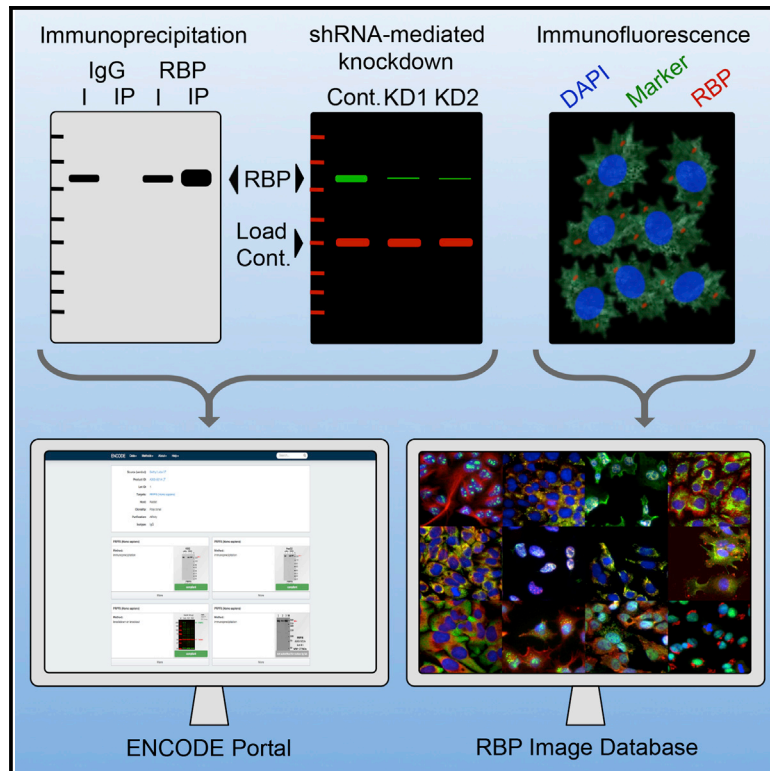# Resources for the Comprehensive Discovery of Functional RNA Elements

Graphical Abstract

Authors

Balaji Sundararaman, Lijun Zhan, Steven M. Blue, ..., Eric Lécuyer, Brenton R. Graveley, Gene W. Yeo

Correspondence

graveley@uchc.edu (B.R.G.),
geneyeo@ucsd.edu (G.W.Y.)

In Brief

Sundararaman et al. present a resource of validated antibodies and short-hairpin RNAs that recognize and target human RNA binding proteins (RBPs). RBPs regulate the life cycle of RNA molecules. This resource will enable a deeper understanding of RBP function.

Highlights

- Antibodies against 365 unique RBPs successfully immunoprecipitate RBPs

- Short-hairpin RNAs against 276 unique RBPs confirm the specificity of RBP antibodies

- Antibodies characterize subcellular localization of RBPs

- Antibody and hairpin RNA information are provided at https://www.encodeproject.org/

CellPress

Resources for the Comprehensive Discovery of Functional RNA Elements

Balaji Sundararaman,[1,5] Lijun Zhan,[2] Steven M. Blue,[1] Rebecca Stanton,[1] Keri Elkins,[1] Sara Olson,[2] Xintao Wei,[2] Eric L. Van Nostrand,[1] Gabriel A. Pratt,[1] Stephanie C. Huelga,[1] Brendan M. Smalec,[2] Xiaofeng Wang,[3] Eurie L. Hong,[4] Jean M. Davidson,[4] Eric Lécuyer,[3] Brenton R. Graveley,[2,*] and Gene W. Yeo[1,*]

[1]Department of Cellular and Molecular Medicine, Institute for Genomic Medicine, Stem Cell Program, Sanford Consortium for Regenerative Medicine, University of California, San Diego, La Jolla, CA 92037, USA
[2]Department of Genetics and Genome Sciences, Institute for Systems Genomics, UConn Health, Farmington, CT 06030, USA
[3]Département de Biochimie, Université de Montréal; Division of Experimental Medicine, McGill University; Institut de Recherches Cliniques de Montréal (IRCM), Montréal, QC H2W 1R7, Canada
[4]ENCODE Data Coordinating Center (DCC), Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA
[5]Present Address: Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA
*Correspondence: graveley@uchc.edu (B.R.G.), geneyeo@ucsd.edu (G.W.Y.)
http://dx.doi.org/10.1016/j.molcel.2016.02.012

SUMMARY

Transcriptome-wide maps of RNA binding protein (RBP)-RNA interactions by immunoprecipitation (IP)-based methods such as RNA IP (RIP) and cross-linking and IP (CLIP) are key starting points for evaluating the molecular roles of the thousands of human RBPs. A significant bottleneck to the application of these methods in diverse cell lines, tissues, and developmental stages is the availability of validated IP-quality antibodies. Using IP followed by immuno-blot assays, we have developed a validated repository of 438 commercially available antibodies that interrogate 365 unique RBPs. In parallel, 362 short-hairpin RNA (shRNA) constructs against 276 unique RBPs were also used to confirm specificity of these antibodies. These antibodies can characterize sub-cellular RBP localization. With the burgeoning interest in the roles of RBPs in cancer, neurobiology, and development, these resources are invaluable to the broad scientific community. Detailed information about these resources is publicly available at the ENCODE portal (https://www.encodeproject.org/).

INTRODUCTION

RNA-binding proteins (RBPs) belong to a diverse class of proteins that are involved in co- and post-transcriptional gene regulation (Glisovic et al., 2008). RBPs interact with RNA to form ribonucleoprotein complexes (RNPs), governing the maturation of their target RNA substrates, such as splicing, editing, cap and 3′ end modifications, localization, turnover, and translation. Dysregulation of and mutations in RBPs are major causes of genetic diseases such as neurological disorders (Kao et al., 2010; King et al., 2012; Lagier-Tourenne et al., 2010; Nussbacher et al., 2015; Paronetto et al., 2007) as well as cancer (Lukong et al.,

2008; Martini et al., 2002; Paronetto et al., 2007) . Traditionally, RBPs were identified by affinity purification of single proteins (Sonenberg et al., 1979a, 1979b). Recent advancements in high-throughput techniques have identified hundreds of proteins that interact with polyadenylated mRNA in human and mouse cell lines (Baltz et al., 2012; Castello et al., 2012; Kwon et al., 2013).

Genome-wide studies that apply methods such as RNA immunoprecipitation (RIP) (Sephton et al., 2011; Zhao et al., 2010) and crosslinking and immunoprecipitation (CLIP) (Hafner et al., 2010; König et al., 2010; Licatalosi et al., 2008; Yeo et al., 2009), followed by high-throughput sequencing have identified hundreds to thousands of protein-RNA interaction sites in the transcriptome for dozens of individual RBPs. These sites or clusters have revealed new rules for how RBPs affect RNA processing and novel pathways for understanding development and disease (Hoell et al., 2011; Modic et al., 2013; Wilbert et al., 2012). The availability of antibodies that specifically recognize the RBP and enable efficient immunoprecipitation of the protein-RNA complex is critical for the successful application of these large-scale techniques in a wide range of tissues and cell types. Alternatively, expression of a fusion protein of one or more peptide tags such as V5, FLAG, or HA in frame with the open reading frame of the RBP is also routinely used (Hafner et al., 2010; Wilbert et al., 2012; Zhao et al., 2010) but has several practical and scientific disadvantages. First, it precludes studying the endogenous proteins in human tissues and currently available animal models of disease. Second, creating cell lines that stably express the tagged RBP is labor intensive and has to be performed for every RBP and cell type under investigation. Third, the tags might interfere severely with protein function or target recognition. Lastly, ectopic expression of tagged RBPs typically uses ubiquitously expressed promoters to drive expression, which might alter the endogenous stoichiometry of the RBP to its binding targets. Overexpression in general may complicate the interpretation of results in an irrelevant cell type.

Given these limitations, characterizing antibodies that can specifically enrich for a given RBP is a laborious yet necessary first step for the systematic evaluation of the endogenous RNA

substrates of RBPs. In this study, we obtained 700 commercially available antibodies that were predicted to recognize 535 candidate RBPs and screened each of them for their ability to efficiently and specifically IP the target RBP. For 51% of the RBPs, we have also identified shRNA reagents that efficiently deplete the target mRNA and protein, simultaneously validating the specificity of the antibodies and providing additional validated experimental reagents. Finally, these antibodies were also used in immunofluorescence assays to determine the subcellular localization of the protein. We expect that this catalog of validated antibodies and shRNA constructs will provide a critical resource for the scientific community.

RESULTS AND DISCUSSION

A Comprehensive Human RNA Binding Protein Reagent Resource

To comprehensively characterize the protein-RNA interactions and functions of all human RBPs, it is essential to develop a resource of validated antibodies and shRNAs for each RBP. Each antibody must be validated to demonstrate that it efficiently and specifically immunoprecipitates the intended target protein. The efficiency of enriching for the target protein is measured by performing immunoprecipitation followed by western blotting, while the specificity of the antibody is measured by performing shRNA knockdown experiments followed by western blotting. These validated antibodies and shRNAs can also be used in a variety of experiments including CLIP-seq to characterize transcriptome-wide protein-RNA interactions, immuno-fluorescence (IF) studies to characterize the subcellular localization patterns of each RBP, and shRNA knockdowns followed by RNA-seq to characterize the function of each RBP in RNA metabolism.

We have compiled a list of candidate human RBPs from the PFAM database (http://pfam.xfam.org/), selecting proteins that contain any of the 86 previously known RNA binding domains (Lunde et al., 2007) (Table S1). This list was further filtered to remove proteins containing RNA-binding domains specific for tRNAs, snoRNAs, and rRNAs, with the remaining proteins containing domains predicted to bind pre-mRNA or mRNA sequences. Additional proteins such as UPF1 and MAGOH that do not contain canonical RNA binding motifs were added to the list based on their previously characterized associations with RNA regulation. Our primary list of 476 RBPs was then merged with the 845 candidate mRNA binding proteins identified in HeLa cells using interactome capture (Castello et al., 2012). Half of the RBPs in our domain-based list overlapped with the interactome-captured RBPs. The union of these two lists yielded a final list of 1,072 candidate RBPs (Table S1), which for the remainder of this manuscript will be referred to as the "RBP compilation." Our primary list of RBPs based on Pfam domains and the putative RBPs identified by polyA captured proteomics study (Castello et al., 2012) are exclusively enriched for mRNA-binding proteins (mRBPs). 817 RBPs in the RBP compilation overlap with a growing list of RBPs known to bind all classes of RNA summarized in Gerstberger et al. (2014) (Figure S1).

To begin building a human RBP resource, we acquired antibodies interrogating the RBP compilation from Bethyl Laboratories (330 antibodies), MBL International (129 antibodies), and GeneTex (245 antibodies), largely consisting of rabbit polyclonal antibodies. Details about the antibodies including catalog number, host species, and antigen information are listed in Table S2. We also acquired 1,139 pre-made shRNA vectors for 491 RBPs from The RNAi Consortium (TRC). The shRNA TRCN ID numbers and the RBP target genes are listed in Table S3. Below we describe our efforts to validate the antibody and shRNA reagents.

Immunoprecipitation-Western Blotting Validation of Antibodies

To date, we have evaluated 700 antibodies, intended to recognize 535 unique RBPs by immunoprecipitation followed by western blotting (IP-WB) validation experiments using K562 whole-cell lysate. These 700 antibodies also include antibodies against 410 RBPs that are common in both the RBP compilation as well as in the RBP census (Figure S1) (Gerstberger et al., 2014). We utilized an IP protocol that contains a series of stringent washing steps similar to that used in the iCLIP method (König et al., 2010; Konig et al., 2011). We devised a scheme to score antibodies for their specificity and IP efficiency as outlined in Figure 1, which are largely based on ENCODE ChIP-seq guidelines (Landt et al., 2012). These scores are based on several criteria including the efficiency of IP, the apparent molecular weight (MW) of the target protein (based on the predicted MWs from Genecards (http://www.genecards.org/), and the number of protein species recognized by the antibody (Figure 1A). The highest quality antibodies are given a score of 1, intermediate-quality antibodies are scored 0.5, and low-quality or unacceptable antibodies are given a score of 0. In addition, if the protein recognized by the antibody is detected in the immunoprecipitation lane and not detected in the input lane due to expression and/or detection level thresholds, the indicator "IP" is appended to the score (e.g., 1IP). Similarly, if only one protein is detected in the input and immunoprecipitation lanes and is enriched upon immunoprecipitation, but the observed MW deviates more than 20% from the expected MW, the identifier "MW" is appended to the score (e.g., 1MW). Finally, if multiple proteins are detected in the input lane and/or are also enriched upon immunoprecipitation, the identifier "MB" is appended to the score (e.g., 1MB).

We identified 284 antibodies (40.6% of tested products) that were characterized with an IP score of "1" (represented by U2AF2 in Figure 1B), indicating that they recognize and enrich only one protein during IP, within 20% of the predicted MW. We identified 12 antibodies (1.7%) that recognize only one protein, but for which the size deviated by more than 20% from the predicted MW, and were therefore scored as "1MW" (KHDRBS2 in Figure 1B). For these antibodies, secondary validations by shRNA depletion of the protein followed by western blotting are necessary to confirm the specificity of the antibody. We identified 41 antibodies (5.9% of tested products) that were scored as "1IP" in K562 cells, indicating that the target proteins were not readily detected in the input (DDX19B in Figure 1B), but were nevertheless enriched upon IP. The secondary validations to assess the specificity of these antibodies by depletion of the target protein cannot be performed in K562 cells as the protein level in untreated cells is below the detection limit of western blot analysis. However, these antibodies can be further validated
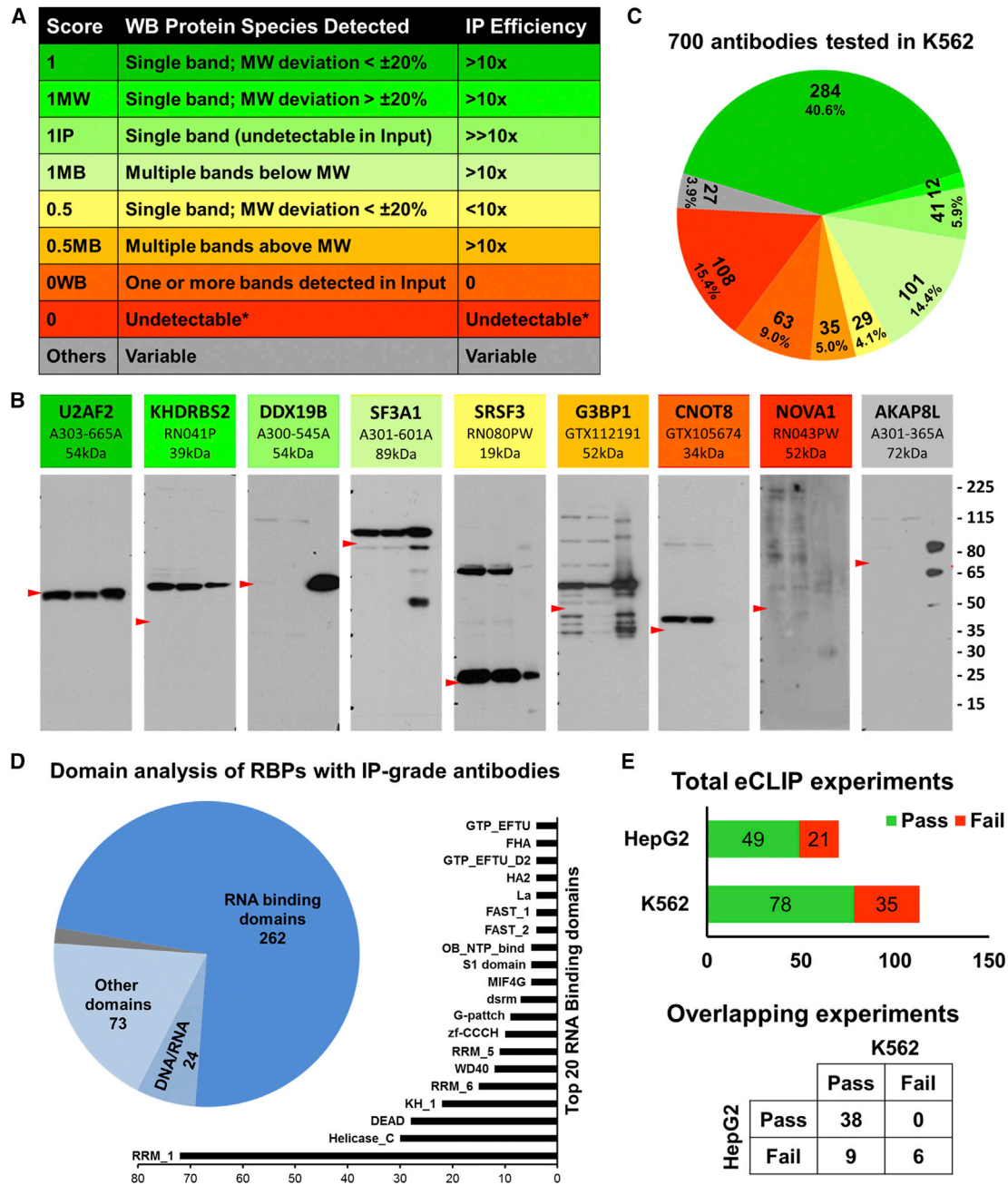
A

Score	WB Protein Species Detected	IP Efficiency
1	Single band; MW deviation < ±20%	>10x
1MW	Single band; MW deviation > ±20%	>10x
1IP	Single band (undetectable in Input)	>>10x
1MB	Multiple bands below MW	>10x
0.5	Single band; MW deviation < ±20%	<10x
0.5MB	Multiple bands above MW	>10x
0WB	One or more bands detected in Input	0
0	Undetectable*	Undetectable*
Others	Variable	Variable

C

700 antibodies tested in K562

284 40.6%
41 5.9%
12
101 14.4%
29 4.1%
35 5.0%
63 9.0%
108 15.4%
27 3.9%

B

U2AF2 A303-665A 54kDa	KHDRBS2 RN041P 39kDa	DDX19B A300-545A 54kDa	SF3A1 A301-601A 89kDa	SRSF3 RN080PW 19kDa	G3BP1 GTX112191 52kDa	CNOT8 GTX105674 34kDa	NOVA1 RN043PW 52kDa	AKAP8L A301-365A 72kDa

Size markers (kDa): 225, 115, 80, 65, 50, 35, 30, 25, 15

D

Domain analysis of RBPs with IP-grade antibodies

RNA binding domains 262
Other domains 73
DNA/RNA 24

Top 20 RNA Binding domains:
GTP_EFTU
FHA
GTP_EFTU_D2
HA2
La
FAST_1
FAST_2
OB_NTP_bind
S1 domain
MIF4G
dsrm
G-pattch
zf-CCCH
RRM_5
WD40
RRM_6
KH_1
DEAD
Helicase_C
RRM_1

(axis: 80 70 60 50 40 30 20 10 0)

E

Total eCLIP experiments

Pass ■ Fail

HepG2: 49 | 21
K562: 78 | 35

(axis: 0 50 100 150)

Overlapping experiments

		K562	
		Pass	Fail
HepG2	Pass	38	0
	Fail	9	6

Figure 1. Immunoprecipitation-Western Blot Validation of Antibodies against RBPs

(A) Scoring scheme with rows of matching colors representing antibodies as in (B). Column 1 is the IP score, column 2 is the description of protein species detected in the WB, and column 3 is IP efficiency deduced from the ratio of band intensities of input and IP pull-down lanes.

(B) Representative blots of antibodies with distinct IP scores. The shades of colors from green to yellow to red represent high-, medium-, and unacceptable-quality of antibodies. Grey antibody represents antibodies with ambiguous IP scores. Each blot contains name of the RBP, catalog number, and expected MW at the top. Within each blot, lane 1 is 2.5% of input of K562 whole-cell lysate, lane 2 is 2.5% of supernatant after IP, and lane 3 is 50% of IP pull-down sample. Red arrowhead in each panel points to the expected MW of the RBP and size marker (in kDa) is at the right.

(C) Distribution of IP scores of 700 antibodies validated in K562 cells. Color pattern is consistent with score description in (A) and representative antibodies in (B).

(D) Domain analysis of 365 unique RBPs that have IP-grade antibodies. Pie chart shows the numbers of RBPs with canonical RNA binding domains, putative DNA/RNA binding domains, and domains with various other functions. The bar chart shows the top 20 RNA Binding Domains represented in the RBPs.

(E) Summary of IP-WB results of eCLIP experiments done to date. Bar chart in the top panel summarizes the total number of antibodies that had either passed or failed the QC step in K562 and HepG2 cells individually. Punnett square at the bottom panel describes the QC results of 53 overlapping eCLIP experiments between K562 and HepG2 cell types.

See also Figure S1 and S2 and Tables S1, S2, and S4.

for specificity in a different cell line that expresses the protein at a detectable level or by analyzing the whole proteome of the immunoprecipitate by mass spectrometry analysis. For example, antibody A300-864A (Bethyl Laboratories), which recognizes RBFOX2, was scored as "1IP" in validations using the K562 cell line that does not express RBFOX2 (Figure S2A), but scored "1" when the validation experiment was performed in the HepG2 cell line, which expresses RBFOX2 at a detectable level (Figure S2B). 101 (14.4%) of the antibodies recognized multiple proteins below the expected MW in the input, supernatant, and/or enriched upon IP and were scored as "1MB" (SF3A1 in Figure 1B). On the other hand, 35 (5.0%) of the antibodies recognized multiple proteins above the predicted MW and were scored as "0.5MB" (G3BP1 in Figure 1B). Antibodies scored as "1MB" may be suitable for CLIP experiments after passing the secondary validation experiments, as the CLIP-seq protocol involves a size selection step for selecting RNP complexes above the MW of the RBP. However, due to non-specific bands above the predicted MW, antibodies with a "0.5MB" score cannot be used for CLIP-seq. Because there is no size selection for the RNP complex after immunoprecipitation in the technique, 1MB and 0.5MB products should not be used in RIP-seq experiments but are nonetheless useful reagents for western blotting.

We scored 29 (4.1%) antibodies as "0.5" (Low Enrichment-Low Priority), for which we observed that the efficiency of immunoprecipitation is such that the intensity of the bands detected in 50% of the immunoprecipitate (Lane 3, SRSF3 in Figure 1B) is less than the band intensity of the 2.5% Input sample (Lane 1). An additional 3.9% of the tested antibodies represented by AKAP8L in Figure 1B, had multiple complicating criteria including multiple bands (MB), MW discrepancy (MW), and detected only upon enrichment (IP) and were designated as "1MBMW," "1IPMW," etc. Due to ambiguity in the specificity of these antibodies, these are considered low priority ("Others" in the Table S4). Finally, there are 171 (24.4%) antibodies that were scored as failing IP validation in K562 cells, because they either do not recognize the correct protein or do not enrich the target protein upon immunoprecipitation. These antibodies were further grouped into two categories. Antibodies that recognize the correct protein in the input lane but failed to enrich the protein in IP are scored as "0WB" (GTX105674 recognizing CNOT8 in Figure 1B) that can only be used for western blotting (63, 9.0%). Additionally there are 108 (15.4%) antibodies that neither recognized nor enriched the correct protein in K562 cells (RN043PW against NOVA1 in Figure 1B). This failure might be due to the absence of the protein recognized by the antibody in K562 cells or that the antibody is simply ineffective for IP and/or WB. The cell type specificity of each antibody could be evaluated in the future by validating the antibody in other cell types that express the RBP.

IP-WB images of all 700 antibodies characterized in K562 cells are available at the ENCODE project portal (https://www.encodeproject.org/) and can be identified using ENCODE Antibody (ENCAB) accession IDs. The results of the IP-WB validations performed to date are summarized in Figure 1C and in Table S4, which includes 438 (62.5%) antibodies against 365 unique RBPs that scored 1, 1MW, 1IP, and 1MB and thus are categorized as "IP-grade" based on our protocol and scoring

criteria in the indicated cell type. The domain diversity of these 365 RBPs was analyzed by searching for Pfam-annotated domains (pfam.xfam.org) associated with these proteins. We identified 322 domains associated with 359 RBPs with 680 total occurrences (see Table S4). 322 Pfam domains were classified into three groups. First, 159 domains are either associated with the direct interaction with RNA or are associated with RNA processing. 268 RBPs that contain at least one of these domains are considered as conventional RBPs (Figure 1D, pie chart). Second, 23 RBPs contain 34 Pfam domains that are either putative DNA/RNA binding domains or bind DNA/chromatin to regulate transcription. Third, another 68 RBPs contain 139 distinct Pfam domains that have no known role in RNA/DNA binding or RNA processing. Six RBPs did not have any annotated Pfam domains. The 20 most frequently occurring domains present in the 268 conventional RBPs are shown in Figure 1D bar chart. RRM_1 domain (PF00076) is the most frequent domain present in 72 RBPs followed by Helicase_C and DEAD domains.

As part of the ENCODE project, we are performing CLIP-seq experiments using our eCLIP method (Van Nostrand et al., 2016). Briefly, cells are subjected to UV-mediated crosslinking, lysis, and treatment with limiting amount of RNAase, followed by IP of protein-RNA complexes using the antibodies described above. RNA fragments protected from RNase digestion by the RBP of interest are then subjected to 3′ RNA linker ligation, reverse-transcription, and 3′ DNA linker ligation to generate eCLIP libraries. The eCLIP method features an optional use of radioactivity, optimizations on enzymatic steps including ligations and, unlike previous methods, also includes a paired size-matched input control which enables the removal of false positive binding sites (Van Nostrand et al., 2016). The IP-WB images that are generated during eCLIP experiments have an additional lane of IP using host-species matched normal IgG antibody as control and can be accessed using the same ENCAB IDs in the portal. We observe that 70% of our *IP-grade* antibodies pass this quality control (QC) step during eCLIP experiments, due to differences in the IP protocol in our initial validation compared to the IP method used by eCLIP. Our initial IP protocol consists of an overnight incubation with the antibody for immunoprecipitation (see Supplemental Experimental Procedures for details). For reasons of standardization and scalability, our CLIP efforts show that immunoprecipitation of the RNP complexes for 2 hr at 4°C is generally sufficient (Van Nostrand et al., 2016). Nevertheless, a fraction of the antibodies prefer an overnight incubation. In summary, of the 113 eCLIP experiments attempted in K562 cells to date, 70% passed the IP-WB QC step (Figure 1E).

As cell-type-specific expression and post-translational modifications add additional layers of complexity that affect the success of immunoprecipitation of RBP-RNA complexes, we tested the utility of IP-grade antibodies validated in K562 cells for eCLIP experiments in HepG2 cells. Of the 53 RBPs that were subjected to eCLIP experiments in both K562 and HepG2 cells, 72% (38) of them passed the QC step in both cell types; nine antibodies failed the IP-WB step in HepG2 cells but passed in K562 cells and another six of them failed in both cell types (Figure 1E). Of the 70 eCLIP experiments performed in HepG2 cells, which

included the 53, 70% (49) passed the QC step, which is comparable to the success rate in K562 cells. Thus, we conclude that the majority of antibodies that are considered *IP-grade* by our criteria in K562 are likely to work in HepG2 cells. If an eCLIP experiment was attempted in HepG2 cell line, the portal will also have IP-WB images from the HepG2 experiments under the same accession ID.

Accessing Information about the Antibodies within the ENCODE Portal

Western blot images of both IP and shRNA knockdown experiments (described in the next section) are uploaded on to ENCODE portal (Figure 2A), which is built and maintained by the Data Coordination Center (DCC) (Sloan et al., 2016). Each unique pair of antibody catalog and lot numbers is given an ENCODE accession number (ENCAB ID). The text search box on the top right of the page of the portal can be used to search for a particular RBP of interest as well as an antibody or shRNA construct using their catalog or TRC numbers, respectively (Figure 2A). Alternatively, the user can browse the entire collection of antibodies by opening the drop-down menu "Data" on the top of the page and then choosing "Antibodies". In the Data/Antibodies page, the results can be filtered based on a number of criteria ranging from "Eligibility status," "Target of antibody," (filters by purported role of target protein), "Characterization method," (filters by method used to validate an antibody), "Source," (filters by antibody manufacturer), and "Lab" (filters by which lab tested the antibody) etc. as shown in Figure S3. The Eligibility status categorizes antibodies as "not pursued," "awaiting lab characterization," "eligible for new data," and "not eligible for new data" based on whether the characterizations met ENCODE standards (Detailed process of how ENCODE reviews antibody characterizations can be found in https://www.encodeproject.org/help/antibody_characterization_process/). Each antibody is rigorously reviewed by the DCC, a subgroup in ENCODE consortium independent of the labs that characterized the antibodies, according to the ENCODE Standards document (https://www.encodeproject.org/about/experiment-guidelines/#antibody) (Landt et al., 2012) and assigned to one of the four eligibility statuses described below.

"Not pursued" indicates that the lab has planned to characterize the antibody and may have done preliminary primary characterizations in one cell type, but it is not interested in doing further characterizations at present. Antibodies under this category are open for further validation either in the same or different cell type. "Awaiting lab characterization" indicates that the lab has completed a primary characterization in one cell type and is expecting to do further validations. "Eligible for new data" indicates that both primary and secondary characterizations are performed in at least one cell type, both of which met the ENCODE standards. Antibodies under this category are eligible for new data generation in the same cell type. A primary validation in the second cell type is necessary before generating new data in that cell type. "Not eligible for new data" indicates that both primary and secondary validations are done in one or more cell types, but they did not meet the ENCODE standards. These antibodies are not eligible for new data generation.

An antibody can be selected by clicking the RBP name listed in the Data/Antibodies page, directing the user to an "antibodies/ENCABnnnxxx" page (Figure 2B) or by entering the RBP name in the search box (Figure S4). The ENCABnnnxxx page contains information regarding species and cell type in which that antibody was characterized, antibody metadata (vendor, host, antigen etc.), and the characterizations. The characterization sub-page is expandable by clicking the "more" option, which lists further information such as characterization methods (immunoprecipitation, knockdown-WB etc), image caption, submitter, and lab names as well as links to download the characterization image and the version controlled standards documents used to set the eligibility status. Each antibody characterization is also classified into "compliant," "not compliant," or "not submitted for review by lab," which indicate that they met, did not meet the standards, or were not reviewed based on the standards document. These characterization statuses determine the eligibility status of the antibody. For example, all antibodies with "eligible for new data" status are expected to have both primary and secondary characterizations with "compliant" status. As additional antibody validation experiments are performed, the results will also be added to the ENCODE project portal under the same accession IDs. The eligibility status of an antibody on the portal will be updated with further characterization. For the antibodies with "eligible for new data" status, the page also will contain links to experiments (eCLIP, for example) in which that antibody might have been used. For more information about data available at the ENCODE Portal, we recommend the help page (https://www.encodeproject.org/help/getting-started/) and Sloan et al. (2016).

Secondary Antibody Validation by Short Hairpin RNA Transduction

To verify the target specificities of these antibodies, we performed secondary validations using shRNA-mediated RNAi. Specifically, to conclude that an antibody recognizes the intended protein, and not a different protein that migrates at the same size range as the intended protein, the band identified in the IP-western blot must be decreased by at least 50% in shRNA knockdown cells compared to cells expressing a control shRNA. To do this, we first identified 1,139 shRNAs from the RNAi Consortium (TRC) that target 491 RBPs (Table S3). To date, we have tested a total of 370 shRNAs against 273 unique RBPs. 274 shRNAs against 242 unique RBPs have been tested in K562 cells and 333 shRNAs against 265 unique RBPs have been tested in HepG2 cells. Of these, 237 shRNAs against 234 RBPs have been tested in both cell lines. We defined a successful knockdown as shRNAs resulting in >50% reduction of the target mRNA or protein, compared to control cells transduced with a non-target control (NTC) shRNA, depending on whether depletion is monitored by qRT-PCR or western blotting. Of the 274 shRNAs tested in K562 cells, 70% passed the validation criteria by RT-qPCR and 60% passed the western blotting validation (Figure 3A). Similarly, in HepG2 cells, 62% and 55% of the target mRNAs and proteins were depleted >50%, respectively, as monitored by qRT-PCR or western blotting (Figure 3B). In most cases, we observed reasonable correlation between the extents of depletion of both the mRNA and protein between both cell lines, though overall, the protein depletion efficiency is an
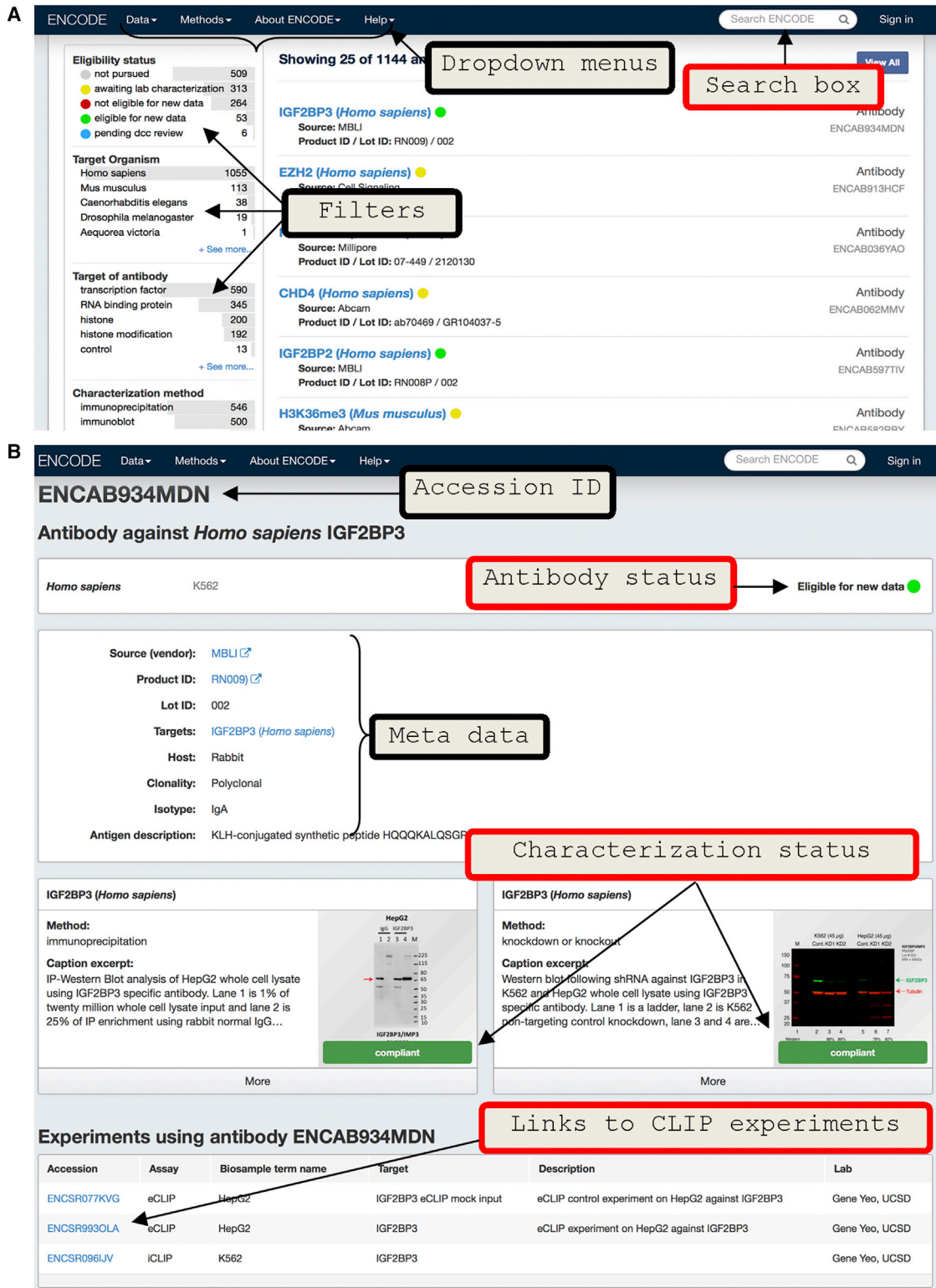
Figure 2. Accessing Antibody Characterizations in ENCODE Portal

(A) Screen shot of the "Antibodies" page of the ENCODE portal. Arrows point to dropdown menus at the top and filtering criteria (to narrow the search) in the left side of the page.
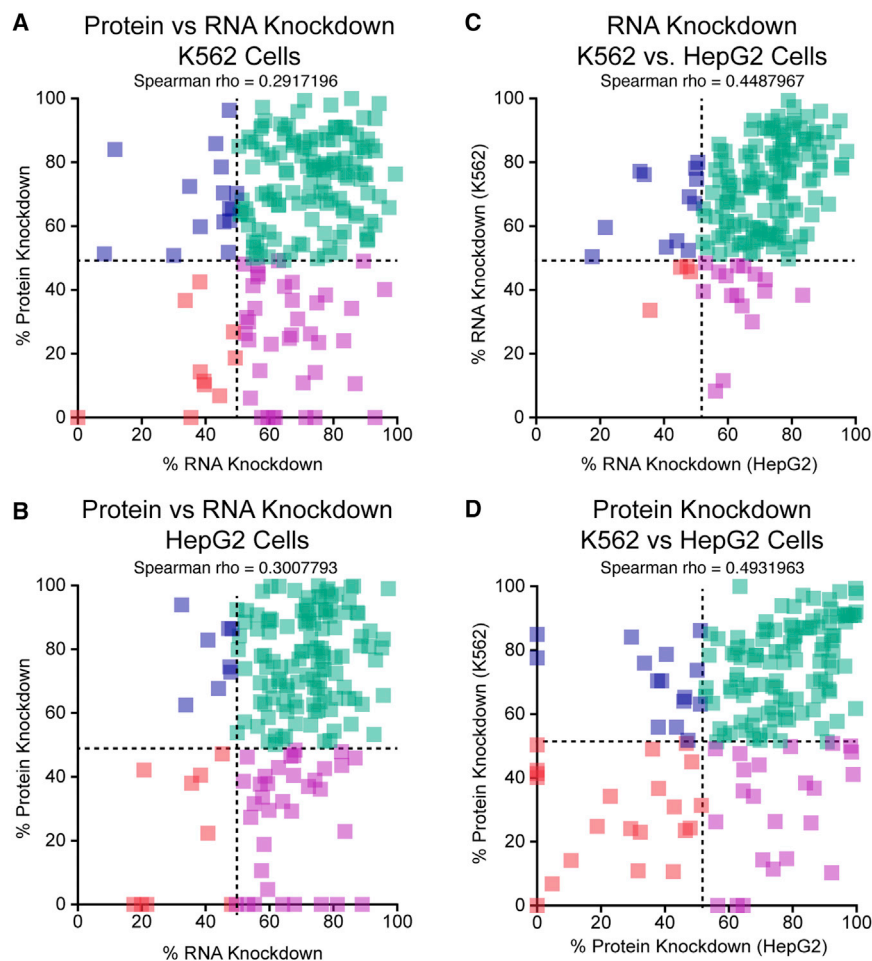
(legend continued on next page)

Figure 3. Comparison of mRNA and Protein Depletion in shRNA Knockdown Experiments

(A) Comparison of protein and RNA knockdown efficiency in K562 cells.

(B) Comparison of protein and RNA knockdown efficiency in HepG2 cells.

(C) Comparison of RNA knockdown efficiency between K562 and HepG2 cells.

(D) Comparison of protein knockdown efficiency between K562 and HepG2 cells. See also Tables S3 and S5.

RBP shRNA knockdown lanes. The antibody that recognizes KHSRP is an example of such a case and recognizes proteins of ∼40 kDa and ∼80 kDa, but only the ∼80 kDa band, near the predicted MW of KHSRP of 73 kDa, is depleted in the KHSRP shRNA samples (Figure 4B). We presume that the difference in the reagents (e.g., secondary antibodies) and detection sensitivity between chemiluminescence used for the IP-WB experiments and fluorescence used for the KD-WB experiments are most likely the source of different banding patterns observed between the experiments. In these cases, the multiple bands detected by fluorescence are likely to be background or cross-reactivity, because only one band was detected when the same sample was analyzed using chemiluminescence. For the four antibodies we

average of 1.25-fold greater in HepG2 cells than in K562 cells (median depletion efficiency of 72% versus 68%) (Figures 3C and 3D). Overall, 68.4% of RBPs tested in both cells were depleted at the protein level by more than 50% in both K562 and HepG2 cells, while 21% of RBPs are depleted >50% in one cell but < 50% in the other cell type (Figure 3D).

Of the 284 antibodies that were scored as "1" during IP validation, 183 were tested by KD-WB in K562 cells and 184 were tested in HepG2 cells. Many of these antibodies, as exemplified by the antibody recognizing PABPC4 (Figure 4A), recognize only a single protein in the control shRNA-treated sample lane and the band intensity was reduced >50% in the RBP shRNA knockdown lanes. 74 of the antibodies that were scored as "1" during IP validation were found to recognize multiple bands in the KD-WB experiments in both the control and RBP shRNA lanes, but only the band of the predicted MW was depleted >50% in the

have tested so far that detected and immunoprecipitated a protein with an aberrant MW in the IP validation experiments (and received a score of 1MW), we also detected a band with the same aberrant MW in the control sample of KD-WB. In each case, the band with the aberrant MW was depleted upon shRNA knockdown, confirming the identity of the protein detected by the antibody. For example, the predicted MW of FASTKD2 is 81 kDa, yet the FASTKD2 antibody recognizes a protein of ∼60 kDa in both K562 and HepG2 cells (Figure 4C). The protein band recognized by FASTKD2 antibody is likely to either be an alternative protein isoform or post-translationally modified form of FASTKD2. For six of the antibodies we have tested by KD-WB that detected multiple bands (1MB) in the IP-WB validation, we also detected multiple bands in the control shRNA sample of KD-WB. For five of these, the band closest to the predicted MW was depleted >50% in the RBP shRNA sample, and the intensity

(B) Screen shot of a representative antibody characterization page from the portal. Top of the page contains the ENCAB accession ID and antibody status. Antibody metadata like catalog number, link to vendor page, lot number, and other information are listed in the top middle panel. Characterizations are in the middle of the page, which can be expanded by clicking "more" option. These characterization subpages lists the submitter lab name, link to download image file and ENCODE standards document, which was used to review the characterization. Links to any experiments that have used the antibody are also listed at the bottom of the page.
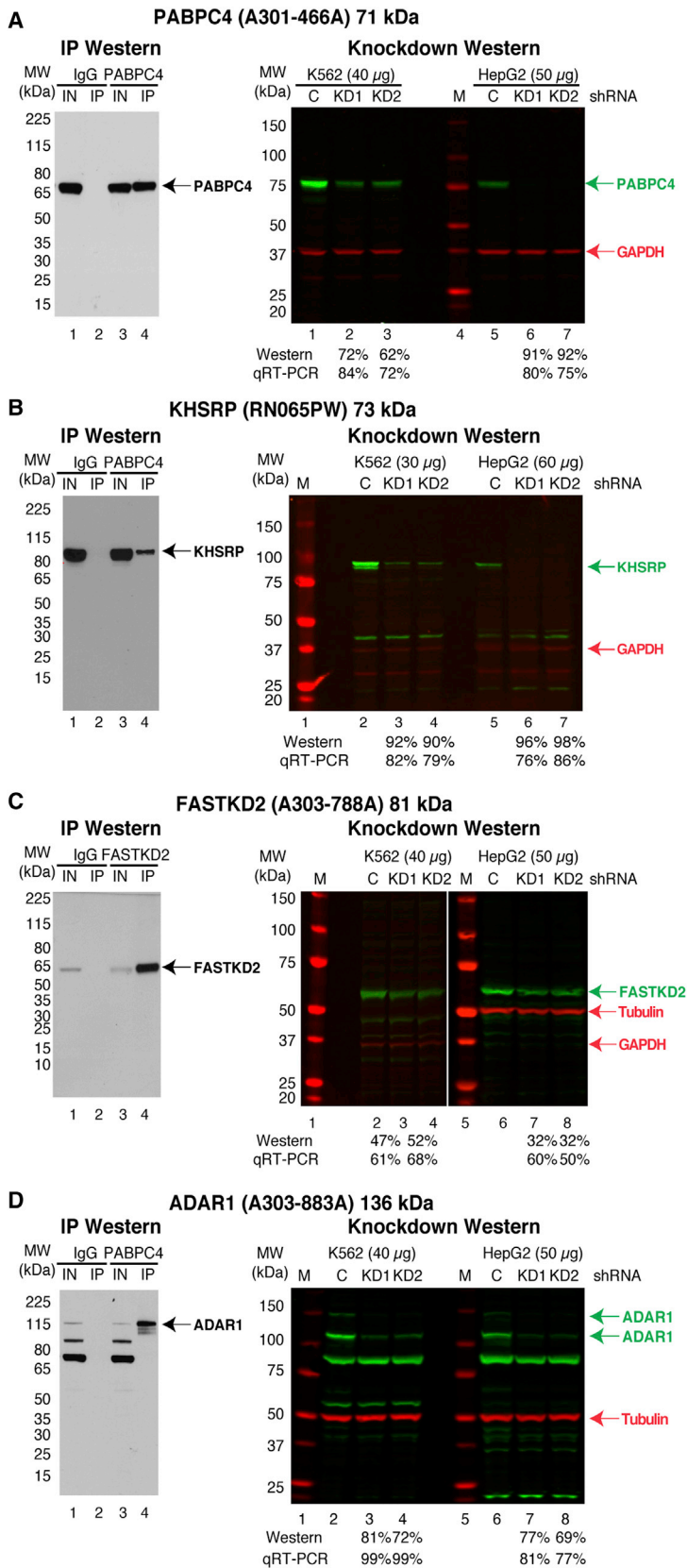
See also Figures S3 and S4.

Figure 4. shRNA Knockdown-Western Blot Validation of Antibodies against RBPs

(A and D) Representative IP-WB (left) and shRNA knockdown-western blot validations (right) for antibodies that cover a spectrum of band patterns. Experiments are shown for antibodies that recognize PABPC4 (A), KHSRP (B), FASTKD2 (C), and ADAR1 (D). For each experiment, the MW markers are shown along with the percent depletion in the knockdown sample compared to the control shRNA sample for both the western blot and qRT-PCR experiments. The position of the RBP (green) and the loading controls of GAPDH or Tubulin (red) are shown. See also Tables S3 and S5.
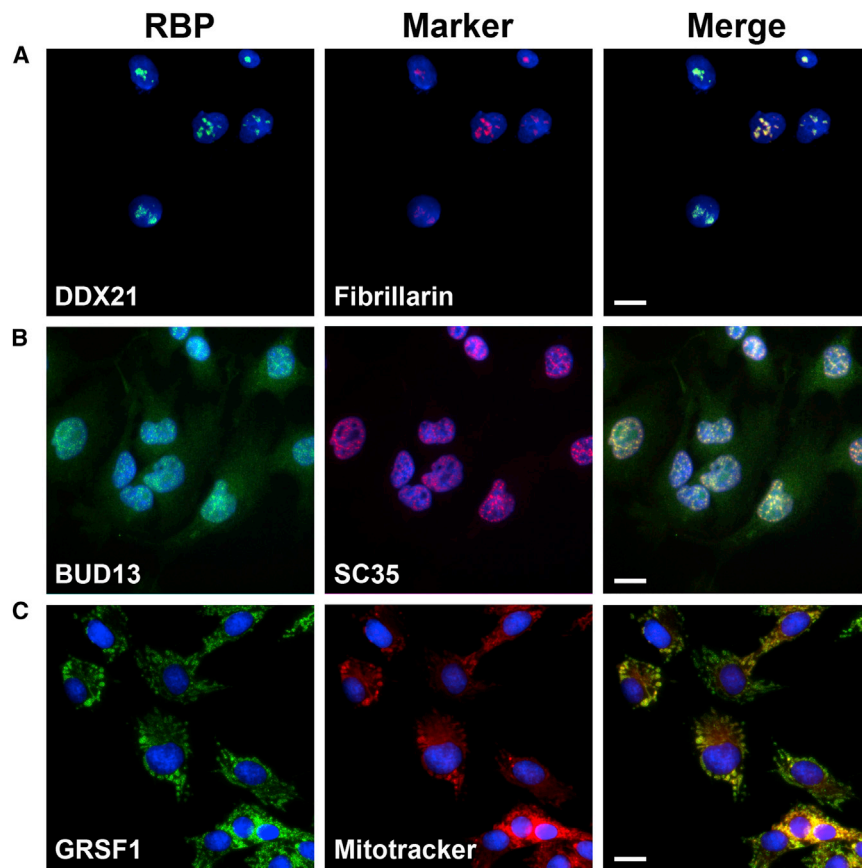
Figure 5. Immunofluorescence Characterization of Antibodies

(A–C) Representative images of immunofluorescence characterizations of antibodies. Left column is the images of RBPs pseudo-colored in green; center column is the sub-cellular markers pseudo-colored in red; and the right column is the merged image of RBP, subcellular marker, and nuclear stain (blue). Scale bar in the merged image represents 20 nm. (A) DDX21 antibody (RN090PW) co-stained with nucleolar marker fibrillarin. (B) BUD13 antibody (A303-320A) co-stained with nuclear speckles marker SC35 and (C) GRSF1 antibody (RN050PW) co-stained with mitochondrial marker mitotracker. See also Table S6.

the RBP compilation antibodies in conjunction with different subcellular markers. We observed clear subcellular distribution patterns for the majority (263 of 274; 96%) of antibodies tested in HepG2 cells. The antibody concentrations employed and staining intensities observed for IF studies are summarized in Table S6. These results are generally consistent with known subcellular localization features of previously characterized RBPs. For example, the DDX21, BUD13, and GRSF1 proteins are respectively localized to nucleoli, nuclear speckles, and mitochondria (Figure 5), consistent with the known functions of these RBPs in rRNA maturation (Calo et al., 2015), splicing control (Zhou et al., 2013), and mitochondrial biogenesis (Antonicka et al., 2013; Jourdain et al., 2013). The full repertoire of results obtained through these systematic imaging studies have been organized within a resource imaging database that will be described in more detail in a separate manuscript (E. Lécuyer, unpublished data).

In conclusion, we have comprehensively validated antibodies and shRNA constructs for hundreds of unique human RBPs. The scoring schema described for the IP-WB validations can be extended to future large-scale antibody characterization studies. The publicly accessible reagent collections serve as key resources for the illumination of functional RNA elements in the human transcriptome.

of most other bands are comparable between the control and RBP shRNA lanes, indicating that the target RBP is one among the multiple bands detected by the antibody. Some cases are additionally complicated. For example, the ADAR antibody scored a 1MB in the IP-WB experiments, recognizes multiple bands in the KD-WB experiments, and more than one band is observed to be depleted in the ADAR1 shRNA samples (Figure 4D). However, ADAR1 is annotated as expressing multiple protein isoforms, and it is likely that the antibody recognizes more than one ADAR1 isoform. The specificity of "1MB" antibodies, which have multiple bands in the IP enrichment that are not depleted via shRNA knockdown, are not considered to be fully validated and therefore should not be used for CLIP experiments. The KD-WB data images can also be publicly accessed through the ENCODE portal and up-to-date report on the status of the RBP antibody validation experiments can be obtained at https://goo.gl/pZqDR5. Table S5 summarizes the results of the RT-PCR and KD-WB experiment for 370 shRNA constructs and also contains the TRC number of the shRNA plasmids, the target sequence of the shRNA, sequences of primers used for RT-qPCR validation, and catalog numbers of antibodies used in the KD-WB characterization.

Immuno-Labeling Studies

As an additional level of validation, and to gain deeper biological insights into RBP function, we have conducted IF studies using

EXPERIMENTAL PROCEDURES

Immunoprecipitation-Western blot validation

Five million human K562 cells were lysed, sonicated (instead of DNase treatment), and the whole-cell lysate used for IP characterizations. Five micrograms of antibody coated on Dynabeads (coupled with either Protein A or anti-Rabbit IgG or anti-mouse IgG) was used to IP overnight at 4°C. Protein-enriched beads were washed twice with a high-salt wash buffer containing 1 M NaCl and detergents to reduce non-specific interactions. For the western blot analysis, aliquots of the input (2.5%), supernatant (2.5%), and bound fractions (50%) ran on 4%–12% SDS-PAGE gel and transferred onto PVDF membrane. Membrane was incubated with 0.2–0.5 μg/ml (see Table S4) of the same

antibody used for IP as the primary antibody. TrueBlot HRP secondary antibodies were used to avoid IgG heavy- and light-chain immunoreactivity. See the Supplemental Experimental Procedures for detailed protocol.

shRNA Knockdown-Western Blot Validation
The shRNA constructs are in pLKO plasmids to facilitate the production of lentiviral particles following co-transfection with appropriate packaging vectors in 293T cells. Lentiviral particles were tittered by qPCR and used to transduce 0.5–0.7 million K562 cells or 0.5 million HepG2 cells in biological duplicate at a MOI of 10. One day after transduction, puromycin was added to the media, and the cells were subjected to selection for 5 to 6 days after which we harvested both RNA and protein (see Supplemental Experimental Procedures for detailed protocol).

Immunofluorescences Imaging Validation
HepG2 cells grown in Poly-L-lysine coated 96-well clear bottom plates were fixed with 3.7% formaldehyde and permeabilized by 0.5% Triton X-100. Cells were then incubated overnight at 4°C with primary antibodies against RBPs (all rabbit antibodies) and marker proteins at 2–10 μg/ml (concentration details provided in Table S6). Cells were washed and incubated with secondary antibodies for 90 min at RT. Imaging was conducted on an ImageXpress Micro high content screening system (Molecular Devices Inc) using a 40× objective (see Supplemental Experimental Procedures for detailed protocol).

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures, six tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2016.02.012.

AUTHOR CONTRIBUTIONS

B.S. and G.W.Y. collected antibodies. E.L.V.N. and S.C.H. collected shRNA constructs. B.S., S.M.B., R.S., and K.E. performed IP characterizations. L.Z., S.O., and B.M.S. performed knockdown characterizations. X. Wang and E.L. performed immunofluorescence characterization. B.S., S.M.B., X. Wei, and B.R.G. curated images and metadata. S.C.H., G.A.P., and G.W.Y. compiled the RBP list; E.L.H. and J.M.D. created and maintain the database. B.S., B.R.G., and G.W.Y. wrote the manuscript. B.R.G. and G.W.Y. directed the project.

REFERENCES

Antonicka, H., Sasarman, F., Nishimura, T., Paupe, V., and Shoubridge, E.A. (2013). The mitochondrial RNA-binding protein GRSF1 localizes to RNA granules and is required for posttranscriptional mitochondrial gene expression. Cell Metab. 17, 386–398.

Baltz, A.G., Munschauer, M., Schwanhäusser, 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.

Calo, E., Flynn, R.A., Martin, L., Spitale, R.C., Chang, H.Y., and Wysocka, J. (2015). RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. Nature 518, 249–253.

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.

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. Nat. Rev. Genet. 15, 829–845.

Glisovic, T., Bachorik, J.L., Yong, J., and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett. 582, 1977–1986.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141, 129–141.

Hoell, J.I., Larsson, E., Runge, S., Nusbaum, J.D., Duggimpudi, S., Farazi, T.A., Hafner, M., Borkhardt, A., Sander, C., and Tuschl, T. (2011). RNA targets of wild-type and mutant FET family proteins. Nat. Struct. Mol. Biol. 18, 1428–1431.

Jourdain, A.A., Koppen, M., Wydro, M., Rodley, C.D., Lightowlers, R.N., Chrzanowska-Lightowlers, Z.M., and Martinou, J.C. (2013). GRSF1 regulates RNA processing in mitochondrial RNA granules. Cell Metab. 17, 399–410.

Kao, D.I., Aldridge, G.M., Weiler, I.J., and Greenough, W.T. (2010). Altered mRNA transport, docking, and protein translation in neurons lacking fragile X mental retardation protein. Proc. Natl. Acad. Sci. USA 107, 15601–15606.

King, O.D., Gitler, A.D., and Shorter, J. (2012). The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. Brain Res. 1462, 61–80.

König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat. Struct. Mol. Biol. 17, 909–915.

Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2011). iCLIP–transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. J. Vis. Exp. 50, e2638.

Kwon, S.C., Yi, H., Eichelbaum, K., Föhr, S., Fischer, B., You, K.T., Castello, A., Krijgsveld, J., Hentze, M.W., and Kim, V.N. (2013). The RNA-binding protein repertoire of embryonic stem cells. Nat. Struct. Mol. Biol. 20, 1122–1130.

Lagier-Tourenne, C., Polymenidou, M., and Cleveland, D.W. (2010). TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. Hum. Mol. Genet. 19 (R1), R46–R64.

Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 22, 1813–1831.

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., et al. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature 456, 464–469.

Lukong, K.E., Chang, K.W., Khandjian, E.W., and Richard, S. (2008). RNA-binding proteins in human genetic disease. Trends Genet. 24, 416–425.

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.

Martini, A., La Starza, R., Janssen, H., Bilhou-Nabera, C., Corveleyn, A., Somers, R., Aventin, A., Foà, R., Hagemeijer, A., Mecucci, C., and Marynen, P. (2002). Recurrent rearrangement of the Ewing's sarcoma gene, EWSR1, or its homologue, TAF15, with the transcription factor CIZ/NMP4 in acute leukemia. Cancer Res. 62, 5408–5412.

Modic, M., Ule, J., and Sibley, C.R. (2013). CLIPing the brain: studies of protein-RNA interactions important for neurodegenerative disorders. Mol. Cell. Neurosci. 56, 429–435.

Nussbacher, J.K., Batra, R., Lagier-Tourenne, C., and Yeo, G.W. (2015). RNA-binding proteins in neurodegeneration: Seq and you shall receive. Trends Neurosci. 38, 226–236.

Paronetto, M.P., Achsel, T., Massiello, A., Chalfant, C.E., and Sette, C. (2007). The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x. J. Cell Biol. 176, 929–939.

Sephton, C.F., Cenik, C., Kucukural, A., Dammer, E.B., Cenik, B., Han, Y., Dewey, C.M., Roth, F.P., Herz, J., Peng, J., et al. (2011). Identification of neuronal RNA targets of TDP-43-containing ribonucleoprotein complexes. J. Biol. Chem. 286, 1204–1215.

Sloan, C.A., Chan, E.T., Davidson, J.M., Malladi, V.S., Strattan, J.S., Hitz, B.C., Gabdank, I., Narayanan, A.K., Ho, M., Lee, B.T., et al. (2016). ENCODE data at the ENCODE portal. Nucleic Acids Res. 44 (D1), D726–D732.

Sonenberg, N., Morgan, M.A., Testa, D., Colonno, R.J., and Shatkin, A.J. (1979a). Interaction of a limited set of proteins with different mRNAs and protection of 5′-caps against pyrophosphatase digestion in initiation complexes. Nucleic Acids Res. 7, 15–29.

Sonenberg, N., Rupprecht, K.M., Hecht, S.M., and Shatkin, A.J. (1979b). Eukaryotic mRNA cap binding protein: purification by affinity chromatography on sepharose-coupled m7GDP. Proc. Natl. Acad. Sci. USA 76, 4345–4349.

Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods. http://dx.doi.org/10.1038/nmeth.3810.

Wilbert, M.L., Huelga, S.C., Kapeli, K., Stark, T.J., Liang, T.Y., Chen, S.X., Yan, B.Y., Nathanson, J.L., Hutt, K.R., Lovci, M.T., et al. (2012). LIN28 binds messenger RNAs at GGAGA motifs and regulates splicing factor abundance. Mol. Cell 48, 195–206.

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.

Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol. Cell 40, 939–953.

Zhou, Y., Chen, C., and Johansson, M.J.O. (2013). The pre-mRNA retention and splicing complex controls tRNA maturation by promoting TAN1 expression. Nucleic Acids Res. 41, 5669–5678.