Characterization of R-Loop–Interacting Proteins in Embryonic Stem Cells Reveals Roles in rRNA Processing and Gene Expression

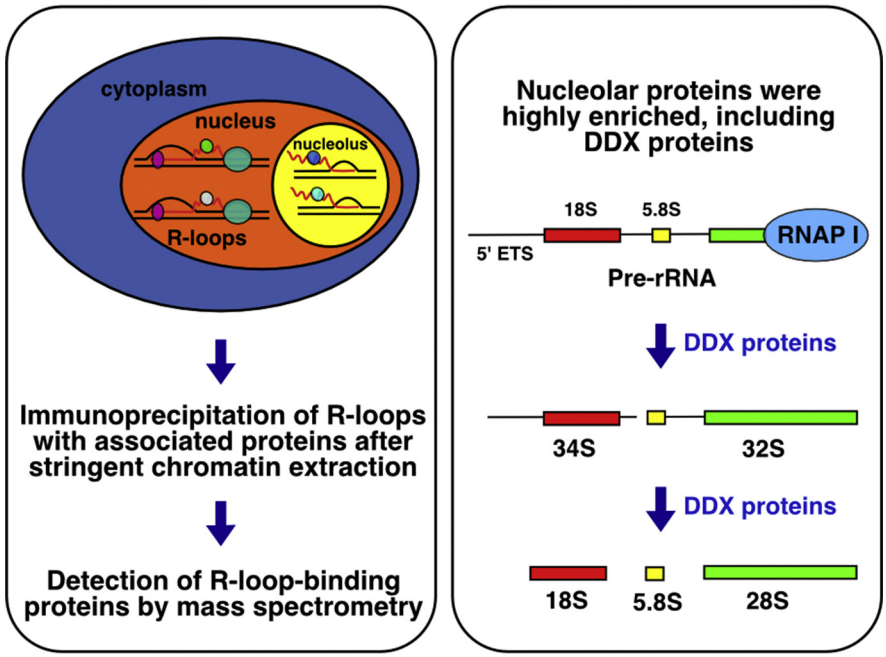
Authors

Tong Wu, Jennifer Nance, Feixia Chu, and Thomas G. Fazzio

Correspondence

thomas.fazzio@umassmed.edu

Graphical Abstract



In Brief

Wu *et al.* performed stringent purification of proteins associated with R-loops in mouse embryonic stem cells, uncovering 364 high-confidence R-loop–associated proteins. Nucleolar proteins, including numerous DEAD-box family helicase proteins, were highly enriched within this group. Closer examination of several DEAD-box helicases revealed post-transcriptional roles in production of mature rRNAs and direct or indirect roles in regulation of differentiation-associated genes. These findings reveal a vast network of R-loop–associated proteins with key functions in stem cell homeostasis.

Highlights

- Three hunderd sixty-four proteins identified as R-loop–interacting proteins.
- Nucleolar proteins were highly enriched, including numerous DEAD-box proteins.
- Several DEAD-box proteins were shown to regulate rRNA processing.
- 4SU labeling/ultraviolet B crosslinking captures weak R-loop–interacting proteins.

2021, Mol Cell Proteomics *20*, 100142

https://doi.org/10.1016/j.mcpro.2021.100142

Characterization of R-Loop–Interacting Proteins in Embryonic Stem Cells Reveals Roles in rRNA Processing and Gene Expression

Tong Wu[1], Jennifer Nance[2], Feixia Chu[2], and Thomas G. Fazzio[1,*]

Chromatin-associated RNAs have diverse roles in the nucleus. However, their mechanisms of action are poorly understood, in part because of the inability to identify proteins that specifically associate with chromatin-bound RNAs. Here, we address this problem for a subset of chromatin-associated RNAs that form R-loops—RNA–DNA hybrid structures that include a displaced strand of ssDNA. R-loops generally form cotranscriptionally and have important roles in regulation of gene expression, immunoglobulin class switching, and other processes. However, unresolved R-loops can lead to DNA damage and chromosome instability. To identify factors that may bind and regulate R-loop accumulation or mediate R-loop–dependent functions, we used a comparative immunoprecipitation/MS approach, with and without RNA–protein crosslinking, to identify a stringent set of R-loop–binding proteins in mouse embryonic stem cells. We identified 364 R-loop–interacting proteins, which were highly enriched for proteins with predicted RNA-binding functions. We characterized several R-loop–interacting proteins of the DEAD-box family of RNA helicases and found that these proteins localize to the nucleolus and, to a lesser degree, the nucleus. Consistent with their localization patterns, we found that these helicases are required for rRNA processing and regulation of gene expression. Surprisingly, depletion of these helicases resulted in misregulation of highly overlapping sets of protein-coding genes, including many genes that function in differentiation and development. We conclude that R-loop–interacting DEAD-box helicases have nonredundant roles that are critical for maintaining the normal embryonic stem cell transcriptome.

R-loops are found throughout the genome in both prokaryotes and eukaryotes (1–5). The majority of R-loops occur when nascent RNAs anneal with the DNA templates from which they are transcribed (6). The unique structure of an R-loop—consisting of an RNA–DNA hybrid (RDH) together with ssDNA—not only affects access of proteins that bind dsDNA but also creates potential sites of recruitment for RNA-, ssDNA-, or RDH-binding proteins.

Although the rules governing R-loop formation and stability are not completely understood, studies have shown that several nucleic acid features favor R-loop formation, including purine-rich (especially G-rich) sequences in transcripts (7), the formation of G-quadruplexes (G4s) (8), negative supercoiling (9), pausing of RNA polymerase (RNAP), and DNA nicks on nontemplate strand (10). For example, upon annealing with the template DNA strand, the G-rich transcripts will form stabilized RDHs because of the higher thermodynamic stability of GC pairs as well as the tendency of G-rich RNA/C-rich DNA of RDHs to exhibit higher stability than equivalent G-rich DNA/C-rich DNA of dsDNA (11, 12). Negative supercoiling, which stabilizes unwound DNA, promotes annealing of nascent RNA to form RDHs (13, 14). In some cases, these features work coordinately: generation of G4s in the G-rich nontemplate ssDNA will induce and stabilize R-loops on the template strand (15), leading to significant colocalization of G4s and R-loops in the genome (16).

Methods of R-loop mapping have been developed to understand the distribution, occurrence, and dynamics of R-loops in the genome. Since the development of DNA–RNA immunoprecipitation (IP) sequencing, most early studies use the RDH-specific monoclonal antibody S9.6 to detect R-loops (17–20), whereas in recent years, alternative methods have been carried out. An alternative approach uses a catalytic dead mutant RNase H1 that binds to but does not digest RDHs to map R-loops (21–23). Coupling of S9.6 mapping with bisulfite sequencing allows for single-nucleotide resolution mapping because of the fact that ssDNA occurs coordinately with RDHs (24). Although these studies all vary somewhat, they all find that a large fraction of R-loops colocalize with nascent transcripts, showing the highest enrichment at the promoter proximal, and to a lesser degree, transcription termination sites and gene bodies (5, 21, 25). In agreement with these findings, R-loops have been shown to have

From the [1]Department of Molecular, Cell, and Cancer Biology, University of Massachusetts Medical School, Worcester, Massachusetts, USA; [2]Department of Molecular, Cellular and Biomedical Sciences, University of New Hampshire, Durham, New Hampshire, USA
*For correspondence: Thomas G. Fazzio, thomas.fazzio@umassmed.edu.

https://doi.org/10.1016/j.mcpro.2021.100142

important regulatory roles in transcription initiation and termination (26–28). Although the roles of R-loops at coding genes have been the focus of many studies (4, 5, 18, 28), R-loops are also prevalent at sites of RNAP I (rRNA) and RNAP III (tRNA) transcription (20, 29). Moreover, R-loops have been shown to regulate epigenetic marks such as DNA methylation and histone modifications near gene promoters (4, 19, 30), as well as immunoglobulin class switch recombination (31, 32), RNA splicing, and DNA replication (33, 34). Despite these important roles, R-loops can promote genomic instability as ssDNA is prone to mutagenesis, and collisions of R-loops with DNA replication forks can cause DNA breaks (35–37). In addition, the accumulation of R-loops beyond normal levels is thought to play roles in amyotrophic lateral sclerosis, Friedreich's ataxia, and fragile X syndrome (38–40). Therefore, R-loop levels must be tightly regulated throughout the genome to maintain genome integrity.

R-loop levels are known to be regulated by several different mechanisms, including relaxation of negative supercoiling by Topoisomerases 1 and 2 (29), resolution of RDHs by helicases such as senataxin (28) and DDX21 (41), and degradation of the RNA portion of RDHs by nucleases, such as RNase H1 (42, 43) and RNase H2 (44, 45). Perhaps as a cellular response to the contributions of R-loops to genomic stress, R-loops often have a relatively short half-life of 10 to 20 min (5). This can be attributed, in part, to a number of helicases and nucleases known to remove or prevent R-loop formation (46).

R-loops have also been shown to regulate local chromatin structure. For example, R-loops regulate the localization of the Tip60–p400 and PRC2 chromatin remodeling complexes (4) as well as the DNA methyltransferase 1 (47), DNA demethylase TET1 (48), and RNA methyltransferase METTL3 (27, 49). Besides these examples, studies in mouse and human cells have identified additional proteins within R-loop–enriched chromatin fractions obtained by IP of RDHs by an RDH-specific monoclonal antibody named S9.6 (17, 50, 51). Although these studies have uncovered several new regulators of R-loop formation or stability, these approaches invariably suffer from a high background of chromatin proteins found within the same chromatin fragment as the R-loop. Consequently, no definitive list of *bona fide* R-loop–interacting proteins has been identified in any cell type.

Several studies have profiled the R-loop interactomes of human or mouse cell lines (50, 51). For studies that focus on endogenous R-loops, an RDH-specific monoclonal antibody named S9.6 has been used to enrich for RDHs from a pool of fragmented chromatin (18). However, because of the fact that R-loop-proximal chromatin—encompassing hundreds of base pairs—is invariably copurified with RDHs, it is difficult to distinguish R-loop–binding proteins from nearby chromatin proteins using these approaches. Although more aggressive DNA fragmentation may reduce the fraction of chromatin proteins copurified with RDHs, RDHs are sensitive to extensive sonication and numerous nonsequence-specific nucleases. Therefore, a different strategy is required to profile R-loop–binding proteins with higher specificity.

Here, we describe two proteomics approaches for more stringent identification of R-loop–associated proteins in mouse embryonic stem cells (mESCs). We identify overlapping sets of R-loop–interacting proteins and many new potential regulators of R-loops. Interestingly, we show that R-loop–binding proteins identified by these approaches are highly enriched in nucleolar proteins, consistent with the high levels of R-loops found within this compartment (29, 52, 53). We find that several R-loop–binding helicase proteins appear to have highly overlapping roles in processing of rRNA, as well as expression of coding genes, suggesting they function in a common pathway. Finally, we show that RNA–protein crosslinking traps a set of proteins that are lost in the absence of crosslinking because of transient or weak association with the RDHs. These studies provide a resource of stringent R-loop–associating proteins in mESCs, including multiple new potential regulators of R-loop formation or stability. In addition, our studies reveal that introduction of selective RNA–protein crosslinking can identify R-loop–binding proteins that are missed by standard approaches.

EXPERIMENTAL PROCEDURES

Experimental Design and Statistical Rationale

For all IP/LC–MS/MS, we performed three biological replicates, with p values calculated by fitting the data to a linear model using ProStar (ProStar Software, Inc) (54) and using a moderated t test, which is robust for large datasets, including proteomics (55). For IP/Western blotting, we performed two or three biological replicates, using mouse IgG2a (same isotype as the S9.6 antibody) as negative controls. For comparison of two proteomics datasets, p values were calculated by hypergeometric tests, which are appropriate for testing the overlaps of moderately sized lists of proteins or genes. For Gene Ontology (GO) term analysis, the PANTHER Classification System was used, with Bonferroni-corrected Fisher's exact tests used to calculate significance. The five GO pathways with the highest observed/expected ratio were shown. For RT-quantitative PCR (qPCR), we typically performed three biological replicates, each with three technical replicates, except where indicated, and Student's t tests were used to test for statistical significance, assuming a normal distribution of RT-qPCR measurements. Three biological replicates were performed for RNA-Seq. Quantitative Western blotting was performed in two biological replicates with both trials shown. Nonquantitative immunofluorescence experiments were performed 2 to 3 independent times per antibody, and representative results are shown.

Cell Culture

E14 mESCs were maintained in tissue culture plates coated with 0.2% gelatin, in medium that contained Dulbecco's modified Eagle's medium–high glucose (MilliporeSigma; D6546-500ML), supplemented with 10% fetal bovine serum (MilliporeSigma; F2442-500ML), 2 mM L-glutamine (Corning; 25-005-CI), minimum essential medium nonessential amino acids (Corning; 25-025-CI), β-mercaptoethanol (MilliporeSigma; M6250-500ML), and recombinant leukemia inhibitory factor.

Antibodies

Antibodies used included DDX18 (Bethyl Laboratories; A300-535A), DDX24 (Abcam; ab70463), DDX27 (Bethyl Laboratories; A302-216A), DDX54 (MilliporeSigma; AV36498-100UL), DHX9 (Abcam; ab26271), RPB1 (Santa Cruz Biotechnology; sc-899x), fibrillarin (FBRL; Novus Biologicals; NB300-269), CEBPZ (Proteintech; 25612-1-AP), SFPQ (Abcam; ab38148), CTCF (MilliporeSigma; 07-729), and mouse IgG2a (Abcam; ab18413). The S9.6 monoclonal antibody was purified from the HB-8730 hybridoma, obtained from American Type Culture Collection.

S9.6 co-IP

For uncrosslinked IP, mESCs were resuspended in 10 mM Tris–HCl, pH 7.0, 150 mM NaCl, and 0.15% NP-40 with 1× Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific; 78429), layered onto sucrose buffer (10 mM Tris–HCl, pH 7.0, 150 mM NaCl, 25% sucrose, and protease inhibitor) and centrifuged at 1000*g*. The pellet was resuspended in stringent wash buffer (10 mM Tris–HCl, pH 7.9, 1.5 mM $MgCl_2$, 420 mM NaCl, 25% glycerol, 0.2 mM EDTA, 0.5 mM DTT, and protease inhibitor) and incubated on ice for 30 min. Nuclei were centrifuged at 7000*g* and resuspended in AM-150/0.1% NP-40 buffer (150 mM KCl, 20 mM Tris–HCl, pH 7.9, 5 mM $MgCl_2$, 0.2 mM EDTA, 10% glycerol, 0.1% NP-40, with 1× Halt Protease Inhibitor Cocktail). The salt-extracted nuclei were then sonicated using a Bioruptor (Diagenode) for 15 cycles, 30 s on/30 s off with the intensity set at medium. Based on protein concentration, 1 mg of sonicated chromatin was mixed with 20 μg S9.6 or mouse IgG2a as control and incubated overnight at 4 °C. About 20 μg of sonicated chromatin was set aside untreated as input (2% of IP samples). Before mixing sonicated chromatin with antibody, 0.2 μl of RNase A (Thermo Fisher Scientific; EN0531) was added and treated at 37 °C for 15 min. For DNase I treatment, 10 μl of DNase I (New England Biolabs; M0303L) was added and treated at 37 °C for 2 h. The next day, the mixture was incubated with prewashed Protein G magnetic beads (New England Biolabs; S1430S), washed three times in AM-150/0.1% NP-40 buffer for 5 min each, and eluted in 1× SDS loading buffer (0.2 M Tris–HCl, pH 6.8, 277 mM SDS, 40% glycerol, and 6 mM bromophenol blue) by boiling for 10 min. For crosslinked IP, mESCs were cultured in stable isotope labeling by amino acids in cell culture (SILAC) medium (Thermo Fisher Scientific; A33972) either supplemented with standard lysine and arginine or $^{13}C_6$ $^{15}N_2$ lysine and $^{13}C_6$ $^{15}N_4$ arginine. About 500 μM 4-thiouridine (4SU) (Biosynth Carbosynth; NT06186) was added to heavy isotope-treated cells for 2 h while the light isotope-cultured cells were treated with dimethyl sulfoxide vehicle. Both sets of cells were then UV treated at a wavelength of 312 nm for 1 J/cm^2, cells were lysed and combined after sonication at 1:1 ratio based on protein amount, and immunoprecipitated as aforementioned. After washing, IP samples were subjected to on-beads RNase A treatment before elution as described previously. Both uncrosslinked and crosslinked S9.6 co-IP were performed with three biological replicates.

LC–MS/MS

S9.6 co-IP elution was run on a 10% SDS-PAGE gel, and gel slices were recovered with care to exclude the majority of the IgG heavy and light chains. Gel bands were in-gel digested and analyzed by LC–MS and LC–MS–MS as described previously (56, 57). The digestion mixture was separated on a 75 μm × 25 cm PepMap Rapid Separation Liquid Chromatography column (100 Å; 2 μm) at a flow rate of ~450 nl/min, and the eluant was analyzed by an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific). LC–MS data were acquired in a data-dependent acquisition mode, cycling between an MS scan (*m/z* 315–2000) acquired in the Orbitrap, followed by collision-induced dissociation analysis on the three most intensely multiply charged precursors acquired in the linear ion trap. The centroided peak lists of the collision-induced dissociation spectra were generated using PAVA searched against a database that is consisted of the Swiss-Prot protein database (version 2017.11.01; 16,942/556,006 entries searched for *Mus Musculus*), using Batch-Tag, a program of the University of California San Francisco Protein Prospector software, version 5.9.2. A precursor mass tolerance of 15 ppm and a fragment mass tolerance of 0.5 Da were used for protein database searches (trypsin as enzyme; one miscleavage; carbamidomethyl [C] as constant modification; acetyl [protein N-term], acetyl + oxidation [protein N-term M], Gln->pyro-Glu [protein N-term Q], Met-loss [protein N-term M], Met-loss + acetyl [protein N-term M], oxidation [M] as variable modifications). Protein hits were reported with a Protein Prospector protein score of ≥22, a protein discriminant score of ≥0.0, and a peptide expectation value of ≤0.01 (58). This set of thresholds of protein identification parameters did not return any substantial false-positive protein hits from the randomized half of the concatenated database. Data are available *via* ProteomeXchange (59) with identifier PXD022697. For differential analysis of IP *versus* input, we used the R package ProStar (54) to calculate the normalized fold change of spectra counting and *p* value and to make volcano plots.

For crosslinked co-IP, cells were differentially labeled, combined, immunoprecipitated, and processed for MS as described previously, and fold-change values were normalized based on the SILAC ratio. For comparison of the crosslinked and uncrosslinked S9.6 co-IP, we first included proteins that appeared in either or both the input and immunoprecipitates in the uncrosslinked data (587 proteins), filtered this list to include only those proteins also found in the crosslinked dataset, and compared both datasets after *Z*-score normalization.

Immunofluorescence Staining

mESCs were cultured on gelatinized coverslips and fixed with 4% paraformaldehyde (Electron Microscopy Sciences; 15710) for 10 min. Cells were then permeabilized with 0.5% NP-40 and blocked with 5% normal goat serum (Vector Laboratories; S-1000) and 0.3% Triton X-100 (Amresco; M143-1L). Primary antibodies were diluted in 1% bovine serum albumin (BSA) and 0.3% Triton X-100 in PBS. The dilutions were as follows: DDX18 (1:250), DDX24 (1:50), DDX27 (1:250), FBRL (1:250), SFPQ (1:250), and CTCF (1:500). Secondary antibodies were Alexa Fluor 488 Goat anti-rabbit IgG (Thermo Fisher Scientific; A-11008, 1:250) and Alexa Fluor 594 Goat antimouse IgG (Thermo Fisher Scientific; A-11005, 1:250). DNA was stained by 1 μg/ml 4′,6-diamidino-2-phenylindole, and slides were observed under Nikon Eclipse E400.

S9.6 Dot Blotting

mESCs were resuspended in ES cell lysis buffer (10 mM Tris, pH 7.5, 10 mM EDTA, 10 mM NaCl, and 0.5% sarkosyl) with 160 μg/ml Proteinase K (Bioline; BIO-37084) added freshly, 37 °C overnight. The next day, phenol/chloroform extraction was performed, the supernatant was mixed with NaOAc pH 5.2 to 0.3 M and then in 2.5 volume of cold ethanol, and kept at −80 °C for 30 min. Spin at maximum speed for 30 min at 4 °C, DNA pellet was washed with 70% ethanol and resuspended in 200 μl meter of water. For DNase I treatment, 1.2 μg of genomic DNA was resuspended in 1× DNase I buffer in a total volume of 20 μl, with 2.5 μl of DNase I. For RNase H treatment, the same amount of genomic DNA was resuspended in 1× RNase H buffer with 1 μl of RNase H (New England Biolabs; M0297L). For RNase A treatment, no buffer was supplied with 1 μl of RNase A. All the treatments were done at 37 °C for 2 h, then 500/250/125/62.5 ng of genomic DNA was spotted onto Amersham Hybond-N+ membrane (Cytiva; RPN303B) using the Bio-Dot Apparatus (Bio-Rad; 1706545). To achieve annealed RDHs for dot blotting, 3.5 μM of 45-nt ssRNA

and ssDNA that are complementary to each other were mixed and slowly cooled down (~5 °C/min) from 95 to 25 °C. About 500 ng of annealed RDHs were treated using the same amount of DNase I/RNase H/RNase A as aforementioned. The samples were loaded onto Hybond-N+ membrane at 200/100/50/25 ng, followed by UV treatment at a wavelength of 254 nm for 120 mJ/cm^2. The genomic DNA blots and RDH blots were blocked with 5% milk for 1 h and then incubated with S9.6 monoclonal antibody (1:1000 dilution in PBS with Tween-20 [PBST] with 2.5% BSA) overnight. The next day, blots were washed three times with PBST, incubated with goat antimouse antibody conjugated with horseradish peroxidase (1:10,000 dilution in PBST with 2.5% BSA) for 1 h, and then washed three times. Blots were ready for development.

Northern Blotting

DNA probes were amplified using complementary DNA (cDNA) from mESCs as a template, radiolabeled by [α-32P]dCTP (PerkinElmer; 3000 Ci/mmol; 10 mCi/ml, BLU013H100UC) using Prime-a-Gene Labeling System (Promega; U1100). Unincorporated dCTP was removed by Bio-Spin P-30 columns (Bio-Rad; 7326231). Radioactivity was determined by a scintillation counter, 5×10^6 cpm of labeled probes were used for a Northern blotting assay. About 1 μg of total RNA was heated at 65 °C for 15 min, cooled on ice, and load into gel containing 1% agarose and 6% formaldehyde, ran in 1× Mops buffer for 100 V for 1.5 h. RNA was transferred from gel to Amersham Hybond-N+ membrane using 10× saline-sodium citrate (SSC) buffer (Invitrogen; 15557044) in a TurboBlotter (Cytiva; 10416316) overnight. The next day, blots were crosslinked at a wavelength of 254 nm for 120 mJ/cm^2 and then transferred into hybridization tubes. Blots were prehybridized with 10 ml PerfectHyb Plus hybridization buffer (MilliporeSigma; H7033-50ML) at 68 °C for 10 min. Radiolabeled probes were added and hybridized at 68 °C overnight with rotation. The next day, blots were quickly washed twice with 5 ml 2× SSC-0.1% SDS wash buffer, then washed with 10 ml 2× SSC-0.1% SDS wash buffer, and rotated at room temperature for 10 min. The blots were exposed to X-ray film in a Kodak BioMax cassette (MilliporeSigma; C4729) at −80 °C. Two biological replicates were performed with similar results.

Western Blotting

The following were the dilution for antibodies: DDX18 (1:4000), DDX27 (1:1000), DDX54 (1:500), DHX9 (1:1000), RPB1 (1:1000), CEBPZ (1:2000), and SFPQ (1:1000). Western blotting was quantified by ChemiDoc Touch Imaging System (Bio-Rad).

Endoribonuclease-Prepared siRNA Preparation and Transfection

Endoribonuclease-prepared siRNAs were prepared as described (60). Briefly, cDNA from mESCs was used as a template to amplify cDNA that targets each gene with T7 anchor sequence added. *In vitro* transcription was then performed using T7 RNAP (New England Biolabs; M0251L). RNA was digested by ShortCut RNase III (New England Biolabs; M0245L) to generate a pool of small siRNAs, which was purified using a PureLink RNA Mini Kit (Thermo Fisher Scientific; 12183020). For transfection, 400 ng of endoribonuclease-prepared siRNA was mixed with 0.4 ml of serum-free medium and 4 μl of Lipofectamine 2000 (Thermo Fisher Scientific; 11668-019), and after 15 min of incubation, 2.8×10^5 mESCs were added and the mixture was plated in one well of a gelatinized 6-well plate. Media were replaced ~16 h later, and cells were harvested 48 h after transfection.

RT-qPCR and RNA-Seq

RNA was extracted using an RNA Clean & Concentrator-25 Kit (Zymo Research; 11-353B) with on-column DNase I digestion for 1 h. For RT-qPCR, cDNA was synthesized using purified mouse mammary tumor virus reverse transcriptase. Quantification was performed using primers targeting cDNA, with *Gapdh* used as a loading control (supplemental Table S11). RNA-Seq libraries were prepared by BGI Genomics Company. mRNA was enriched from total RNA by oligo(dT)-attached magnetic beads, followed by fragmentation, first- and second-strand cDNA synthesis, end repair, add A, adapter ligation, and PCR amplification. PCR products were purified with Ampure XP beads (Beckman Coulter; A63881). The PCR products were denatured and circularized by the splint oligo sequence, and the single-strand circle DNAs were formatted as the final library. Libraries were amplified with phi29 to make DNA nanoball, loaded into the patterned nanoarray, sequenced on BGISEQ-500 using 100-base paired-end sequencing. Reference genome mapping, transcript assignment, quantification, and differential analysis were done using RSEM (https://deweylab.github.io/RSEM/) (61). Heat maps were made by Java TreeView (http://jtreeview.sourceforge.net/). GO term analyses were done by PANTHER Classification System, and overlapping of RNA-Seq data was illustrated by UpSet (https://jku-vds-lab.at/tools/upset/) (62).

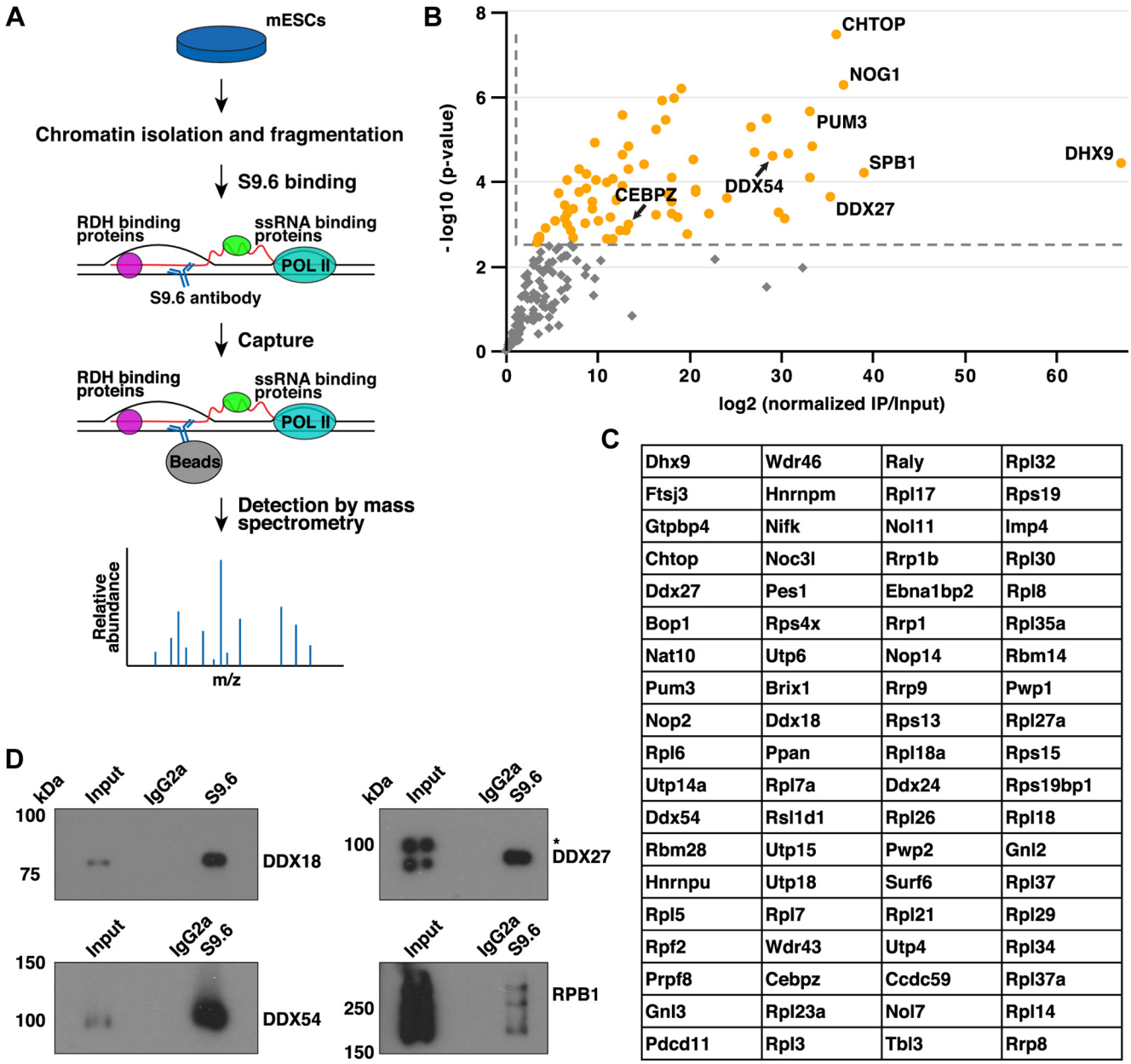
RESULTS

Stringent Capture of R-Loop–Binding Proteins

To identify a stringent set of R-loop–associated proteins, we developed an S9.6 co-IP protocol that first uses high salt washes to remove most chromatin-associated proteins, followed by chromatin fragmentation and IP of RDHs (Fig. 1*A*). To validate this approach, we first tested whether S9.6 can pull down the RNAP II core subunit RPB1, RNA helicase DHX9, and the splicing factor SFPQ. DHX9 has been shown to bind to RDHs and R-loops *in vitro* (63), as well as associate with R-loops *in vivo*, where it regulates their accumulation (50, 64). SFPQ is a splicing factor that was shown to reduce the accumulation of R-loops (64). We observed reproducible enrichment of these proteins (supplemental Fig. S1*A*), validating our approach. In addition, we verified that R-loops were enriched within S9.6 immunoprecipitates by qPCR, observing higher signals at genes known to form R-loops (4) in comparison to genomic regions that lack R-loops (supplemental Fig. S1*B*). As an additional control, we showed that addition of DNase I strongly reduced enrichment of coimmunoprecipitated proteins (supplemental Fig. S1*C*), further validating this approach. Finally, after confirming that RDHs were also highly sensitive to RNase A (supplemental Fig. S1, *E* and *F*), as demonstrated in two previous studies (65, 66), we found that treatment of chromatin with RNase A was able to deplete endogenous R-loops, as expected (supplemental Fig. S1*D*). These studies demonstrate the high specificity with which R-loops and known interacting proteins are enriched by our approach.

Identification of R-Loop–Binding Proteins by MS

Our laboratory previously mapped the genomic locations of R-loops in mESCs and showed that they regulate the recruitment of two chromatin remodeling complexes, making mESCs a good platform for studying R-loop functions (4). To identify the R-loop interactome of mESCs in an unbiased



Dhx9	Wdr46	Raly	Rpl32
Ftsj3	Hnrnpm	Rpl17	Rps19
Gtpbp4	Nifk	Nol11	Imp4
Chtop	Noc3l	Rrp1b	Rpl30
Ddx27	Pes1	Ebna1bp2	Rpl8
Bop1	Rps4x	Rrp1	Rpl35a
Nat10	Utp6	Nop14	Rbm14
Pum3	Brix1	Rrp9	Pwp1
Nop2	Ddx18	Rps13	Rpl27a
Rpl6	Ppan	Rpl18a	Rps15
Utp14a	Rpl7a	Ddx24	Rps19bp1
Ddx54	Rsl1d1	Rpl26	Rpl18
Rbm28	Utp15	Pwp2	Gnl2
Hnrnpu	Utp18	Surf6	Rpl37
Rpl5	Rpl7	Rpl21	Rpl29
Rpf2	Wdr43	Utp4	Rpl34
Prpf8	Cebpz	Ccdc59	Rpl37a
Gnl3	Rpl23a	Nol7	Rpl14
Pdcd11	Rpl3	Tbl3	Rrp8

FIG. 1. **Identification of R-loop–binding proteins.** *A*, schematic diagram of S9.6 co-IP experiment. *B*, volcano plot of the proteins enriched by S9.6 IP. The *vertical dashed line* denotes two-fold normalized enrichment, and the *horizontal dashed line* denotes a *p* value of 0.003 [–log10 (*p*) = 2.5]. Highly enriched proteins of interest were labeled with their protein name. Three biological replicates were included. *C*, gene names of the 76 most enriched proteins by S9.6 co-IP, ordered by normalized IP/input. *D*, Western blots of several DEAD-box family proteins enriched by S9.6 co-IP. mIgG2a is a negative control for IP, and the RNAP II subunit RPB1 is shown for comparison. The *asterisk* denotes a nonspecific band. IP, immunoprecipitation; RNAP, RNA polymerase.

manner, we performed S9.6 co-IP with or without RNase A treatment in three biological replicates. Samples were fractionated by SDS-PAGE (supplemental Fig. S2*A*) and subjected to tryptic digestion after isolating gel slices that excluded the majority of IgG heavy and light chains. Input and IP samples from each treatment were subjected to LC–MS/MS to identify the repertoire of proteins that interact with R-loops. A total of 709 proteins were detected in any of the three biological replicates in input or IP samples, with high correlation among the replicates (supplemental Fig. S2*B*). As expected, very few proteins were identified in RNase A-treated samples, demonstrating high specificity of the IP conditions (supplemental Table S1 and supplemental Fig. S2*C*). After filtering for proteins present in at least two IP replicates and removal of contaminating IgG peptides, 335 proteins were retained for downstream analysis (supplemental Table S2).

Next we used the ProStar software package to identify the most significantly enriched R-loop–interacting proteins (54). We observed 76 proteins that were enriched more than twofold (relative to normalized input) with p value <0.003 (supplemental Table S3 and Fig. 1, *B* and *C*). One of the most highly enriched proteins is DHX9, which was previously shown to interact with RDHs in HeLa cells, further validating our approach (50). Other highly enriched proteins included SPB1 (gene name: *Ftsj3*), a nucleolar protein that regulates pre-rRNA processing (67) and NOG1 (gene name *Gtpbp4*), a nucleolar GTP-binding protein with crucial roles in 60S ribosome biogenesis (68). Another highly enriched protein was CHTOP, a component of the THO–TREX complex previously shown to inhibit the formation of R-loops (69, 70). In addition, numerous known or predicted RNA-binding proteins were identified, including DDX18, DDX21, DDX27, DDX54, which belong to the DEAD-box family of RNA helicases (71–73). Another group of R-loop-interacting proteins, including PUM3 and HNRPU, have both DNA- and RNA-binding activities, suggesting possible roles in R-loop-dependent regulation of chromatin architecture or local epigenomic features. To validate our findings, we performed S9.6 co-IP followed by Western blotting and detected DDX18, DDX27, DDX54, with high IP efficiency (Fig. 1*D*). In all, 49% of proteins (164 of 335) identified in our S9.6 co-IP/MS analysis are known RNA-binding proteins, including several known regulators of R-loop formation, suggesting a high proportion of hits are *bona fide* R-loop-interacting proteins.

R-loops and other non-canonical DNA structures, such as G4s and cruciform DNA, have previously been shown to colocalize throughout the genome (74, 75). We therefore investigated if R-loop-interacting proteins identified in this study are known to bind to and regulate other DNA or RNA structures. To this end, we took advantage of G4IPDB, a database of G4-binding proteins (76). We examined the overlap of 77 known G4-binding proteins from *Homo sapiens* (77) with the 335 R-loop-interacting proteins identified above, where one-to-one mouse to human homologs could be identified. We observed that 17 of the 77 G4-binding proteins overlap with the 335 mESC R-loop binding proteins (supplemental Table S4, observed/expected = 13.2; p value = 8.472×10^{-15}). Among the 17 overlapping proteins is the DDX21 helicase, consistent with *in vitro* studies showing that DDX21 resolves R-loops and G4s (41, 78). Of the 17 known cruciform-binding proteins from *H. sapiens* (79), we found that two—TOP2A and YWHAB—overlapped with our set of 335 R-loop-binding proteins (supplemental Table S5, observed/expected = 7.0; p value = 0.032). Yeast TOP2, a homolog of the mouse TOP2A (DNA topoisomerase 2-alpha), is known to regulate R-loops formed during rRNA synthesis (29). In contrast, the YWAHB protein has no known function in regulation of R-loops. The presence of YWAHB in both datasets could simply reflect the colocalization of cruciform structures and R-loops form at sites with negative supercoiling (14, 79).

We next examined the enrichment of GO categories among the 335 RDH-interacting proteins identified by our approach. As expected, multiple protein classifications associated with RNA processing were highly enriched (supplemental Fig. S3*A*). We observed similar enrichment by analyzing the 76 most prominent R-loop interacting proteins (enriched over two-fold, with p value < 0.003, data not shown). Notably, GO terms associated with rRNA processing were especially prominent, in agreement with the fact that R-loops are frequently observed in nucleoli (52, 53).

To examine the cellular localization of several stringent R-loop–interacting proteins, we performed immunofluorescence staining in mESCs. Consistent with the GO term analysis, DDX18, DDX24, and DDX27 showed substantial overlap with the nucleolar marker FBRL, although nuclear localization outside the nucleolus was also observed, especially for DDX24 and DDX27 (Fig. 2*A*). Conversely, two poorly enriched proteins, CTCF and SFPQ, exhibited largely diffuse nuclear staining (supplemental Fig. S3*B*). This raised the possibility that the strongest hits may be largely nucleolar, whereas hits with lower enrichment were more likely to interact with nuclear R-loops. Consistent with this possibility, R-loop–interacting proteins contributing to the enrichment of nucleolus- or rRNA-related GO terms exhibited stronger overall enrichment in our MS dataset (Fig. 2*B*). Finally, we compared our hits to a dataset describing the nucleolar proteome, which was previously measured by an independent group using isolated nucleoli from mouse fibroblasts followed by protein extraction and MS quantification (80). Of the 320 nucleolar proteins previously identified, we found that 122 overlapped with the 335 stringent R-loop–interacting proteins identified in our study (supplemental Table S6 and Fig. 2*C*, O/E: observed/expected = 22.8; $p = 3.957 \times 10^{-138}$). Furthermore, of the 76 most prominently enriched R-loop–associated proteins (described previously), 39 overlapped with the nucleolar proteome (supplemental Table S7 and Fig. 2*C*, O/E = 32.1; $p = 3.367 \times 10^{-50}$). These data indicate that many of the most strongly enriched R-loop–interacting proteins function largely within the nucleolus, likely because of the high abundance of cotranscriptional R-loops during rRNA synthesis. Accordingly, proteins with important roles in the regulation or functions of R-loops at protein coding genes may be enriched to lower levels in the LC–MS/MS dataset.

Shared Functions of R-Loop–Associated DEAD-Box Proteins

The DEAD-box protein family, named after the conserved D-E-A-D amino acid sequence within the Walker B motif, consists of known or putative ATP-dependent RNA helicases that are conserved throughout eukaryotes (71–73). DEAD-box proteins have been found to contribute to RNA metabolism *in vivo*, including processes such as mRNA transcription and degradation, splicing, mRNA export, and ribosome biogenesis (71).

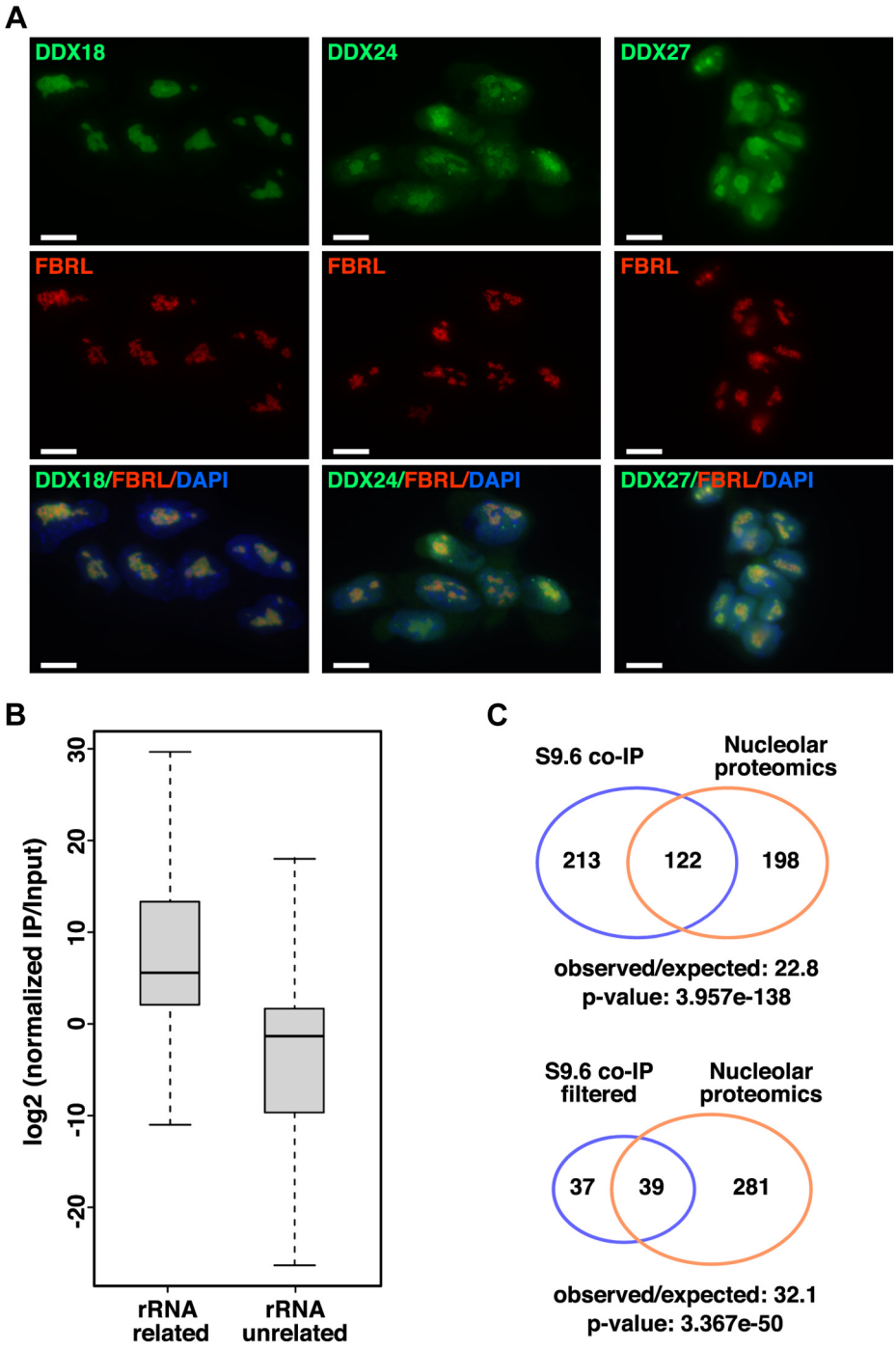


FIG. 2. **Many R-loop–binding proteins are enriched within the nucleolus.** *A*, immunofluorescence staining of three R-loop–associated DEAD-box proteins, DDX18, DDX24, and DDX27, costained with FBRL to mark the nucleolus. DNA was stained by DAPI. The scale bar represents 10 μm. *B*, comparison of rRNA-related and rRNA-unrelated proteins enriched by S9.6. R-loop–associated proteins were classified as rRNA related or unrelated by GO term analysis, using the biological process term "ribosomebiogensis," "rRNAmetabolicprocess," and "rRNAprocessing." *Boxes* of the box plot represent the first and third quartiles, the band represents the median, and the whiskers depict 1.5 times the interquartile range. *C*, Venn diagram of the overlap between the R-loop interactomes (with and without filtering) and the nucleolar proteome from Kar *et al.* (80). The *upper diagram* includes the 335 proteins enriched in S9.6 co-IP samples with the 76 most enriched (corresponding to the *orange dots* in Fig. 1*B*) depicted in the *lower diagram*. *p* Values were calculated using hypergeometric tests. DAPI, 4′,6-diamidino-2-phenylindole; FBRL, fibrillarin; GO, Gene Ontology; IP, immunoprecipitation.

Of the 335 stringent R-loop–associated proteins we identified from our MS, 13 belong to the DEAD-box protein family, including known R-loop regulators such as DDX5 (81) and DDX21 (41). To better understand the functions of this protein family in the regulation of R-loops, we used RNA interference to examine the cellular consequences of partial depletion of R-loop–associated proteins DDX10, DDX24, DDX27, and DDX54. Several DEAD-box family members were previously shown to localize to the nucleolus and/or function in ribosome biogenesis (82–86). For example, DDX5 has been shown to promote rRNA transcription (87), whereas DDX51 and DDX54 were shown to affect different steps of rRNA processing (88, 89). We therefore examined the roles of DDX10, DDX24, DDX27, and DDX54 in the production and processing of rRNAs. To this end, we performed Northern blotting on total RNA isolated from mESCs, using probes specific to the 18S or 28S rRNAs. In addition to fully processed rRNAs, these probes also hybridize to the 45S pre-rRNA and multiple smaller rRNA processing intermediates (90–92), as outlined in Figure 3*A*. The 18S rRNA Northern probe, which also detects the 45S, 41S, and 34S pre-rRNAs, uncovered alterations in levels of 45S and 34S pre-rRNAs in *Ddx24* and *Ddx10* knockdown (KD), respectively (Fig. 3*B*). In contrast, the 28S probe revealed increased 36S pre-rRNA in *Ddx24* KD and *Ddx27* KD mESCs (Fig. 3*C*). In addition, *Ddx54* KD cells exhibited increased 32S pre-rRNA relative to *enhanced GFP* KD control cells (Fig. 3*C*).

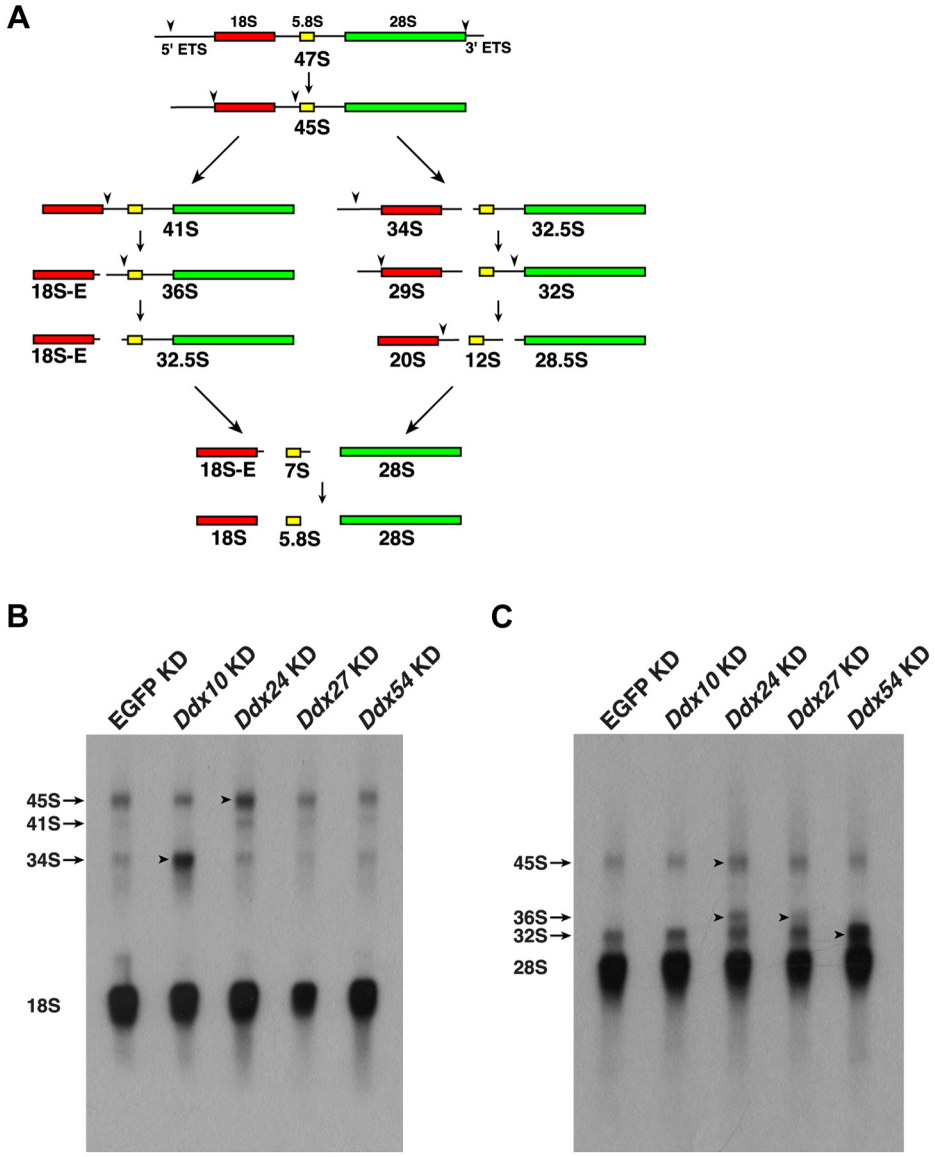


FIG. 3. **Regulation of rRNA processing by R-loop–interacting DEAD-box family helicases.** *A*, schematic diagram of rRNA processing in mouse. *B*, relative abundance of 18S rRNA and indicated pre-rRNA transcripts in enhanced GFP and *Ddx* knockdowns (KDs), as shown in a Northern blot using a probe targeting the 18S sequence. *Arrowheads* indicate alterations in pre-mRNA transcripts in some KDs. *C*, Northern blot using a probe targeting the 28S sequence, as depicted in *B*.

Collectively, these data demonstrate that DDX10, DDX24, DDX27, and DDX54 contribute to production of mature 18S and 28S rRNAs at the level of rRNA processing, as previously observed for DEAD-box helicases DDX5, DDX17, and DDX51 (82, 88, 93).

In addition to their substantial nucleolar localization, several R-loop–interacting DEAD-box proteins exhibited some localization within non-nucleolar regions of the nucleus (Fig. 2*A*). To test these proteins for potential roles in expression of protein-coding genes, we performed mRNA-Seq upon KD of the same set of DEAD-box proteins. We observed very high correlation among replicates of each individual KD as well as high correlation among different *Ddx* KDs (supplemental Fig. S4*A*). Using a cutoff of twofold upregulated or downregulated in any KD relative to control and adjusted p value <0.05, we observed 737 genes that were differentially expressed, with more genes upregulated than downregulated, in one or more *Ddx* KD relative to controls (supplemental Table S8 and Fig. 4*A*). Consistent with the high correlation of their overall expression profiles, the sets of genes upregulated or downregulated by depletion of each factor were highly overlapping. Of the 603 genes significantly upregulated in any of the four KDs, 209 were shared among all four of the KDs (Fig. 4*D*), including transcripts, such as *Cdkn1a*, *Cdkn2b*, and *Wnt5a*. The observed changes in expression of these transcripts upon KD of each DEAD-box protein were further confirmed by RT-qPCR (supplemental Fig. S4, *B–D*). Of the 134 genes significantly downregulated in at least one KD, 19 were downregulated in all four KDs (Fig. 4*E*). This overlap was also significantly more than expected ($p = 1.27 \times 10^{-133}$), albeit lower than observed for upregulated genes. The substantial overlap of genes misregulated by KD of each factor suggested either that DDX10, DDX24, DDX27, and DDX54 regulate mRNA levels through a shared pathway or that the genes observed to be misregulated were more sensitive to the alterations in rRNA levels observed upon KD of these proteins. Notably, GO term analysis of upregulated genes revealed enrichment for multiple terms related to cellular differentiation, including "neuronal differentiation," "cell fate commitment," and "cell migration" (Fig. 4*B*). Similarly, downregulated genes were also enriched for terms related to development and differentiation, including genes that regulate lipid and lipoprotein homeostasis, such as *Adipoq* (94) and *Pcsk9* (95) (Fig. 4*C* and supplemental Fig. S4*E*). These findings raise the possibility that some alterations in gene expression upon depletion of these DEAD-box family helicases may be specific to ESCs,

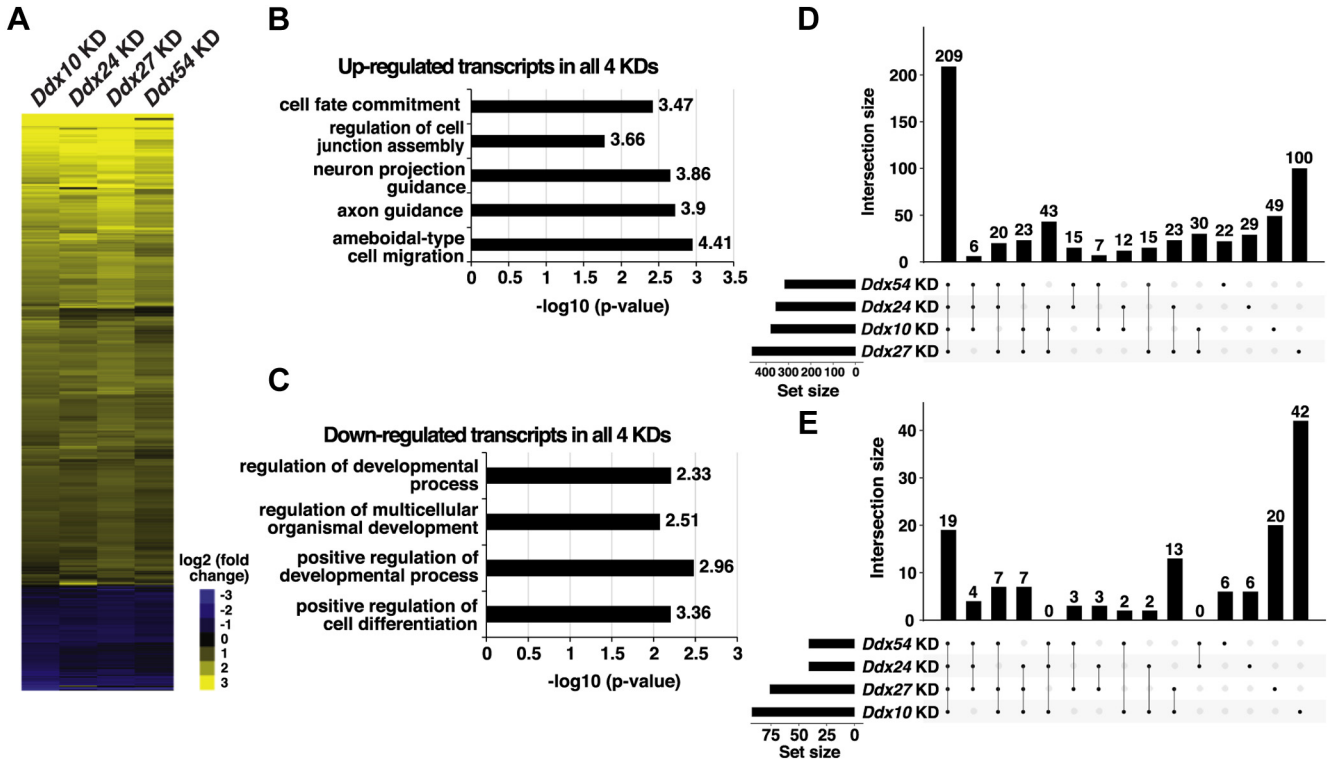


FIG. 4. **Regulation of mRNA expression by R-loop–interacting DEAD-box proteins.** *A*, genes significantly misregulated by *Ddx* knockdown (KD) as measured by mRNA-Seq. The 737 transcripts at least twofold upregulated or downregulated in any KD relative to control (with a $p < 0.05$) are depicted in the heat map. *B*, Gene Ontology (GO) term analysis of the genes significantly upregulated in all four KDs. The top five GO pathways are shown. *C*, GO term analysis of the genes significantly downregulated. The only four pathways reaching statistical significance are shown. *D*, overlap of genes significantly upregulated in any of the four KDs. *Linked dots* represent genes shared between the indicated KDs. *E*, overlap of the genes significantly downregulated in any of the four KDs, depicted as in *D*.

with potential implications for maintenance of the pluripotent state or regulation of ESC differentiation.

RNA–Protein Crosslinking Uncovers Transiently or Weakly Interacting R-Loop–Associated Proteins

Although proteins that bind to or near R-loops can be identified by co-IP using the S9.6 antibody, it is difficult to distinguish proteins that bind directly to RDHs from the proteins that bind to chromatin flanking the RDHs. In addition, while our stringent chromatin wash reduces background in our co-IP procedure, it likely also removes RDH-associated proteins that interact transiently or weakly with these structures. To systematically and specifically identify proteins that directly bind to the RDH structure of R-loop, we adapted a crosslinking protocol previously used to enrich for RNA-associated proteins (96) and coupled this procedure with S9.6 co-IP.

To this end, we cultured mESCs with 4SU to incorporate this nucleotide analog into RNA transcripts, followed by irradiation with intermediate wavelength (312 nm) UV light, which has been shown to induce crosslinks between 4SU-labeled RNA and their direct binding proteins (96, 97). To precisely quantify the extent to which crosslinking affected enrichment of proteins, we introduced SILAC (98, 99) to directly compare protein abundance in the presence or the absence of 4SU addition. After 4SU or vehicle addition and UV treatment, cells were lysed and their

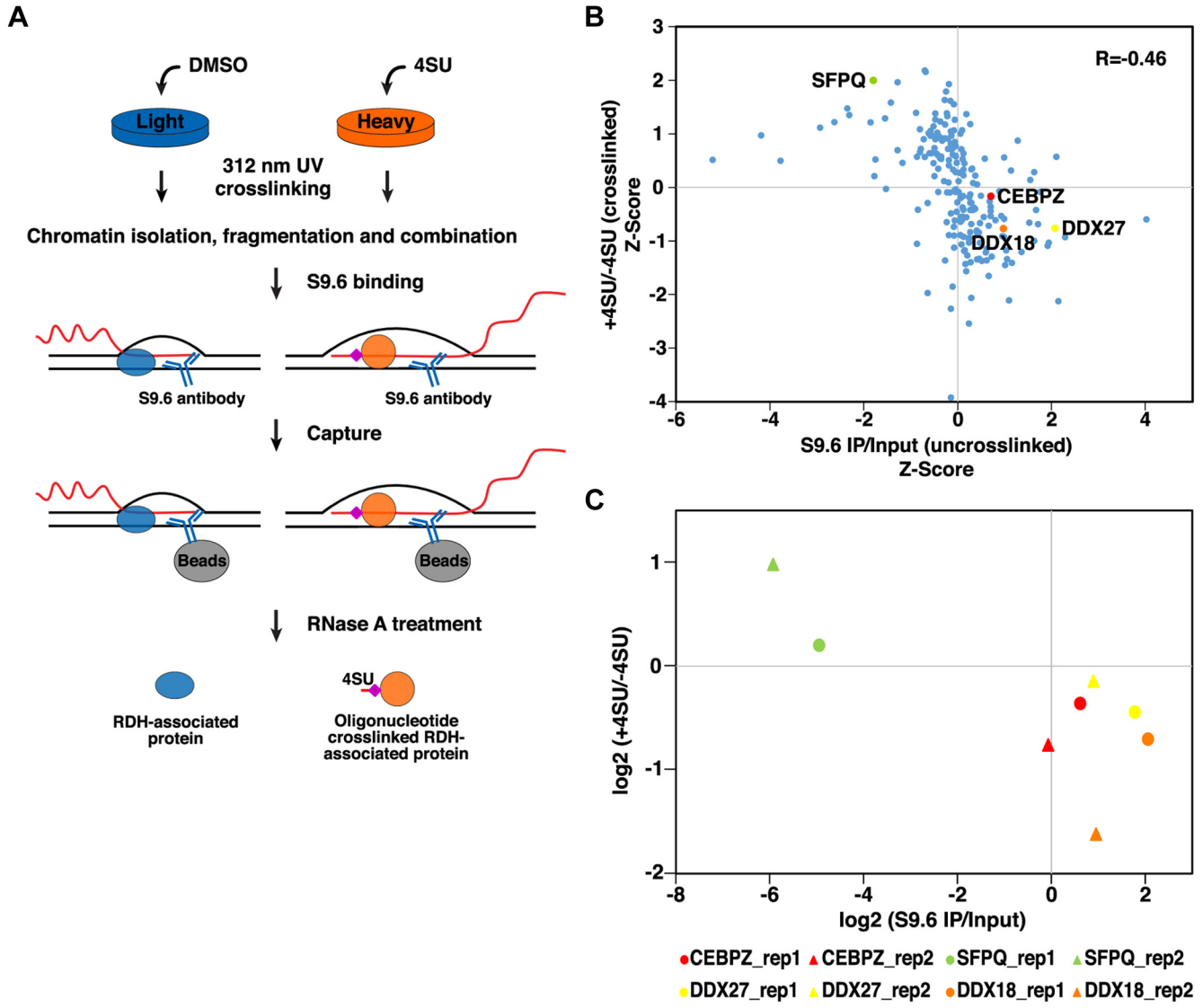


FIG. 5. **RNA–protein crosslinking enables identification of weakly interacting R-loop–binding proteins.** *A*, schematic diagram of 4SU-labeled crosslinking R-loop purification approach. *B*, comparison of protein enrichment by S9.6 IP in the presence or the absence of crosslinking. Normalized enrichment of the 214 proteins identified by both methods is plotted, and the Pearson's correlation coefficient is indicated. *C*, measurement of the effect of crosslinking on selected R-loop–interacting proteins by quantitative Western blotting. Two biological replicates were performed. 4SU, 4-thiouridine; IP, immunoprecipitation.

chromatin was mixed in equal amounts, followed by S9.6 IP and LC–MS/MS for each of three independent replicates (Fig. 5*A*).

After removing proteins that appeared only in one replicate, 231 proteins were left for downstream analysis (supplemental Tables S9 and S10). Interestingly, although we observed enrichment of 116 proteins upon 4SU addition, another group of 115 proteins was reduced in the 4SU samples relative to non-4SU-containing samples. Whereas the enriched proteins were likely proteins that directly interact with the RNA part of R-loops, either transiently or weakly, those that were reduced in the crosslinked samples may represent proteins that interact very strongly with R-loops with multiple RNA contacts throughout each polypeptide. For proteins in this latter category, multiple protein–RNA crosslinks may render a substantial fraction of peptides "unreadable" by MS because of the covalent attachment of oligonucleotides of unknown mass that remain even after bulk digestion of RNA in the immunoprecipitates. Although 202 of 231 proteins identified in the crosslinking dataset overlapped with the 335 R-loop–interacting proteins identified in our initial studies, an additional 29 proteins were recovered in the crosslinked samples, for a total of 364 R-loop–binding proteins overall (supplemental Table S10). Based on GO term analysis, these 29 additional proteins included proteins that function in mRNA splicing or transport, raising the possibility that some of these proteins were only weakly or transiently associated with R-loops because of active roles in removing transcripts from contexts conducive to R-loop formation.

To explore the possibility that crosslinking may enrich weakly interacting proteins while depleting strongly interacting proteins, we directly compared the two datasets, hereafter referred to as "crosslinked" and "uncrosslinked." For this comparison, we included proteins that appeared in two or more replicates in the uncrosslinked data and further filtered this list to include only those proteins identified by both MS approaches, leaving 214 total proteins. Interestingly, we found that the proteins with higher enrichment in the presence of 4SU than in its absence were among the most poorly enriched in the uncrosslinked data relative to input (Fig. 5*B*, *upper left quadrant*). Conversely, the proteins for which enrichment was reduced by crosslinking were often among the highly enriched in the uncrosslinked data (Fig. 5*B*, *lower right quadrant*). Consequently, we observed a negative correlation between the two datasets (Fig. 5*B*). To further validate these results, we performed S9.6 co-IP followed by quantitative Western blotting, with or without crosslinking, on a sample of proteins across the enrichment spectrum. We examined three proteins (CEBPZ, DDX18, and DDX27) reduced in the presence of 4SU-dependent crosslinking and one (SFPQ) enriched in the presence of crosslinking (Fig. 5*B*). Consistent with the MS data, the quantitative Western blotting showed an anti-correlation pattern (Fig. 5*C*). Overall, these findings suggest that crosslinking improves enrichment of weakly or transiently RDH-interacting proteins that bind directly to RNA but reduces enrichment of some strongly associated proteins. Previous studies of RNA–protein interactions revealed that proteins and protein domains directly associated with RNA can be underrepresented in the spectra upon crosslinking because of the covalent attachment of RNA nucleotides (96). A similar phenomenon may explain the crosslinking-dependent reduction we observe for strongly enriched R-loop–associated factors.

DISCUSSION

In this study, we established approaches for rigorous identification of R-loop–associated proteins, using a modified IP approach that includes stringent extraction of chromatin proteins prior to pull down, with or without selective RNA–protein crosslinking. These approaches uncovered several hundred associated proteins, including several known R-loop regulatory factors and a number of novel R-loop–interacting proteins. Consistent with their potential roles in regulation of RDHs or association with R-loops once they form, many of these proteins were known or predicted RNA-binding proteins. One of the most enriched proteins was DHX9, which has previously been shown to bind R-loops and regulate their formation (50, 64). Moreover, DHX9 is also known to bind G-quadruplexes within DNA (63), suggesting a possible role for DHX9 in coordinate regulation of R-loops and G-quadruplexes where they colocalize in the genome (16).

Upon examining the functions of R-loop–interacting proteins identified in this study, GO terms related to rRNA were enriched, and factors that localize within the nucleolus and function in rRNA processing were particularly prominent. The proteins most highly enriched by S9.6 tend to be nucleolar, often with known or predicted functions in rRNA production or processing, in contrast to the relatively lowly enriched proteins with no apparent nucleolar localization or functions in rRNA metabolism. These findings lend additional support to the idea that a significant fraction of R-loops within cells are located within nucleolus because of RNAP I–mediated transcription (29, 52).

Members of the DEAD-box family of RNA helicases were overrepresented among the R-loop–associated RNA-binding proteins identified. DEAD-box helicases are known to be involved in RNA processing, including transcription, splicing, and RNA decay (100). Several DEAD-box proteins have established roles in regulation of rRNA (82). For example, DDX5 functions in rRNA transcription and processing (87, 93), whereas DDX51 was shown to be necessary for normal cleavage of pre-rRNAs (88). Consistent with these roles, we found that R-loop–interacting proteins DDX18, DDX24, and DDX27 localized largely to the nucleolus, with lower levels of localization to non-nucleolar chromatin (Fig. 2*A*). To further study the functions of several R-loop–interacting DEAD-box proteins, we tested the effects of DDX10, DDX24, DDX27,

and DDX54 loss using RNA interference. Northern blotting revealed increased accumulation of different pre-rRNAs in different DEAD-box KD cells, suggesting these DEAD-box proteins impact rRNA maturation at several different steps. Each of these DEAD-box proteins has previously been implicated in aspects of rRNA production in multiple organisms. Human DDX10 was previously shown to reside within a complex that contributes to 18S rRNA processing (84). Consistent with this finding, we showed that *Ddx10* KD caused accumulation of 34S pre-rRNA, a precursor of mature 18S rRNA. Mak5p, the yeast homolog of mouse DDX24, has been shown to facilitate processing of the yeast homolog of 28S rRNA (86), whereas human *Ddx24* KD showed accumulation of 36S pre-rRNA (101), consistent with our results. KD of *Ddx27* in U2OS cells was previously shown to reduce the abundance of 28S mature rRNA (85). Our results showed that *Ddx27* KD caused accumulation of 36S, a precursor of 28S, providing a potential explanation for this earlier finding. Finally, DDX54 was previously shown to regulate levels of precursors of 18S rRNA (89), while we found that DDX54 regulates pre-rRNA precursors of 28S rRNA. These differences could potentially be due to differences in experimental system or cell type. In total, our findings confirm and add additional insights into the known roles of DEAD-box helicases in rRNA processing. These data suggest that resolution of R-loops by multiple RNA helicases acting nonredundantly may be necessary to efficiently process the pre-rRNA transcript into mature rRNAs.

In addition, by performing mRNA-Seq, we observed misregulation of a highly overlapping set of genes in all four KDs, including genes implicated in cellular differentiation and migration. These findings raise the possibility that DEAD-box helicases act in a common pathway to regulate the levels of hundreds of mRNAs. These effects could potentially occur through direct effects on a shared set of mRNA-encoding genes. Consistent with this possibility, DDX17, which was known to regulate rRNA processing, has also been shown to interact with the estrogen receptor-alpha transcription factor and regulate its target genes (102). Similarly, DDX21 has been shown to release pausing of RNAP II by resolving R-loops associated with estrogen-regulated genes (41). Alternatively, the shared set of genes we found to be misregulated upon KD of DDX10, DDX24, DDX27, or DDX54 in mESCs could be particularly sensitive to changes in rRNA levels.

Although stringent high salt extraction of chromatin prior to IP of R-loops may reduce contamination with general chromatin-binding proteins, this harsh treatment may also result in loss of dynamic or weakly binding proteins that nevertheless play roles in R-loop regulation or function. An example of one such protein was SFPQ, which was detectable by S9.6 IP/Western blotting but was only significantly enriched in our S9.6 IP/MS data upon crosslinking. To identify such proteins, we introduced a specific RNA–protein crosslinking step into our IP protocol and utilized SILAC labeling to increase sensitivity. By comparing the array of R-loop–interacting proteins identified with and without crosslinking, we observed that crosslinking increased enrichment of numerous proteins that were weakly enriched in the absence of crosslinking, as predicted, as well as a number of proteins that were not identified in the absence of crosslinking. 4SU is a "zero length" crosslinking reagent, suggesting proteins enriched upon crosslinking may directly interact with RDHs (97, 103, 104). One such protein was SFPQ, a splicing factor known to regulate R-loops (64). In addition, STRBP, a protein that binds to both DNA and RNA, along with HNRH1 and HNRH2 was also enriched by crosslinking. The latter two proteins belong to the heterogeneous nuclear ribonucleoprotein family that functions in various RNA-dependent processes, including regulation of R-loop abundance (105). Perhaps more interestingly, a large fraction of proteins highly enriched in the uncrosslinked dataset exhibited reduced enrichment in the presence of crosslinking, suggesting that many proteins that strongly interact with the RNA component of R-loops may become less easily detected by MS approaches, because of the covalent addition of RNA nucleotides of unknown size. These two methods in combination enabled us to identify 364 stringent R-loop–interacting proteins, including both known regulators of R-loops and proteins previously not shown to bind these structures. Taken together, these findings suggest that both uncrosslinked and crosslinked S9.6 co-IP offer advantages for identification of the R-loop interactome *in vivo*. These studies serve as a resource for uncovering the mechanisms by which R-loops are regulated as well as the means by which R-loops might affect regulation or processing of RNA. Given the key roles of R-loops in gene regulation, DNA damage, and genomic instability, a better understanding of the factors that bind and regulate R-loop formation will lead to a better understanding of developmental gene regulation and diseases.

DATA AVAILABILITY

Primary MS data are available *via* ProteomeXchange with identifier PXD022697. RNA-Seq data are deposited in Gene Expression Omnibus (GSE161890).

Supplemental data—This article contains supplemental data.

Acknowledgments—We thank J. Benanti, S. Gopalan, E. Hass, C.-H. Chao, and A. MacDonagh for critical comments on the article. We thank C. Song for a detailed S9.6 IP protocol and S. Wieczorek for advice on use of ProStar. We thank A. Luciano and M. Gura for their assistance with Northern blotting experiments.

Funding and additional information—These studies were supported by R01HD072122 (to T. G. F.) and R01HD093783 (to F. C. and T. G. F.). T. W. was supported in part by Li Weibo PhD Scholarship Fund. The content is solely the responsibility

of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions—T. W. conceptualization; T. W. and F. C. methodology; T. W. validation; T. W. and J. N. investigation; T. W. and T. G. F. writing–original draft; T. W. and T. G. F. writing–review and editing; F. C. and T. G. F. supervision.

Conflict of interest—The authors declare no competing interests.

Abbreviations—The abbreviations used are: 4SU, 4-thiouridine; BSA, bovine serum albumin; cDNA, complementary DNA; FBRL, fibrillarin; G4, G-quadruplex; GO, Gene Ontology; IP, immunoprecipitation; KD, knockdown; mESC, mouse embryonic stem cell; PBST, PBS with Tween-20; qPCR, quantitative PCR; RDH, RNA–DNA hybrid; RNAP, RNA polymerase; SILAC, stable isotope labeling by amino acids in cell culture; SSC, saline-sodium citrate.

Received November 25, 2020, and in revised from, August 21, 2021
Published, MCPRO Papers in Press, August 31, 2021, https://doi.org/10.1016/j.mcpro.2021.100142

REFERENCES

1. Lang, K. S., Hall, A. N., Merrikh, C. N., Ragheb, M., Tabakh, H., Pollock, A. J., Woodward, J. J., Dreifus, J. E., and Merrikh, H. (2017) Replication-transcription conflicts generate R-loops that orchestrate bacterial stress survival and pathogenesis. *Cell* **170**, 787–799
2. Wahba, L., Costantino, L., Tan, F. J., Zimmer, A., and Koshland, D. (2016) S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev.* **30**, 1327–1338
3. Xu, W., Xu, H., Li, K., Fan, Y., Liu, Y., Yang, X., and Sun, Q. (2017) The R-loop is a common chromatin feature of the Arabidopsis genome. *Nat. Plants* **3**, 704–714
4. Chen, P. B., Chen, H. V., Acharya, D., Rando, O. J., and Fazzio, T. G. (2015) R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nat. Struct. Mol. Biol.* **22**, 999–1007
5. Sanz, L. A., Hartono, S. R., Lim, Y. W., Steyaert, S., Rajpurkar, A., Ginno, P. A., Xu, X., and Chédin, F. (2016) Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol. Cell* **63**, 167–178
6. Belotserkovskii, B. P., Tornaletti, S., D'Souza, A. D., and Hanawalt, P. C. (2018) R-loop generation during transcription: Formation, processing and cellular outcomes. *DNA Repair (Amst.)* **71**, 69–81
7. Roy, D., and Lieber, M. R. (2009) G clustering is important for the initiation of transcription-induced R-loops *in vitro*, whereas high G density without clustering is sufficient thereafter. *Mol. Cell. Biol.* **29**, 3124–3133
8. Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F., and Maizels, N. (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.* **18**, 1618–1629
9. Stolz, R., Sulthana, S., Hartono, S. R., Malig, M., Benham, C. J., and Chedin, F. (2019) Interplay between DNA sequence and negative superhelicity drives R-loop structures. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 6260–6269
10. Allison, D. F., and Wang, G. G. (2019) R-loops: Formation, function, and relevance to cell stress. *Cell Stress* **3**, 38–46
11. Gyi, J. I., Conn, G. L., Lane, A. N., and Brown, T. (1996) Comparison of the thermodynamic stabilities and solution conformations of DNA·RNA hybrids containing purine-rich and pyrimidine-rich strands with DNA and RNA duplexes. *Biochemistry* **35**, 12538–12548
12. Conn, G. L., Brown, T., and Leonard, G. A. (1999) The crystal structure of the RNA/DNA hybrid r(GAAGAGAAGC)·d(GCTTCTCTTC) shows significant differences to that found in solution. *Nucleic Acids Res.* **27**, 555–561
13. Li, D., Peng, P., Yang, Z., and Lv, B. (2019) Formation of G-quadruplex structure in supercoiled DNA under molecularly crowded conditions. *RSC Adv.* **9**, 26248–26251
14. Ma, J., and Wang, M. D. (2016) DNA supercoiling during transcription. *Biophys. Rev.* **8**, 75–87
15. Lee, C.-Y., McNerney, C., Ma, K., Zhao, W., Wang, A., and Myong, S. (2020) R-loop induced G-quadruplex in non-template promotes transcription by successive R-loop formation. *Nat. Commun.* **11**, 3392
16. Miglietta, G., Russo, M., and Capranico, G. (2020) G-quadruplex-R-loop interactions and the mechanism of anticancer G-quadruplex binders. *Nucleic Acids Res.* **48**, 11942–11957
17. Boguslawski, S. J., Smith, D. E., Michalak, M. A., Mickelson, K. E., Yehle, C. O., Patterson, W. L., and Carrico, R. J. (1986) Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *J. Immunol. Methods* **89**, 123–130
18. Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I., and Chédin, F. (2012) R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **45**, 814–825
19. Ginno, P. A., Lim, Y. W., Lott, P. L., Korf, I., and Chedin, F. (2013) GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Res.* **23**, 1590–1600
20. El Hage, A., Webb, S., Kerr, A., and Tollervey, D. (2014) Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genet.* **10**, e1004716
21. Chen, L., Chen, J. Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., Tang, P., Qian, H., Luo, D., Li, H., Zhou, Y., Zhang, D. E., and Fu, X. D. (2017) R-ChIP using inactive RNase H reveals dynamic coupling of R-loops with transcriptional pausing at gene promoters. *Mol. Cell* **68**, 745–757.e5
22. Chen, J.-Y., Zhang, X., Fu, X.-D., and Chen, L. (2019) R-ChIP for genome-wide mapping of R-loops by using catalytically inactive RNASEH1. *Nat. Protoc.* **14**, 1661–1685
23. Yan, Q., Shields, E. J., Bonasio, R., and Sarma, K. (2019) Mapping native R-loops genome-wide using a targeted nuclease approach. *Cell Rep.* **29**, 1369–1380
24. Dumelie, J. G., and Jaffrey, S. R. (2017) Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. *Elife* **6**, e28306
25. Chédin, F., Hartono, S. R., Sanz, L. A., and Vanoosthuyse, V. (2021) Best practices for the visualization, mapping, and manipulation of R-loops. *EMBO J.* **40**, e106394
26. Tan-Wong, S. M., Dhir, S., and Proudfoot, N. J. (2019) R-loops promote antisense transcription across the mammalian genome. *Mol. Cell* **76**, 600–616
27. Yang, X., Liu, Q.-L., Xu, W., Zhang, Y.-C., Yang, Y., Ju, L.-F., Chen, J., Chen, Y.-S., Li, K., Ren, J., Sun, Q., and Yang, Y.-G. (2019) m6A promotes R-loop formation to facilitate transcription termination. *Cell Res.* **29**, 1035–1038
28. Skourti-Stathaki, K., Proudfoot, N. J., and Gromak, N. (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* **42**, 794–805
29. El Hage, A., French, S. L., Beyer, A. L., and Tollervey, D. (2010) Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev.* **24**, 1546–1558
30. Skourti-Stathaki, K., Torlai Triglia, E., Warburton, M., Voigt, P., Bird, A., and Pombo, A. (2019) R-loops enhance polycomb repression at a subset of developmental regulator genes. *Mol. Cell* **73**, 930–945
31. Daniels, G. A., and Lieber, M. R. (1995) RNA: DNA complex formation upon transcription of immunoglobulin switch regions: Implications for the mechanism and regualtion of class switch recombination. *Nucleic Acids Res.* **23**, 5006–5011
32. Yu, K., Chedin, F., Hsieh, C. L., Wilson, T. E., and Lieber, M. R. (2003) R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat. Immunol.* **4**, 442–451
33. Wan, Y., Zheng, X., Chen, H., Guo, Y., Jiang, H., He, X., Zhu, X., and Zheng, Y. (2015) Splicing function of mitotic regulators links R-loop-mediated DNA damage to tumor cell killing. *J. Cell Biol.* **209**, 235–246
34. Holmes, J. B., Akman, G., Wood, S. R., Sakhuja, K., Cerritelli, S. M., Moss, C., Bowmaker, M. R., Jacobs, H. T., Crouch, R. J., and Holt, I. J. (2015) Primer retention owing to the absence of RNase H1 is catastrophic for

mitochondrial DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 9334–9339
35. Sollier, J., and Cimprich, K. A. (2015) Breaking bad: R-loops and genome integrity. *Trends Cell Biol.* **25**, 514–522
36. Crossley, M. P., Bocek, M., and Cimprich, K. A. (2019) R-loops as cellular regulators and genomic threats. *Mol. Cell* **73**, 398–411
37. Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T., and Cimprich, K. A. (2017) Transcription-replication conflict orientation modulates R-loop levels and activates distinct DNA damage responses. *Cell* **170**, 774–786.e19
38. Richard, P., and Manley, J. L. (2017) R loops and links to human disease. *J. Mol. Biol.* **429**, 3168–3180
39. Salvi, J. S., and Mekhail, K. (2015) R-loops highlight the nucleus in ALS. *Nucleus* **6**, 23–29
40. Groh, M., Lufino, M. M. P., Wade-Martins, R., and Gromak, N. (2014) R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genet.* **10**, e1004318
41. Song, C., Hotz-Wagenblatt, A., Voit, R., and Grummt, I. (2017) SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. *Genes Dev.* **31**, 1370–1381
42. Stein, H., and Hausen, P. (1969) Enzyme from calf thymus degrading the RNA moiety of DNA-RNA hybrids: Effect on DNA-dependent RNA polymerase. *Science* **166**, 393–395
43. Cerritelli, S. M., and Crouch, R. J. (2009) Ribonuclease H: The enzymes in eukaryotes. *FEBS J.* **276**, 1494–1505
44. Cornelio, D. A., Sedam, H. N. C., Ferrarezi, J. A., Sampaio, N. M. V., and Argueso, J. L. (2017) Both R-loop removal and ribonucleotide excision repair activities of RNase H2 contribute substantially to chromosome stability. *DNA Repair (Amst.)* **52**, 110–114
45. Lockhart, A., Pires, V. B., Bento, F., Kellner, V., Luke-Glaser, S., Yakoub, G., Ulrich, H. D., and Luke, B. (2019) RNase H1 and H2 are differentially regulated to process RNA-DNA hybrids. *Cell Rep.* **29**, 2890–2900.e5
46. Santos-Pereira, J. M., and Aguilera, A. R. (2015) Loops: New modulators of genome dynamics and function. *Nat. Rev. Genet.* **16**, 583–597
47. Grunseich, C., Wang, I. X., Watts, J. A., Burdick, J. T., Guber, R. D., Zhu, Z., Bruzel, A., Lanman, T., Chen, K., Schindler, A. B., Edwards, N., Ray-Chaudhury, A., Yao, J., Lehky, T., Piszczek, G., *et al*. (2018) Senataxin mutation reveals how R-loops promote transcription by blocking DNA methylation at gene promoters. *Mol. Cell* **69**, 426–437
48. Arab, K., Karaulanov, E., Musheev, M., Trnka, P., Schäfer, A., Grummt, I., and Niehrs, C. (2019) GADD45A binds R-loops and recruits TET1 to CpG island promoters. *Nat. Genet.* **51**, 217–223
49. Zhang, C., Chen, L., Peng, D., Jiang, A., He, Y., Zeng, Y., Xie, C., Zhou, H., Luo, X., Liu, H., Chen, L., Ren, J., Wang, W., and Zhao, Y. (2020) METTL3 and N6-methyladenosine promote homologous recombination-mediated repair of DSBs by modulating DNA-RNA hybrid accumulation. *Mol. Cell* **79**, 425–442
50. Cristini, A., Groh, M., Kristiansen, M. S., and Gromak, N. (2018) RNA/DNA hybrid interactome identifies DXH9 as a molecular player in transcriptional termination and R-loop-associated DNA damage. *Cell Rep.* **23**, 1891–1905
51. Li, Y., Song, Y., Xu, W., Li, Q., Wang, X., Li, K., Wang, J., Liu, Z., Velychko, S., Ye, R., Xia, Q., Wang, L., Guo, R., Dong, X., Zheng, Z., *et al*. (2020) R-loops coordinate with SOX2 in regulating reprogramming to pluripotency. *Sci. Adv.* **6**, eaba0777
52. Shen, W., Sun, H., De Hoyos, C. L., Bailey, J. K., Liang, X. H., and Crooke, S. T. (2017) Dynamic nucleoplasmic and nucleolar localization of mammalian RNase H1 in response to RNAP I transcriptional R-loops. *Nucleic Acids Res.* **45**, 10672–10692
53. Velichko, A. K., Petrova, N. V., Luzhin, A. V., Strelkova, O. S., Ovsyannikova, N., Kireev, I. I., Petrova, N. V., Razin, S. V., and Kantidze, O. L. (2019) Hypoosmotic stress induces R loop formation in nucleoli and ATR/ATM-dependent silencing of nucleolar transcription. *Nucleic Acids Res.* **47**, 6811–6825
54. Wieczorek, S., Combes, F., Lazar, C., Giai Gianetto, Q., Gatto, L., Dorffer, A., Hesse, A. M., Couté, Y., Ferro, M., Bruley, C., and Burger, T. (2017) DAPAR & ProStaR: Software to perform statistical analyses in quantitative discovery proteomics. *Bioinformatics* **33**, 135–136
55. Ting, L., Cowley, M. J., Hoon, S. L., Guilhaus, M., Raftery, M. J., and Cavicchioli, R. (2009) Normalization and statistical analysis of quantitative proteomics data generated by metabolic labeling. *Mol. Cell. Proteomics* **8**, 2227–2242
56. Chen, G. (2013) *Characterization of Protein Therapeutics Using Mass Spectrometry*. Springer, New York, NY
57. Chu, F., Hogan, D., Gupta, R., Gao, X.-Z., Nguyen, H. T., and Cote, R. H. (2019) Allosteric regulation of rod photoreceptor phosphodiesterase 6 (PDE6) elucidated by chemical cross-linking and quantitative mass spectrometry. *J. Mol. Biol.* **431**, 3677–3689
58. Chalkley, R. J., Baker, P. R., Huang, L., Hansen, K. C., Allen, N. P., Rexach, M., and Burlingame, A. L. (2005) Comprehensive analysis of a multidimensional liquid chromatography mass spectrometry dataset acquired on a quadrupole selecting, quadrupole collision cells, time-of-flight mass spectrometer: II. New developments in protein prospector allow for reliable and comprehensive automatic analysis of large datasets. *Mol. Cell. Proteomics* **4**, 1194–1204
59. Deutsch, E. W., Bandeira, N., Sharma, V., Perez-Riverol, Y., Carver, J. J., Kundu, D. J., García-Seisdedos, D., Jarnuczak, A. F., Hewapathirana, S., Pullman, B. S., Wertz, J., Sun, Z., Kawano, S., Okuda, S., Watanabe, Y., *et al*. (2020) The ProteomeXchange consortium in 2020: Enabling 'big data' approaches in proteomics. *Nucleic Acids Res.* **48**, D1145–D1152
60. Fazzio, T. G., Huff, J. T., and Panning, B. (2008) An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* **134**, 162–174
61. Li, B., and Dewey, C. N. (2011) RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323
62. Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014) UpSet: Visualization of intersecting sets. *IEEE Trans. Vis. Comput. Graph.* **20**, 1983–1992
63. Chakraborty, P., and Grosse, F. (2011) Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair (Amst.)* **10**, 654–665
64. Chakraborty, P., Huang, J. T. J., and Hiom, K. (2018) DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing. *Nat. Commun.* **9**, 4346
65. Smolka, J. A., Sanz, L. A., Hartono, S. R., and Chédin, F. (2021) Recognition of RNA by the S9.6 antibody creates pervasive artifacts when imaging RNA:DNA hybrids. *J. Cell Biol.* **220**, e202004079
66. Halász, L., Karányi, Z., Boros-Oláh, B., Kuik-Rózsa, T., Sipos, É., Nagy, É., Mosolygó-L, Á., Mázló, A., Rajnavölgyi, É., Halmos, G., and Székvölgyi, L. (2017) RNA-DNA hybrid (R-loop) immunoprecipitation mapping: An analytical workflow to evaluate inherent biases. *Genome Res.* **27**, 1063–1073
67. Morello, L. G., Coltri, P. P., Quaresma, A. J., Simabuco, F. M., Silva, T. C., Singh, G., Nickerson, J. A., Oliveira, C. C., Moore, M. J., and Zanchin, N. I. (2011) The human nucleolar protein FTSJ3 associates with NIP7 and functions in pre-rRNA processing. *PLoS One* **6**, e29174
68. Jensen, B. C., Wang, Q., Kifer, C. T., and Parsons, M. (2003) The NOG1 GTP-binding protein is required for biogenesis of the 60 S ribosomal subunit. *J. Biol. Chem.* **278**, 32204–32211
69. Gómez-González, B., García-Rubio, M., Bermejo, R., Gaillard, H., Shirahige, K., Marín, A., Foiani, M., and Aguilera, A. (2011) Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J.* **30**, 3106–3119
70. Huertas, P., and Aguilera, A. (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell* **12**, 711–721
71. Linder, P. (2006) Dead-box proteins: A family affair-active and passive players in RNP-remodeling. *Nucleic Acids Res.* **34**, 4168–4180
72. Cordin, O., Banroques, J., Tanner, N. K., and Linder, P. (2006) The DEAD-box protein family of RNA helicases. *Gene* **367**, 17–37
73. Linder, P., Lasko, P. F., Ashburner, M., Leroy, P., Nielsen, P. J., Nishi, K., Schnier, J., and Slonimski, P. P. (1989) Birth of the D-E-A-D box. *Nature* **337**, 121–122
74. De Magis, A., Manzo, S. G., Russo, M., Marinello, J., Morigi, R., Sordet, O., and Capranico, G. (2019) DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 816–825
75. MacKay, R. P., Xu, Q., and Weinberger, P. M. (2020) R-loop physiology and pathology: A brief review. *DNA Cell Biol.* **39**, 1914–1925
76. Mishra, S. K., Tawani, A., Mishra, A., and Kumar, A. (2016) G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins. *Sci. Rep.* **6**, 38144

77. Brázda, V., Červeň, J., Bartas, M., Mikysková, N., Coufal, J., and Pečinka, P. (2018) The amino acid composition of quadruplex binding proteins reveals a shared motif and predicts new potential quadruplex interactors. *Molecules* **23**, 2341
78. McRae, E. K. S., Booy, E. P., Moya-Torres, A., Ezzati, P., Stetefeld, J., and McKenna, S. A. (2017) Human DDX21 binds and unwinds RNA guanine quadruplexes. *Nucleic Acids Res.* **45**, 6656–6668
79. Brázda, V., Laister, R. C., Jagelská, E. B., and Arrowsmith, C. (2011) Cruciform structures are a common DNA feature important for regulating biological processes. *BMC Mol. Biol.* **12**, 33
80. Kar, B., Liu, B., Zhou, Z., and Lam, Y. W. (2011) Quantitative nucleolar proteomics reveals nuclear re-organization during stress- induced senescence in mouse fibroblast. *BMC Cell. Biol.* **12**, 33
81. Villarreal, O. D., Mersaoui, S. Y., Yu, Z., Masson, J.-Y., and Richard, S. (2020) Genome-wide R-loop analysis defines unique roles for DDX5, XRN2, and PRMT5 in DNA/RNA hybrid resolution. *Life Sci. Alliance* **3**, e202000762
82. Martin, R., Straub, A. U., Doebele, C., and Bohnsack, M. T. (2013) DExD/H-box RNA helicases in ribosome biogenesis. *RNA Biol.* **10**, 4–18
83. Wild, T., Horvath, P., Wyler, E., Widmann, B., Badertscher, L., Zemp, I., Kozak, K., Csucs, G., Lund, E., and Kutay, U. (2010) A protein inventory of human ribosome biogenesis reveals an essential function of exportin 5 in 60S subunit export. *PLoS Biol.* **8**, e1000522
84. Turner, A. J., Knox, A. A., Prieto, J.-L., McStay, B., and Watkins, N. J. (2009) A novel small-subunit processome assembly intermediate that contains the U3 snoRNP, nucleolin, RRP5, and DBP4. *Mol. Cell. Biol.* **29**, 3007–3017
85. Kellner, M., Rohrmoser, M., Forné, I., Voss, K., Burger, K., Mühl, B., Gruber-Eber, A., Kremmer, E., Imhof, A., and Eick, D. (2015) DEAD-box helicase DDX27 regulates 3' end formation of ribosomal 47S RNA and stably associates with the PeBoW-complex. *Exp. Cell Res.* **334**, 146–159
86. Zagulski, M., Kressler, D., Bécam, A.-M., Rytka, J., and Herbert, C. J. (2003) Mak5p, which is required for the maintenance of the M1 dsRNA virus, is encoded by the yeast ORF YBR142w and is involved in the biogenesis of the 60S subunit of the ribosome. *Mol. Genet. Genomics* **270**, 216–224
87. Saporita, A. J., Chang, H.-C., Winkeler, C. L., Apicelli, A. J., Kladney, R. D., Wang, J., Townsend, R. R., Michel, L. S., and Weber, J. D. (2011) RNA helicase DDX5 is a p53-independent target of ARF that participates in ribosome biogenesis. *Cancer Res.* **71**, 6708–6717
88. Srivastava, L., Lapik, Y. R., Wang, M., and Pestov, D. G. (2010) Mammalian DEAD box protein Ddx51 acts in 3′ end maturation of 28S rRNA by promoting the release of U8 snoRNA. *Mol. Cell. Biol.* **30**, 2947–2956
89. Milek, M., Imami, K., Mukherjee, N., Bortoli, F. D., Zinnall, U., Hazapis, O., Trahan, C., Oeffinger, M., Heyd, F., Ohler, U., Selbach, M., and Landthaler, M. (2017) DDX54 regulates transcriptome dynamics during DNA damage response. *Genome Res.* **27**, 1344–1359
90. Morello, L. G., Hesling, C., Coltri, P. P., Castilho, B. A., Rimokh, R., and Zanchin, N. I. T. (2011) The NIP7 protein is required for accurate pre-rRNA processing in human cells. *Nucleic Acids Res.* **39**, 648–665
91. Henras, A. K., Plisson-Chastang, C., O'Donohue, M.-F., Chakraborty, A., and Gleizes, P.-E. (2015) An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip. Rev. RNA* **6**, 225–242
92. Moraleva, A., Magoulas, C., Polzikov, M., Hacot, S., Mertani, H. C., Diaz, J.-J., and Zatsepina, O. (2017) Involvement of the specific nucleolar protein SURF6 in regulation of proliferation and ribosome biogenesis in mouse NIH/3T3 fibroblasts. *Cell Cycle* **16**, 1979–1991
93. Jalal, C., Uhlmann-Schiffler, H., and Stahl, H. (2007) Redundant role of DEAD box proteins p68 (Ddx5) and p72/p82 (Ddx17) in ribosome biogenesis and cell proliferation. *Nucleic Acids Res.* **35**, 3590–3601
94. Hu, E., Liang, P., and Spiegelman, B. M. (1996) AdipoQ is a novel adipose-specific gene dysregulated in obesity. *J. Biol. Chem.* **271**, 10697–10703
95. Wu, N.-Q., and Li, J.-J. (2014) PCSK9 gene mutations and low-density lipoprotein cholesterol. *Clin. Chim. Acta* **431**, 148–153
96. He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D. C., Wilusz, J. E., Garcia, B. A., and Bonasio, R. (2016) High-resolution mapping of RNA-binding regions in the nuclear proteome of embryonic stem cells. *Mol. Cell* **64**, 416–430
97. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M., Ulrich, A., Wardle, G. S., Dewell, S., Zavolan, M., and Tuschl, T. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141
98. Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386
99. Mann, M. (2006) Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* **7**, 952–958
100. Rocak, S., and Linder, P. (2004) Dead-box proteins: The driving forces behind RNA metabolism. *Nat. Rev. Mol. Cell Biol.* **5**, 232–241
101. Yamauchi, T., Nishiyama, M., Moroishi, T., Yumimoto, K., and Nakayama, K. I. (2014) MDM2 mediates nonproteolytic polyubiquitylation of the DEAD-box RNA helicase DDX24. *Mol. Cell. Biol.* **34**, 3321–3340
102. Wortham, N. C., Ahamed, E., Nicol, S. M., Thomas, R. S., Periyasamy, M., Jiang, J., Ochocka, A. M., Shousha, S., Huson, L., Bray, S. E., Coombes, R. C., Ali, S., and Fuller-Pace, F. V. (2009) The DEAD-box protein p72 regulates ERα-/oestrogen-dependent transcription and cell growth, and is associated with improved survival in ERα-positive breast cancer. *Oncogene* **28**, 4053–4064
103. Favre, A., Moreno, G., Blondel, M. O., Kliber, J., Vinzens, F., and Salet, C. (1986) 4-Thiouridine photosensitized RNA-protein crosslinking in mammalian cells. *Biochem. Biophys. Res. Commun.* **141**, 847–854
104. Juzumiene, D., Shapkina, T., Kirillov, S., and Wollenzien, P. (2001) Short-range RNA-RNA crosslinking methods to determine rRNA structure and interactions. *Methods* **25**, 333–343
105. Alfano, L., Caporaso, A., Altieri, A., Dell'Aquila, M., Landi, C., Bini, L., Pentimalli, F., and Giordano, A. (2019) Depletion of the RNA binding protein HNRNPD impairs homologous recombination by inhibiting DNA-end resection and inducing R-loop accumulation. *Nucleic Acids Res.* **47**, 4068–4085