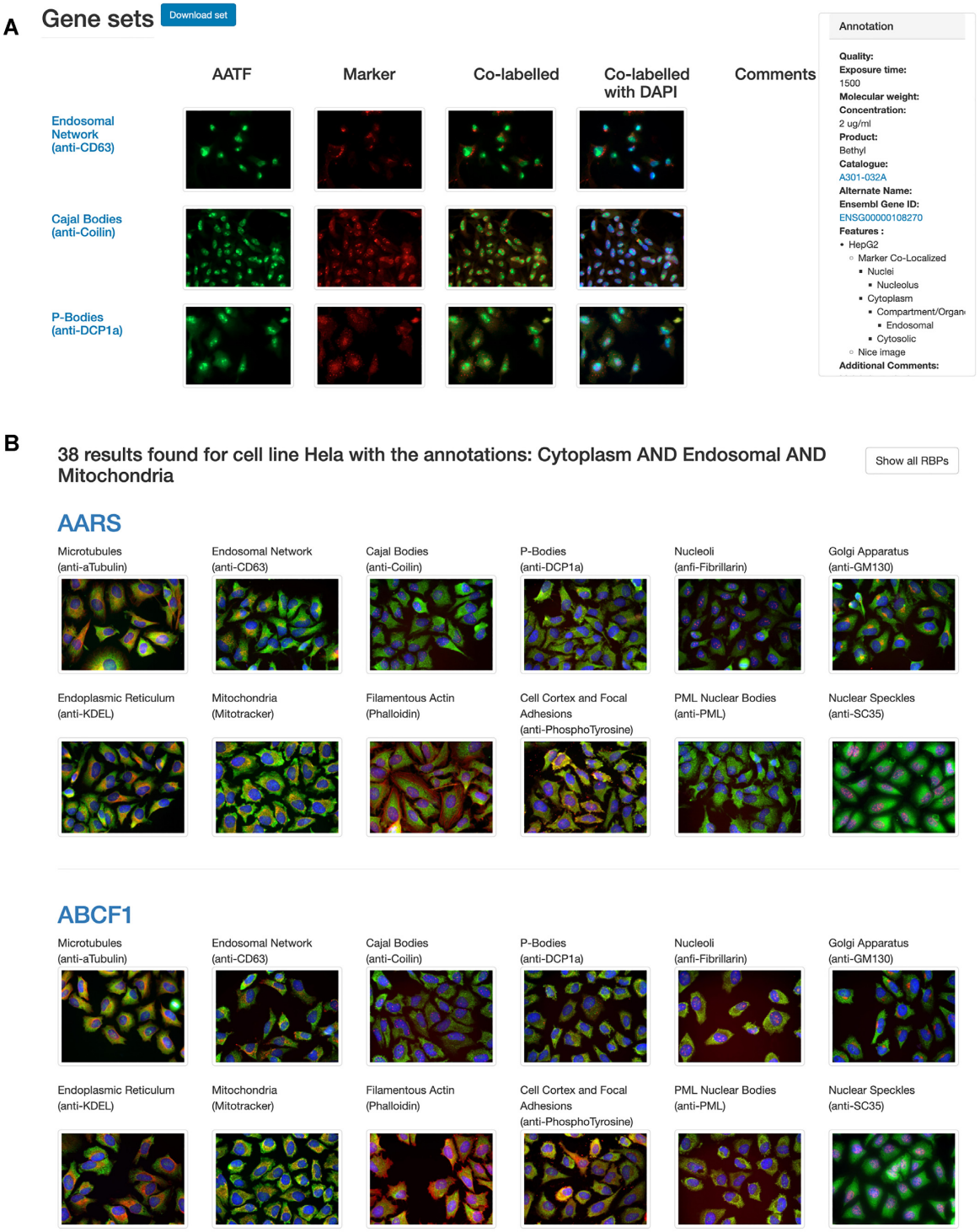

Figure 2. Representation of RBP Image Database data visualization functionalities. Shown are the visualization formats obtained when conducting (A) gene or (B) annotation centric data searches within the RBP Image Database.

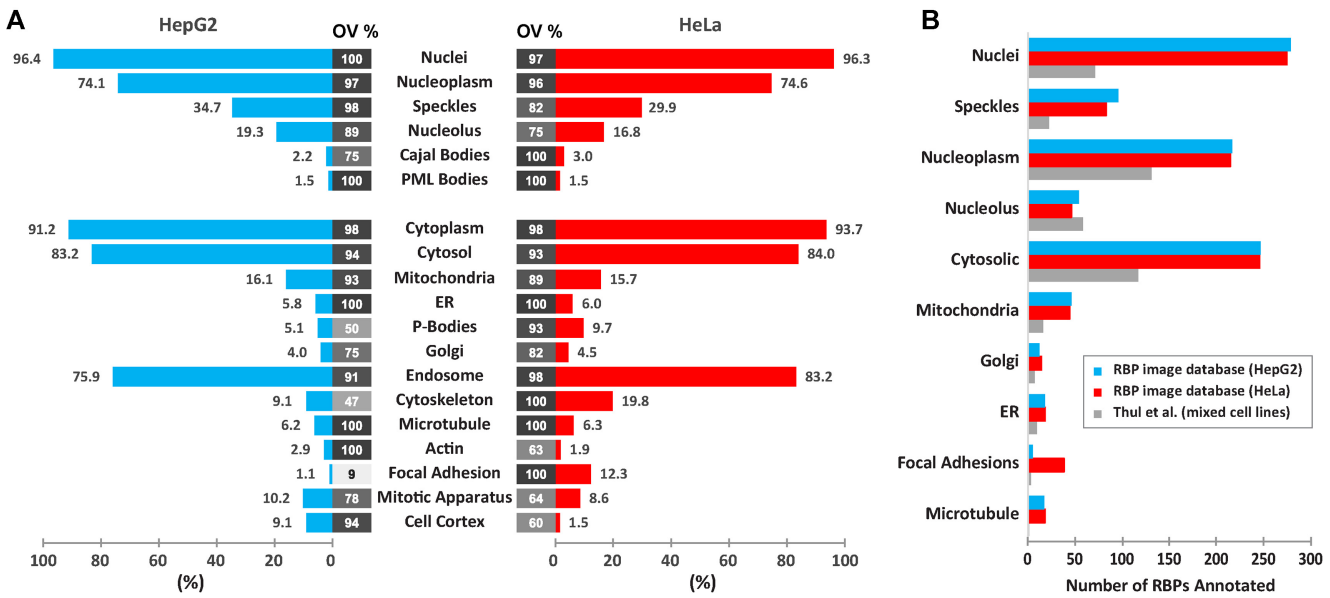

Figure 3. Overall features of RBP localization properties. (A) The percentage of interrogated RBPs (i.e. 299 in HepG2 cells, 294 RBPs in HeLa) localizing to specific subcellular structures in HepG2 (left panel, blue histogram) or HeLa (right panel, red histogram) cells. The percentage of overlap (OV%) values indicate the degree of overlap in lists of RBPs targeted to a given structure in the other cell line. (B) Comparison of commonly employed annotation terms used to describe subcellular localization patterns in the RBP Image Database versus the Human Protein Atlas, as described by Thul et al., 2017 (30), for a set of 269 RBPs commonly analyzed in these datasets.

similarity in RBP repertoires (Figure 4A). This reveals, for example, that cytoskeleton-associated patterns (e.g. microtubules, actin, focal adhesions, centrosomes, cell cortex) tend to be more robustly correlated, as do cytosolic and endosomal RBP subsets, whereas other localization classes exhibit anti-correlated RBP repertoires, including nuclear speckles vs nucleoli or endoplasmic reticulum, or mitochondria vs cytosol. Strikingly, these correlation/anti-correlation signatures are broadly similar between HepG2 and HeLa cells (Figure 4A). We also note that the majority of interrogated RBPs are not restricted to a single compartment, but localize to multiple subcellular destination, with most of these factors having both nuclear and cytoplasmic localization features (Figure 4B). To correlate RBP subcellular localization properties to their functional properties, we next took advantage of recent literature-based functional annotations (21), which we found to provide a deeper description of potential functions in RNA regulation compared to existing Gene Ontology terms. As detailed in (Figure 4C), this analysis revealed several positive and negative correlations in functional attributes of RBP subgroups with specific subcellular localizations. For example, nuclear speckles are enriched for RBPs implicated in mRNA splicing regulation and depleted for factors involved in rRNA processing, while nucleolus-localized RBPs showed the opposite functional profiles. Mitochondria-targeted RBPs showed an enrichment for mitochondrial RNA regulation and a depletion for functions in mRNA splicing regulation and nuclear export. Overall, these analyses suggest that RBPs have complex subcellular distribution features that are generally correlated with their properties in RNA regulation.

Finally, comparative analyses of RBP subcellular distribution features to publicly available enhanced CLIP-seq (eCLIP-seq) datasets (21), which were generated using the same validated antibody reagents, reveals that RBPs with particular cytotopic distribution properties exhibit selective binding features towards particular RNA biotypes (Figure 5). For example, nuclear speckle localized RBPs exhibit prominent binding to pre-mRNA intronic regions and mRNA untranslated regions (UTRs), consistent with the enrichment of these structures for snRNPs and splicing regulatory factors, whereas nucleolar RBPs show depleted binding of these RNA biotypes. Nucleoplasmic RBPs display striking association with different classes of transposable element derived RNAs and micro RNA (MIR). Mitochondria-localized RBPs display enriched binding to mitochondrial chromosome (Chm.M) encoded RNAs, whereas nucleolus-targeted factors show enriched binding to certain snoRNAs (e.g. U14) and certain Vault RNAs (e.g. VTRNA1, VTRNA3). These various examples illustrate how current and future data assembled within the RBP Image Database are likely to be highly impactful on forthcoming studies focused on various aspects of post-transcriptional gene regulation within subcellular space.

CONCLUSION AND FUTURE DIRECTIONS

In this study, we present the RBP Image Database as a user-friendly resource, composed of systematic microscopy imaging data for 301 human RBPs, across 12 subcellular makers and in 2 human cell lines (HepG2 and HeLa), as well as expert curated controlled vocabulary annotations, detailing the subcellular localization features of RBPs. Overall, the data assembled within the RBP Image Database reveals that RBPs display a broad repertoire of subcellular distribution patterns, while also highlighting specific subgroups

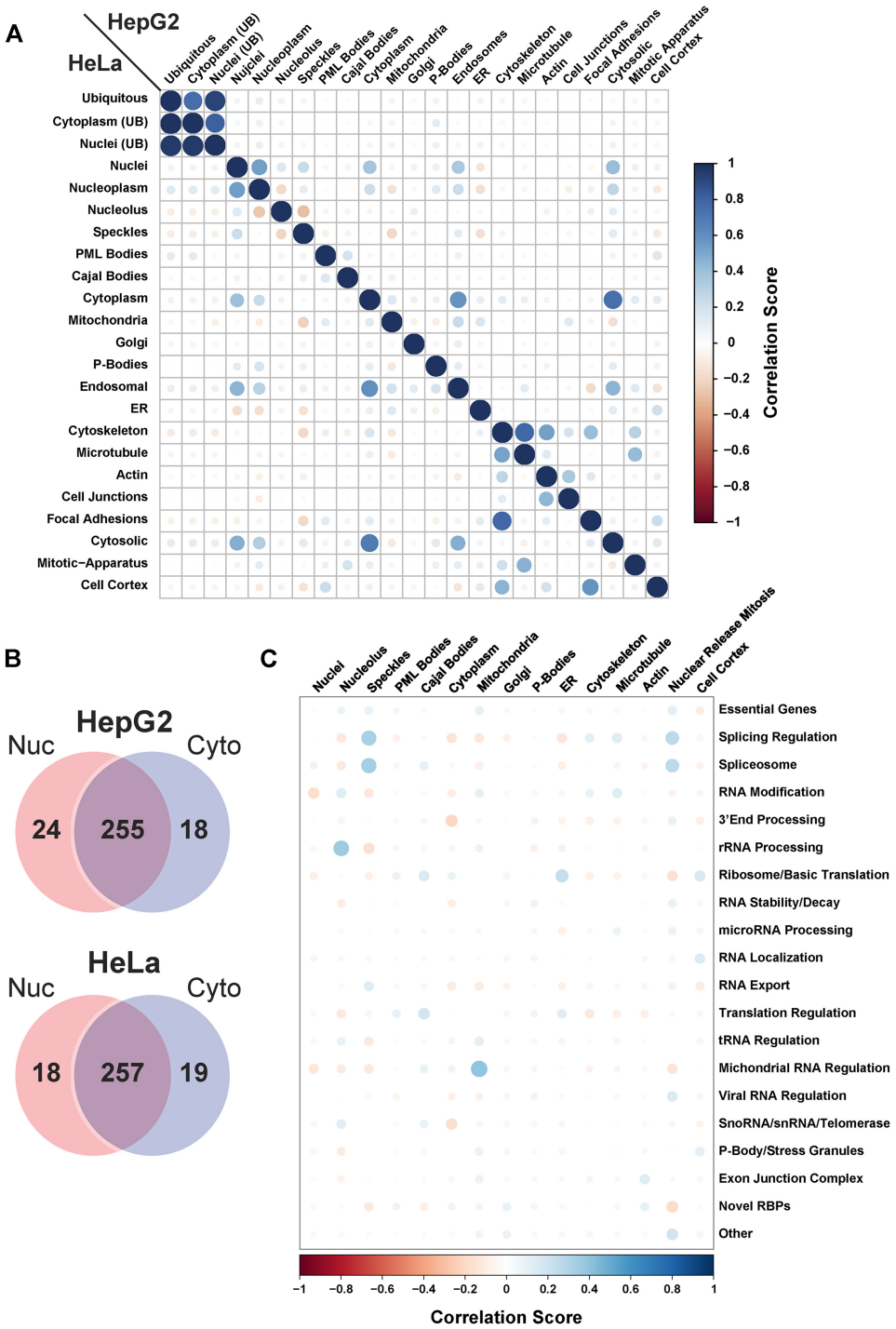

Figure 4. Comparative analysis of the RBP subgroups localized to different subcellular structures. (A) Correlation plot showing all Pearson correlations between the collections of RBPs that localize to specific intracellular sites for HepG2 (upper triangle) and HeLa (lower triangle) cells, as defined by the controlled vocabulary annotations. Note the mirror image correlation scores exhibited for localized RBP repertoires in both cell types. (B) Numbers of RBPs exhibiting nuclear and/or cytoplasmic localization in HepG2 and HeLa cells. (C) Correlation plot showing the relationship of localized RBP subsets from HepG2 cells to functional signatures ascribed through expert literature curation.

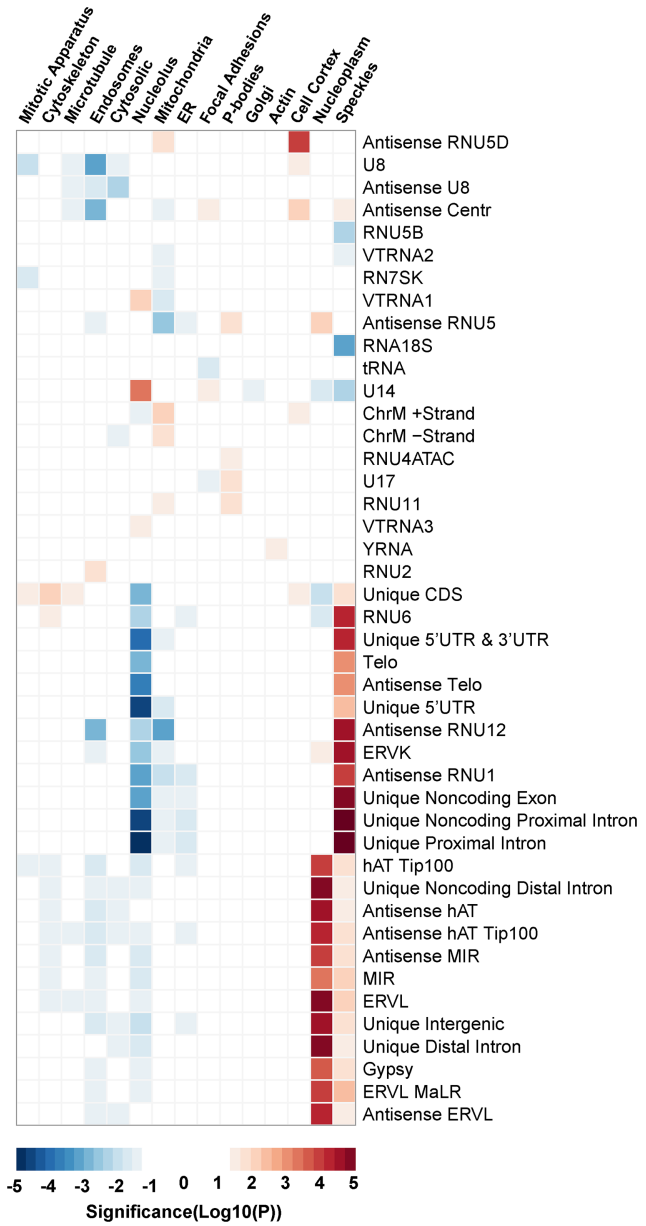

Figure 5. Transcriptome binding properties of RBPs with different localization patterns. Heatmap shows the significance of the relative information (RI) within enhanced CLIP (eCLIP) datasets for groups of localized RBPs, in relation to different RNA biotypes. For each RNA biotype, the RI distribution of RBPs observed at a particular localization was compared against the RI distribution of unlocalized RBPs by a one-sided Wilcoxon rank-sum test. The Log10(p-values) of statistically significant ($p < 0.05$) localized RBPs classes were adjusted to reveal greater (red) or lesser (blue) medians compared to unlocalized RBPs.

of factors that share functionally relevant localization features. The database offers important information to help elucidate the functions of RBPs in RNA regulation within specific subcellular compartments. As such, this resource should prove beneficial for users aiming to address hypotheses and to design experiments to study post-transcriptional regulation. In the coming years, we expect to integrate additional imaging datasets that will expand data coverage across different cell types/models, additional RBPs, stress conditions and genetic perturbations.

DATA AVAILABILITY

The RBP Image Database is freely available at https://rnabiology.ircm.qc.ca/RBPImage/.

ACKNOWLEDGEMENTS

This research was enabled in part through support provided by Calcul Québec and Compute Canada.

FUNDING

This work was funded in part by a grant from the Fonds de Recherche du Québec-Santé (FRQS) to E.L., who is also an FRQS Senior Scholar, as well as a grant from the National Human Genome Research Institute, contract U41HG009889, to B.R.G. (PI), G.W.Y. (PI), C.B.B. (co-PI) and E.L. (co-PI).

Conflict of interest statement. G.W.Y. is a cofounder, member of the board of directors, equity holder, and paid consultant for Locanabio and Eclipse BioInnovations. The terms of these arrangements have been reviewed and approved by the USCD in accordance with its conflict-of-interest policies. ELVN is co-founder, member of the board of directors, on the scientific advisory board, equity holder, and paid consultant for Eclipse BioInnovations. ELVN's interests have been reviewed and approved by UCSD and the Baylor College of Medicine in accordance with their conflict of interest policies. The authors declare that they have no other competing interests.

REFERENCES

1. Anantharaman,V., Koonin,E.V. and Aravind,L. (2002) Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res.*, **30**, 1427–1464.
2. Gerstberger,S., Hafner,M. and Tuschl,T. (2014) A census of human RNA-binding proteins. *Nat. Rev. Genet.*, **15**, 829–845.
3. Mitchell,S.F. and Parker,R. (2014) Principles and properties of eukaryotic mRNPs. *Mol. Cell*, **54**, 547–558.
4. Lunde,B.M., Moore,C. and Varani,G. (2007) RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.*, **8**, 479–490.
5. Stefl,R., Skrisovska,L. and Allain,F.H. (2005) RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle. *EMBO Rep.*, **6**, 33–38.
6. Auweter,S.D., Oberstrass,F.C. and Allain,F.H. (2006) Sequence-specific binding of single-stranded RNA: is there a code for recognition?*Nucleic Acids Res.*, **34**, 4943–4959.
7. Cook,K.B., Kazan,H., Zuberi,K., Morris,Q. and Hughes,T.R. (2011) RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res.*, **39**, D301–D308.
8. Sundararaman,B., Zhan,L., Blue,S., Stanton,R., Elkins,K., Olson,S., Wei,X., Van Nostrand,E.L., Pratt,G.A., Huelga,S.C. *et al.* (2016) Resources for the comprehensive discovery of functional RNA elements. *Mol. Cell*, **61**, 903–913.
9. Brannan,K.W., Jin,W., Huelga,S.C., Banks,C.A., Gilmore,J.M., Florens,L., Washburn,M.P., Van Nostrand,E.L., Pratt,G.A., Schwinn,M.K. *et al.* (2016) SONAR discovers RNA-Binding proteins from analysis of large-scale protein-protein interactomes. *Mol. Cell*, **64**, 282–293.
10. Baltz,A.G., Munschauer,M., Schwanhausser,B., Vasile,A., Murakawa,Y., Schueler,M., Youngs,N., Penfold-Brown,D., Drew,K., Milek,M. *et al.* (2012) The mRNA-bound proteome and its global

occupancy profile on protein-coding transcripts. *Mol. Cell*, **46**, 674–690.
11. Castello,A., Fischer,B., Eichelbaum,K., Horos,R., Beckmann,B.M., Strein,C., Davey,N.E., Humphreys,D.T., Preiss,T., Steinmetz,L.M. *et al.* (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, **149**, 1393–1406.
12. Queiroz,R.M.L., Smith,T., Villanueva,E., Marti-Solano,M., Monti,M., Pizzinga,M., Mirea,D.M., Ramakrishna,M., Harvey,R.F., Dezi,V. *et al.* (2019) Comprehensive identification of RNA-protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nat. Biotechnol.*, **37**, 169–178.
13. Trendel,J., Schwarzl,T., Horos,R., Prakash,A., Bateman,A., Hentze,M.W. and Krijgsveld,J. (2019) The human RNA-Binding proteome and its dynamics during translational arrest. *Cell*, **176**, 391–403.
14. Urdaneta,E.C., Vieira-Vieira,C.H., Hick,T., Wessels,H.H., Figini,D., Moschall,R., Medenbach,J., Ohler,U., Granneman,S., Selbach,M. *et al.* (2019) Purification of cross-linked RNA-protein complexes by phenol-toluol extraction. *Nat. Commun.*, **10**, 990.
15. Caudron-Herger,M., Rusin,S.F., Adamo,M.E., Seiler,J., Schmid,V.K., Barreau,E., Kettenbach,A.N. and Diederichs,S. (2019) R-DeeP: Proteome-wide and quantitative identification of RNA-Dependent proteins by density gradient ultracentrifugation. *Mol. Cell*, **75**, 184–199.
16. Mallam,A.L., Sae-Lee,W., Schaub,J.M., Tu,F., Battenhouse,A., Jang,Y.J., Kim,J., Wallingford,J.B., Finkelstein,I.J., Marcotte,E.M. *et al.* (2019) Systematic discovery of endogenous human ribonucleoprotein complexes. *Cell Rep.*, **29**, 1351–1368.
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. Mukherjee,N., Wessels,H.H., Lebedeva,S., Sajek,M., Ghanbari,M., Garzia,A., Munteanu,A., Yusuf,D., Farazi,T., Hoell,J.I. *et al.* (2019) Deciphering human ribonucleoprotein regulatory networks. *Nucleic Acids Res.*, **47**, 570–581.
20. 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.
21. 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.
22. Caudron-Herger,M., Jansen,R.E., Wassmer,E. and Diederichs,S. (2021) RBP2GO: a comprehensive pan-species database on RNA-binding proteins, their interactions and functions. *Nucleic Acids Res.*, **49**, D425–D436.
23. Liao,J.Y., Yang,B., Zhang,Y.C., Wang,X.J., Ye,Y., Peng,J.W., Yang,Z.Z., He,J.H., Zhang,Y., Hu,K. *et al.* (2020) EuRBPDB: a comprehensive resource for annotation, functional and oncological investigation of eukaryotic RNA binding proteins (RBPs). *Nucleic Acids Res.*, **48**, D307–D313.
24. 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.
25. Blin,K., Dieterich,C., Wurmus,R., Rajewsky,N., Landthaler,M. and Akalin,A. (2015) DoRiNA 2.0–upgrading the doRiNA database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res.*, **43**, D160–D167.
26. Paz,I., Kosti,I., Ares,M. Jr., Cline,M. and Mandel-Gutfreund,Y. (2014) RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res.*, **42**, W361–W367.
27. Giudice,G., Sanchez-Cabo,F., Torroja,C. and Lara-Pezzi,E. (2016) ATtRACT-a database of RNA-binding proteins and associated motifs. *Database (Oxford)*, **2016**, baw035.
28. Benoit Bouvrette,L.P., Bovaird,S., Blanchette,M. and Lecuyer,E. (2020) oRNAment: a database of putative RNA binding protein target sites in the transcriptomes of model species. *Nucleic Acids Res.*, **48**, D166–D173.
29. Maher,P.A., Pasquale,E.B., Wang,J.Y. and Singer,S.J. (1985) Phosphotyrosine-containing proteins are concentrated in focal adhesions and intercellular junctions in normal cells. *Proc. Natl. Acad. Sci. U. S. A.*, **82**, 6576–6580.
30. Thul,P.J., Akesson,L., Wiking,M., Mahdessian,D., Geladaki,A., Ait Blal,H., Alm,T., Asplund,A., Bjork,L., Breckels,L.M. *et al.* (2017) A subcellular map of the human proteome. *Science*, **356**, eaal3321.