

Human proteins that interact with RNA/DNA hybrids

Isabel X. Wang,^{1,2} Christopher Grunseich,³ Jennifer Fox,^{1,2} Joshua Burdick,^{1,2} Zhengwei Zhu,^{2,4} Niema Ravazian,¹ Markus Hafner,⁵ and Vivian G. Cheung^{1,2,4}

¹Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA; ²Life Sciences Institute, University of Michigan, Ann Arbor, Michigan 48109, USA; ³Neurogenetics Branch, National Institute of Neurological Disorders and Stroke, NIH, Bethesda, Maryland 20892, USA; ⁴Department of Pediatrics, University of Michigan, Ann Arbor, Michigan 48109, USA; ⁵Laboratory of Muscle Stem Cells and Gene Regulation, National Institute of Arthritis and Musculoskeletal and Skin Diseases, Bethesda, Maryland 20892, USA

RNA/DNA hybrids form when RNA hybridizes with its template DNA generating a three-stranded structure known as the R-loop. Knowledge of how they form and resolve, as well as their functional roles, is limited. Here, by pull-down assays followed by mass spectrometry, we identified 803 proteins that bind to RNA/DNA hybrids. Because these proteins were identified using *in vitro* assays, we confirmed that they bind to R-loops *in vivo*. They include proteins that are involved in a variety of functions, including most steps of RNA processing. The proteins are enriched for K homology (KH) and helicase domains. Among them, more than 300 proteins preferred binding to hybrids than double-stranded DNA. These proteins serve as starting points for mechanistic studies to elucidate what RNA/DNA hybrids regulate and how they are regulated.

[Supplemental material is available for this article.]

RNA/DNA hybrids are abundant in human cells. They form during transcription when nascent RNA is in close proximity to its DNA template. The resulting RNA/DNA hybrids and the displaced single-stranded (ss) DNA are called R-loops. RNA/DNA hybrids are structurally different and more stable than the corresponding double-stranded DNAs (Bhattacharyya et al. 1990; Roberts and Crothers 1992).

RNA/DNA hybrids are found in origins of replication (Baker and Kornberg 1988; Xu and Clayton 1996), immunoglobulin class-switch regions (Yu et al. 2003), and transcription complexes (Hanna and Meares 1983; Nudler et al. 1997; Skourti-Stathaki et al. 2011). R-loops were mostly viewed as deleterious because they can lead to DNA damage. The unpaired DNA strand is vulnerable to damage (Huertas and Aguilera 2003; Li and Manley 2005; Mischo et al. 2011; Wahba et al. 2011), and improper processing of R-loops such as those mediated by transcription-coupled excision repair also results in DNA damage (Sollier et al. 2014). Increasingly, studies have shown that R-loops have regulatory roles. They are found abundantly in human gene promoters and terminators where RNA processing takes place (Ginno et al. 2012; Chen et al. 2017). Given these opposite impacts of R-loops, their formation and resolution must be regulated tightly. Genome-wide methods have mapped and quantified R-loops in yeast to human cells (Chan et al. 2014; El Hage et al. 2014; Wahba et al. 2016; Chen et al. 2017). With these methods, studies have shown that too many and too few R-loops lead to pathologic consequences. In immunodeficiencies such as Wiskott–Aldrich syndrome (Sarkar et al. 2017), and neurodegenerative diseases such as Friedreich ataxia (Groh et al. 2014), patients have more R-loops, whereas cells from ALS4 patients with the senataxin mutation have fewer R-loops (Grunseich et al. 2018).

Because the number and location of hybrids are critical to maintaining cellular function, most likely there are regulatory proteins that distinguish RNA/DNA hybrids from their double-stranded (ds) DNA counterparts. The structures of RNA/DNA hybrids with different sequences have been studied alone (Benevides et al. 1986; Fedoroff et al. 1993) and in complex with different proteins (Rychlik et al. 2010; Figiel and Nowotny 2014; Nishimasu et al. 2014; Bernecky et al. 2016). The results show that RNA/DNA hybrids do not adopt the traditional B-conformation of DNA or A-conformation of RNA but occur as mixtures or heteromeric duplexes (Fedoroff et al. 1993). It is well known that regulatory proteins recognize their targets by nucleic acid sequences and/or structures. Transcription factors often identify their targets based on sequences. In contrast, there are proteins that recognize their targets by structures and not just by sequences. Some proteins can target specifically different components (the RNA or hybrid) of the R-loops. For example, the conformation of R-loop is critical for the cleavage of the two DNA strands by Cas9 (Jiang et al. 2016). Ribonuclease H1 (also known as RNase H1) cleaves the RNA of RNA/DNA hybrids (Fedoroff et al. 1993; Cerritelli and Crouch 1995), whereas activation-induced cytidine deaminase (AID) favors binding to RNA/DNA hybrids (Abdouni et al. 2018). Recently, we showed that DNA methyltransferase 1 (DNMT1) binds more avidly to dsDNA than to the corresponding RNA/DNA hybrids; thus, the formation of the hybrid promotes transcription by preventing methylation-induced silencing (Grunseich et al. 2018). Presumably, there are other proteins like DNMT1 whose regulatory roles can be influenced by RNA/DNA hybrids.

The roles of RNA/DNA hybrids are beginning to be recognized, but much remains unknown. It is not clear what regulates the formation and resolution of RNA/DNA hybrids. It is also not known how hybrids affect processing of RNA and what transcriptional

Corresponding author: vgcheung@med.umich.edu

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.237362.118>. Freely available online through the *Genome Research* Open Access option.

© 2018 Wang et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

steps they regulate. Naturally occurring mutations and yeast mutant collections have facilitated much of the mechanistic studies of R-loops. But the mutant screens alone have yet to yield a comprehensive view of R-loops. High-throughput methods to identify proteins that interact with nucleic acids have provided valuable information on gene regulation (Hafner et al. 2010; Zhao et al. 2010; Baltz et al. 2012; Panda et al. 2014). A comprehensive list of proteins that interact with R-loops will facilitate studies on formation and processing of R-loops as well as their regulatory roles. Here, we report pull-down assays followed by mass spectrometry and *in vivo* confirmation studies that identified more than 800 proteins that bind to the RNA/DNA hybrids of R-loops in human cells.

Results

To identify proteins that bind to RNA/DNA hybrids, we made hybrids corresponding to two R-loops identified previously by S9.6 DRIP-seq (Grunseich et al. 2018): One is in the 5' end of the *BAMBI* gene, and the second is in the 3' end of the *DPP9* gene. We synthesized 600-mer and 90-mer RNA/DNA hybrids that correspond to sequences underlying R-loops in *BAMBI* and *DPP9*, respectively. Figure 1A shows the locations of the R-loops in *BAMBI* and *DPP9*. The *BAMBI* and *DPP9* regions are GC-rich, with GC content of 76% and 62%, respectively, consistent with findings that regions with G-rich RNA and complementary C-rich DNA are prone to hybrid formation (Roy and Lieber 2009; Skourti-Stathaki et al. 2011; Ginno et al. 2012). The R-loop in the *BAMBI* promoter was extensively characterized previously (Grunseich et al. 2018). We validated the R-loop in the 3' UTR of *DPP9* by S9.6 precipitation in this study (Fig. 1B). To check the integrity of the two RNA/DNA hybrids, we confirmed their sensitivity to RNase H1 (Cerritelli and Crouch 1995, 2009) and resistance to ribonuclease T1 (rntA, also known as RNase T1) (Fig. 1C; Zuo and Deutscher 2002).

To find proteins that bind to these hybrids, we added biotinylated forms of the *BAMBI* and *DPP9* hybrids to human B-cell extracts and carried out pull-down assays (Fig. 2A). Liquid chromatography followed by tandem mass spectrometry (LC-MS/MS) were performed to identify the proteins bound to the two hybrids. We used stringent inclusion criteria (Methods); each protein must be represented by four or more peptides with unique sequences. Despite these criteria, we identified a large number of proteins in the pull-down assays, namely, 1460 proteins with the *BAMBI* hybrid and 1018 proteins with the *DPP9* hybrid, in which 803 proteins were identified by both (Supplemental Table S1). Among the proteins identified in our *BAMBI* and *DPP9* pull-down assays

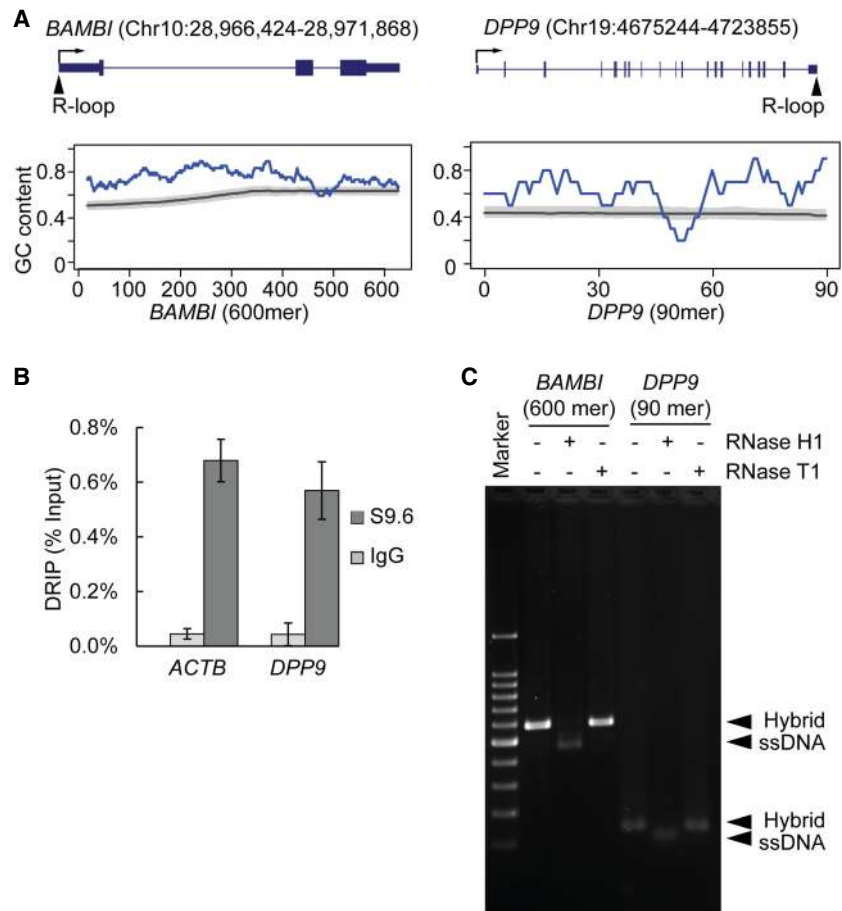


Figure 1. R-loops at GC-rich regions in the *BAMBI* promoter and *DPP9* 3' UTR. (A) Location and GC content of sequences underlying R-loops in *BAMBI* and *DPP9*. (Top) R-loop location is marked on gene models of *BAMBI* and *DPP9*. Boxes represent exons, and lines represent introns. Arrows show transcription start site and direction of transcription. (Bottom) GC content of the 600-mer and 90-mer sequences corresponding to RNA sequence in the R-loops (blue line). GC content is calculated as $(G + C)/(G + C + A + U)$ in the 50-nt sliding window for 600-mer or 10-nt sliding window for 90-mer. Genome background of GC content is calculated from corresponding regions of 14,587 RefSeq genes that are at least 2 kb long and 1 kb away from neighboring genes. The gray line represents median GC content, and the shade represents $\pm 10\%$. (B) The S9.6 antibody specifically pulled down R-loops at the 3' UTR of *DPP9*. DRIP was carried out using an S9.6 antibody or nonspecific IgG. Precipitated DNA was amplified using primers specific for *DPP9* 3' UTR. Primers specific for a previously reported R-loop region at the 3' UTR of *ACTB* were used as positive control. (Error bars) SEM of triplicates. (C) Integrity of the RNA/DNA hybrid was confirmed using RNase H1 and RNase T1. As expected, RNase H1 specifically digested RNA in the hybrids, leaving ssDNA as a product. RNase T1, which is specific for ssRNA, did not cleave the hybrids.

are RNase H1 (RNASEH1) and XRN2, which are known to bind to and modify RNA/DNA hybrids (Table 1), confirming that our approach identifies enzymes that process RNA/DNA hybrids (Stein and Hausen 1969; Keller and Crouch 1972; Skourti-Stathaki et al. 2011). Most of the identified proteins have not been reported to associate with hybrids. We validated the interaction between proteins and hybrids by Western blot (Fig. 2B). To ensure that the proteins are binding to hybrids and not to single-stranded RNAs, we showed that digestion by RNase H1 abolished the interactions, whereas RNase T1 did not interfere with the protein-hybrid interactions.

Many of the hybrid-binding proteins interact with R-loops in human cells. To ensure that the hybrid-interacting proteins we identified reflect *in vivo* interactions, we carried out three independent analyses. First, we looked for overlap between our proteins and those determined by Gromak and colleagues to bind to

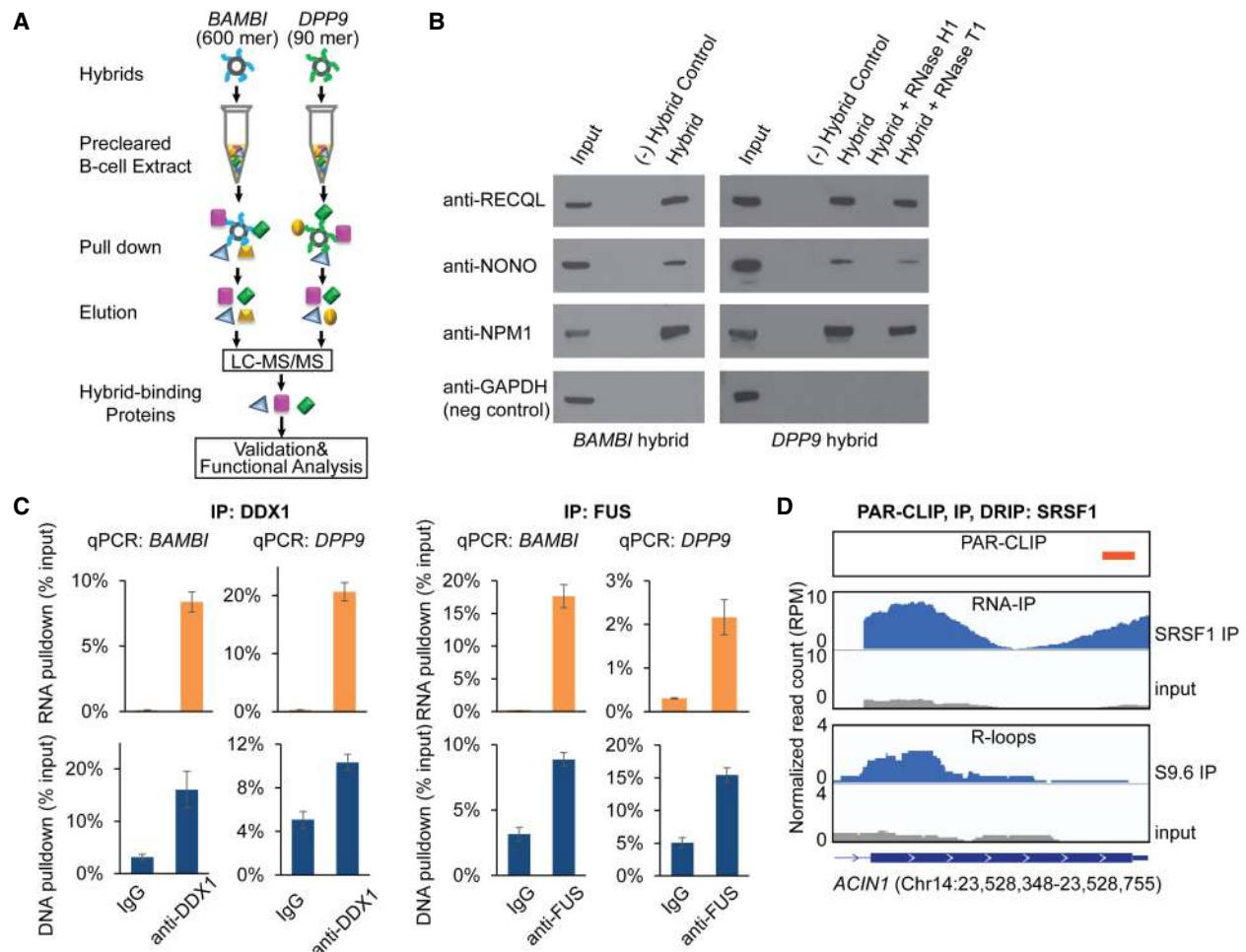


Figure 2. Identification of proteins that bind specifically to RNA/DNA hybrids. (A) Schematic of the experimental procedure. Biotinylated hybrids were conjugated to streptavidin beads and incubated with B-cell extracts. The proteins pulled down by hybrids were identified through proteomic analysis (LC-MS/MS). Only proteins that were bound by both hybrids were retained for further analysis. (B) Hybrid-binding proteins identified by proteomic analysis were validated by Western blot. B-cell extract (input) was incubated with no hybrid as a negative control, *BAMBI* or *DPP9* hybrid, *DPP9* hybrid pretreated with RNase H1 or RNase T1, respectively. Proteins were pulled down by biotinylated hybrids and analyzed by Western blot. Interactions with the pulled down proteins were eliminated by RNase H1 digestion but were not affected by RNase T1 digestion. (C) Validation of protein-hybrid interaction by reverse pull down. DDX1 and FUS and their associated hybrids were pulled down by anti-DDX1 and anti-FUS antibodies, respectively. RNA and DNA were purified from precipitates and quantified by qPCR using primers annealing to *DPP9* or *BAMBI* hybrids. Input amount was normalized against copy numbers of DNA and transcripts. (Error bars) SEM of triplicates. (D) Colocalization of SRSF1 binding with R-loops. SRSF1 binding to the *ACIN1* transcript was identified by PAR-CLIP and RNA-IP using an anti-SRSF1 antibody, and R-loops were identified by S9.6 DRIP. IGV viewer screenshots of data showing sequence reads from PAR-CLIP, RNA-IP, and DRIP-seq.

R-loops immunoprecipitated from HeLa cells with the S9.6 antibody (Cristini et al. 2018). Although their study was performed using HeLa cells and our study was carried out with human B-cell extracts, 197 of the proteins identified in their study were also found in our study (Supplemental Table S1). This provides evidence that the hybrid-binding proteins we identified interact with R-loops in vivo. Second, we validated the hybrid-protein interaction by reverse immunoprecipitation. Using antibodies specific for DDX1 and FUS, we pulled down the protein-nucleic acid complexes, then by quantitative PCR, we showed enrichment of DNA and RNA corresponding to the *BAMBI* and *DPP9* hybrids (Fig. 2C). Thus, the results validate that DDX and FUS bind in vivo to *BAMBI* and *DPP9* hybrids. Third, we assessed globally the binding of SRSF1, one of the hybrid-binding proteins, to R-loops in vivo. SRSF1 is a member of the serine/arginine-rich splicing factors that binds to exon-splicing enhancers (Pandit et al. 2013). We

identified transcripts bound by SRSF1 using two independent methods: PAR-CLIP and RNA-IP. Then we characterized R-loop regions in human cells using DRIP-seq with the S9.6 antibody. We carried out this experiment to assess the number of R-loops with which these hybrid-binding proteins interact. Given the large number of hybrid-binding proteins, each can be interacting with a few or many R-loops. Here, we began by addressing one protein. The results showed that SRSF1 binds to *BAMBI*, *DPP9*, and >20% of R-loops in human B-cells. Figure 2D shows an example of the colocalization of R-loops and SRSF1 binding sites in *ACIN1*. Together, these results support that the hybrid-binding proteins we identified bind to many R-loops in human cells, in addition to the *BAMBI* and *DPP9* hybrids.

The hybrid-binding proteins, such as FUS, matrin 3, and ligase 3, have significant enrichment of domains that bind nucleic acids and participate in a broad spectrum of gene regulation. A search of

Table 1. Examples of proteins that bind to RNA/DNA hybrids

Protein	Description	Number of peptides: <i>BAMBI</i> hybrid	Number of peptides: <i>DPP9</i> hybrid
DDX5	Probable ATP-dependent RNA helicase DDX5	36	18
FUS	RNA-binding protein FUS	14	9
HNRNPM	Heterogeneous nuclear ribonucleoprotein M	40	17
MATR3	Matrin 3	25	17
NCL	Nucleolin	31	25
NONO	Non-POU domain containing octamer binding	21	18
PSPC1	Paraspeckle component 1	7	5
RECQL	ATP-dependent DNA helicase Q1	26	27
RNASEH1	Ribonuclease H1	7	7
SFPQ	Splicing factor, proline- and glutamine-rich	25	24
SRSF1	Serine and arginine rich splicing factor 1	56	8
XRN1	5'-3' exoribonuclease 1	7	6
XRN2	5'-3' exoribonuclease 2	17	32

domains found in the hybrid-binding proteins reveals that 50 have alpha-beta plait and 27 contain OB-fold. A helicase domain, such as that in AQR that resolves hybrids, is also found (Sollner et al. 2014). Examples of the functional domains that are highly enriched in the hybrid-binding proteins are listed in Table 2. Among these 803 hybrid-binding proteins, 354 have disordered protein domains, including 59 proteins with [G/S]Y[G/S] amino acid motif that is hydrophobic and confers the proteins the ability to form hydrogels (Supplemental Table S1; Casas-Finet et al. 1993; Frey et al. 2006). Disordered regions provide protein flexibility in structure and function. In the hybrid-binding proteins, these regions likely allow the proteins to scan for target structures and interact with a range of other proteins and nucleic acid targets (Oldfield and Dunker 2014). The 803 hybrid-binding proteins cover a range of functions. Table 3 shows five functional categories that are highly enriched with hybrid-binding proteins. It shows that these proteins participate in multiple RNA processing steps, including splicing, pre-mRNA processing, and unwinding of RNA. The hybrid-binding proteins include PABPC1 (Kuhn et al. 2003) and CPSF1 (Murthy and Manley 1995) that bind poly(A) sequences, which is somewhat unexpected considering the absence of polyadenine track in our

hybrids. Therefore, these proteins may be recognizing structures that are shared by poly(A) RNA and the hybrids.

The *BAMBI* and *DPP9* hybrids pulled down several protein complexes. These include the *Drosophila* behavior/human splicing (DBHS) complex that comprises the non-POU domain containing octamer binding protein (NONO) and paraspeckle protein component 1 (PSPC1). DBHS proteins form heterodimer and oligomers with multiple domains for RNA binding (Passon et al. 2011, 2012). The resulting combinations likely provide surfaces that facilitate binding to RNA/DNA hybrids. Other protein complexes include RPA1, RNase H1 (Nguyen et al. 2017), and THRAP3/BCLAF1 (Vohhodina et al. 2017). Thus, our pull-down assays identify direct RNA-protein and indirect RNA-protein interactions mediated by protein-protein interactions.

Next, we studied the RNA/DNA hybrid-binding proteins to look for those that are repelled or attracted by hybrids relative to other nucleic acid structures. RNA/DNA hybrids are formed during transcription when nascent RNA hybridizes with their template DNA, thus disrupting the dsDNA. In a previous study, we showed that hybrid formation deters methylation-dependent gene silencing because DNA methyltransferase 1 binds less avidly to RNA/

Table 2. Domains represented in RNA/DNA hybrid-binding proteins

Domain	Proteins	FDR
Alpha-beta plait	SRSF1, RBM3, TRMT2A, U2AF2, U2SURP, RBM5, SYNCRIP, SLIRP, SART3, NONO, RAVR1, DDX21, PTBP1, G3BP2, NCL, TAF15, CPSF7, MATR3, FUS, SRSF10, IGF2BP1, NIFK, IGF2BP3, HNRNPL, HNRNPA3, HNRNPM, SAFB, HNRNPC, TNRC6B, RBM27, RBM25, EWSR1, RBM26, RPS24, HNRNPAB, ALYREF, ELAVL1, RPL23A, SSB, HNRNPDL, LARP4B, HNRNPA0, UPF3B, SRSF7, SFPQ, PSPC1, CIRBP, HNRNPH1, RBM14, RBM15	10 ⁻³⁰
P-loop triphosphate hydrolase	KIFC1, SEPT2, HBS1L, EIF5B, RNF213, MTHFD1L, DDX17, DDX18, DHX38, DDX23, DDX21, DHX36, DHX30, MX1, GNL3, MYO6, RAN, DDX39B, EIF2S3, EEF2, MYH9, HNRNPU, RFC5, RFC3, AQR, DHX29, ASCC3, SMARCAL1, SMARCA5, EEFSEC, SEPT7, SMARCA4, SEPT9, RAB7A, GNAI2, CTPS1, MOV10, DDX46, DDX3X, DDX41, HELLS, EEF1A1, DHX9, MYO1C, MYO1E, MYO1G, DDX1, YTHDC2, DDX5, DDX6, ARF4, SMC1A, DAP3, ABCF1, ABCF2, SKIV2L2, MTHFD1, LONP1, MCM7, DYNC1H1, KIF2A, SRPRB, MCM2, MCM3, MCM4, RAD50, MCM5, MCM6, EIF4A3, RECQL, EIF4A1, SNRNP200, RUVBL2, RUVBL1, MYO5A, TUFM, SKIV2L, TAP1, DHX15, RAB11A, DHX16, EHD1, CHD4, EHD4, UPF1, MSH3, MSH2, SMC6, DRG1, SMC3, SMC4, HNRNPUL2, SAMD9	10 ⁻²³
Nucleic acid-binding, OB-fold	NARS, PNPT1, YBX1, KARS, RPA1, MCM7, CSDE1, CCAR1, YARS, PDCD11, DARS, AIMP1, EXOSC2, LIG3, EXOSC3, MCM2, DIS3L, MCM3, MCM4, MCM5, MCM6, DIS3, EIF2S1, RUVBL2, RPS11, RUVBL1, SRBD1	10 ⁻¹⁷
DNA/RNA helicase, DEAD/DEAH box type, N-terminal	SKIV2L, SKIV2L2, DDX17, DDX46, DDX18, DDX23, DHX38, DDX3X, DHX15, DHX16, DDX21, DHX36, DDX41, DHX30, DHX9, DDX39B, DDX1, YTHDC2, DDX5, DDX6, EIF4A3, RECQL, DHX29, ASCC3, SNRNP200, EIF4A1	10 ⁻¹⁵
K homology domain	KHDRBS1, HDLBP, PNPT1, PNO1, SF1, EXOSC2, EXOSC3, IGF2BP1, ANKHD1, IGF2BP3, FXR1, FUBP1, ANKRD17, PCBP2, KHSRP	10 ⁻⁸

Table 3. Hybrid-binding proteins are enriched for various functions

Functional category	Fold enrichment	Benjamini-Hochberg P-value	Examples
RNA binding	6.3	10 ⁻⁷¹	DIS3, DKC1, ELAVL1, FUS, NCL, NPM1, UPF1, XPO1
mRNA splicing	7.4	10 ⁻³⁶	AQR, FRG1, GEMIN5, MBNL1, SRSF1, SRSF7, SRSF10, SFPQ
ATP-dependent helicase	9.3	10 ⁻¹⁵	DDX1, DDX5, DDX17, SKIV2L, YTHDC2
Termination of RNA polymerase II transcription	19	10 ⁻¹⁰	CPSF1, CPSF7, PCF11, U2AF2, UPF3B, ZC3H11A
Regulation of telomerase RNA localization to Cajal body	2	10 ⁻¹¹	DKC1, RUVBL1, RUVBL2

DNA hybrid than dsDNA (Grunseich et al. 2018). We assume that there are other proteins that are repelled by hybrids or attracted to them. To look for these proteins, we repeated the pull-down assays with dsDNA from the *BAMBI* and *DPP9* regions. We then compared the proteins that were pulled down by the nucleic acids of the same sequences but with different structures, that is, hybrids versus dsDNA. We found proteins like DNMT1 that bind more avidly to dsDNA than the RNA/DNA hybrids. Confirming our previous study that showed DNMT1 is repelled by hybrids, we found DNMT1 bound to both hybrids and the corresponding dsDNA, but dsDNA forms of *BAMBI* and *DPP9* pulled down more DNMT1 than their hybrids. With these results, we looked for similar binding patterns in other proteins and found 84 other candidates, such as PARP1 and UHRF1, that are repelled by hybrids (Supplemental Table S2). The ubiquitin transferase, UHRF1, interacts with and recruits DNMT1 (Sharif et al. 2007); thus, it could be repelled because it complexes with DNMT1. PARP1 can also be part of a complex with UHRF1 and DNMT1 (Sharif et al. 2007), and it regulates UHRF1's interaction with DNMT1 by addition of poly(ADP-ribose) (De Vos et al. 2014). In addition, PARP1 binds gene promoters and regulates transcription (Krishnakumar et al. 2008). Because hybrids disrupt binding of PARP1, their formation can affect PARP1 and lead to direct and indirect transcriptional consequences. Although PARP1 is mostly studied as a DNA-binding protein, a recent study showed that PARP1 binds RNA, in particular, those that are GC-rich (Melikishvili et al. 2017). Thus, it is possible that PARP1 recognizes non-B nucleic acid structures that are shared between GC-rich RNA and RNA/DNA hybrids.

In addition to proteins that are repelled by hybrids, there are proteins that are attracted by RNA/DNA hybrids. We found 364 proteins that are attracted by both *BAMBI* and *DPP9* hybrids compared to corresponding dsDNA (Supplemental Table S3). There are 14 proteins that bound to only the hybrid forms of *BAMBI* and *DPP9* but not to the corresponding dsDNA. These include several members of the nuclear exosomes such as DIS3L, EXOSC3, and EXOSC6. In addition, there are 350 proteins that favor the hybrid forms of *BAMBI* and *DPP9* compared to the corresponding dsDNA. We validated the affinity of some of these proteins to RNA/DNA hybrids versus dsDNA by biolayer interferometry. Figure 3A shows that DDX5, NONO, SPT5H, and RNase H1 (RNASEH1) have a higher affinity for RNA/DNA hybrids than the corresponding dsDNA. These hybrid-binding proteins are significantly enriched for K homology (KH) domains and RNA/DNA helicase. Immunostaining of human cells with S9.6 antibody confirmed colocalization of the nuclear RNA/DNA hybrids with proteins that were identified to be attracted by RNA/DNA hybrid, nucleolin (NCL), and DDX18 (Fig. 3B). There are 24 RNA/DNA helicases and 12 proteins with KH domains. Because the sequences of the *BAMBI* and *DPP9* hybrids are distinct, the proteins that bind to both of them likely recognize their hybrid structures rather than sequences.

Some of these hybrid-binding proteins, such as nucleolin (González et al. 2009; Haesler et al. 2014) and FUS (Takahama et al. 2013), were found in other studies to bind G-quadruplexes, which also are in non-B DNA conformations. However, not all proteins that bind G-quadruplexes bind RNA/DNA hybrids. Most likely, proteins can distinguish between different types of non-B structures; for example, TPM4 (von Hacht et al. 2014) and BLM helicase (Li et al. 2001; Chatterjee et al. 2014) that bind G-quadruplexes did not bind to either hybrid although they are expressed in our B-cell lysates.

Discussion

In this study, we identified more than 800 proteins that bound to RNA/DNA hybrids. R-loops are three-stranded structures that comprise an RNA/DNA hybrid and a displaced ssDNA. Here, we focused on the hybrid because the ssDNA has been the focus of many studies in the DNA repair field. We used two biotinylated hybrids corresponding to R-loops in the promoter of *BAMBI* and 3' UTR of *DPP9* to pull down proteins that were then identified by mass spectrometry. The resulting proteins include the well-characterized RNase H1 (RNASEH1) that is known to bind to hybrids, but most of the proteins were not known to interact with hybrids. We classified these hybrid-binding proteins into those (84) that are repelled by hybrids and those (364) that are attracted to hybrids relative to the underlying dsDNA. Although the hybrid-binding proteins were identified through in vitro studies, we provide evidence that they interact with R-loops in vivo.

Cellular functions rely on proteins and their interactions with each other and with nucleic acids. Although binding does not imply function, this set of hybrid-binding proteins helps to narrow down where one should focus further investigations. Such efforts are particularly useful at the beginning of studies in which preliminary results suggest many pathways could be involved and targeted analysis may miss critical pathways. Studies are elucidating the functions of R-loops, but much remains unknown. The 803 hybrid-binding proteins described in this paper suggest that proteins involved in RNA processing from splicing to unwinding RNA are involved in hybrid-mediated regulation. RNA/DNA hybrids are also key components in DNA replication as Okazaki fragments. Molecular studies of the hybrid-binding proteins identified in this study can elucidate how proteins divide their roles (or not) between transcription and DNA replication. These hybrid-binding proteins serve as a starting point for studying how proteins recognize RNA/DNA hybrids and the functional consequences of these interactions. Delineation of how these proteins interact with RNA/DNA hybrids in transcription, DNA replication, and other cellular processes will deepen our understanding of these crucial biological pathways.

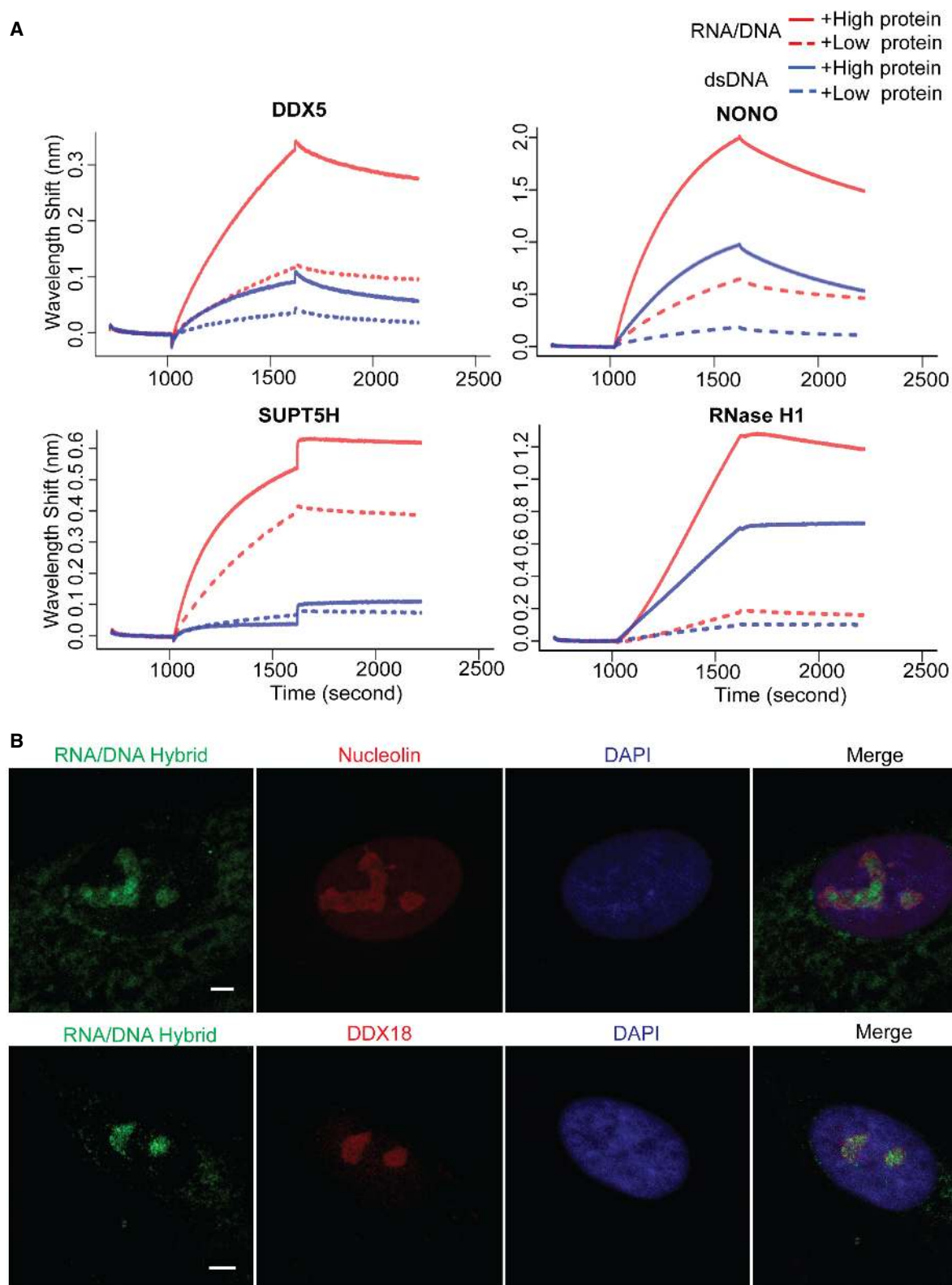


Figure 3. Validation of protein and hybrid interaction. (A) The protein and hybrid interaction is shown by biolayer interferometry. Binding of *DPP9* RNA/DNA hybrid (red lines) or dsDNA (blue lines) to high concentration (solid lines) or low concentration (dotted lines) of each protein was measured. These proteins showed more avid binding to RNA/DNA hybrid than to dsDNA. Baseline was recorded from 0 to 1000 sec, association of protein with dsDNA or RNA/DNA hybrid from 1000 to 1600 sec, followed by dissociation. (B) Immunofluorescence staining in primary human fibroblasts showing colocalization of nucleolin and DDX18 (red) with the R-loops stained by S9.6 RNA/DNA hybrid antibody (green). DAPI staining is in blue. (Scale bar) 1 μ m.

Methods

Cell culture

Immortalized B-cells (Coriell) were cultured to a density of 5×10^5 cells/mL in RPMI 1640 supplemented with 15% fetal bovine serum, 2 mM L-glutamine, and 100 units/mL penicillin-streptomycin.

Biotinylated RNA/DNA hybrids

Ninety-mer RNA and DNA oligos corresponding to *DPP9* 3' UTR (Chr 19: 4,675,244–4,723,855) were synthesized by Integrated DNA Technologies (Supplemental Table S4). Oligos were dissolved in Annealing Buffer (10 mM Tris at pH 8.0; 50 mM NaCl, 1 mM EDTA). To generate the RNA/DNA hybrid, 10 μ M of each oligo was mixed and heated for 5 min at 95°C and cooled down gradually to room temperature.

Six hundred-mer RNA/DNA hybrid corresponding to *BAMBI* promoter (Chr 10: 28,966,424–28,971,868) was generated as previously described (Grunseich et al. 2018). dsDNA was prepared by PCR using biotinylated primers (Supplemental Table S4). The 600-nt RNA transcript was synthesized from this *BAMBI* dsDNA template using MEGAscript T7 Transcription kit (Thermo Fisher Scientific, #AM1334). The T7 promoter sequence in dsDNA was then removed by SfcI (NEB, #R0561S) digestion. The dsDNA and the transcribed ssRNA were dissolved in 10 mM Tris HCl at pH 8.0, 1 mM EDTA, 50 mM NaCl, and incubated for 5 min at 95°C and slowly cooled down to room temperature. Reannealed dsDNA was removed by HpaII digestion (NEB, #R0171S), and the RNA/DNA hybrid was purified using agarose gel electrophoresis.

To confirm the integrity of RNA/DNA hybrids, each hybrid was digested with RNase H1 (a gift from Dr. Robert Crouch at the NIH) or RNase T1 (Thermo Fisher Scientific, #EN0541) for 1 h at 37°C, extracted with phenol/chloroform, and precipitated with ethanol. The digested hybrid was analyzed by agarose gel electrophoresis and used in protein precipitation experiments.

Hybrid-binding protein precipitation and Western blot

Cultured B-cells were lysed in lysis buffer (20 mM Tris HCl at pH 8, 137 mM NaCl, 10% glycerol, 1% NP-40, and 2 mM EDTA) supplemented with 1 \times Complete protease inhibitors (Roche), 1 \times phosphatase inhibitors II and III (Sigma), and 0.1 unit RNase inhibitor (Thermo Fisher Scientific). Cell lysates were precleared for 2 h at 4°C using streptavidin beads (Thermo Fisher Scientific, #65305). Thirty picomoles of biotinylated RNA/DNA hybrid or dsDNA were conjugated with streptavidin beads and incubated with precleared lysates containing 240 μ g total protein for 2 h to overnight at 4°C. Protein-nucleotide complexes were pulled down with streptavidin beads and washed three times in 20 mM Tris at pH 7.5, 10 mM NaCl, 0.1% Tween-20. Proteins were eluted in 1 \times LDS sample buffer (Thermo Fisher Scientific, #NP0007) containing 1 \times sample reducing reagent (Thermo Fisher Scientific, #NP0004) for 5 min at 95°C.

Hybrid-binding protein was validated by Western blot using the following antibodies: anti-NONO (Novus, #NB100-1556), anti-NPM1 (Cell Signaling, #3542), anti-RECQL (Novus, #NB100-619), and anti-GAPDH (Santa Cruz, #sc-25778).

Mass spectrometry analysis

Liquid chromatography–tandem mass spectrometry (LC-MS/MS) analysis was performed. Each eluted sample from protein precipitation was divided into five fractionated by one-dimensional SDS-PAGE. Each fraction was digested in gel, and tryptic peptides were injected onto a UPLC Symmetry trap column (180 μ m i.d. \times

2 cm packed with 5 μ m C18 resin; Waters). A blank gel slice was digested and injected as a background control. Tryptic peptides were separated by reversed phase HPLC on a nanocapillary analytical column (75 μ m i.d. \times 25 cm, 1.7 μ m particle size; Waters). Eluted peptides were analyzed on a Q Exactive HF mass spectrometer (Thermo Fisher Scientific).

MS/MS spectra were searched against the UniProt human database (The UniProt Consortium 2017) with the MaxQuant 1.5.2.8 program (Cox and Mann 2008) using full tryptic specificity allowing up to two missed cleavages. Search parameters include static carboxamidomethylation of Cys, variable protein N-terminal acetylation, and variable Met oxidation. For statistical analysis, we carried out a decoy database search to determine the false discovery rate (Elias and Gygi 2010). FDR for both protein and peptide identifications was set at <1%. In the input B-cell lysate, we identified 22,440 unique peptides.

To identify proteins that specifically bind to hybrids or dsDNA in the pull-down experiments, we required at least four peptides with unique sequences from a given protein to be included. Using this criterion, 1460 and 1018 proteins were pulled down by *BAMBI* and *DPP9* hybrids, respectively, and 1092 and 995 proteins were pulled down by *BAMBI* and *DPP9* dsDNA, respectively. We focused on proteins that were reproducibly pulled down by both *BAMBI* and *DPP9* in downstream analysis. To exclude the possibility that the hybrid-binding proteins are pulled down through nonspecific binding to nucleic acids, we pre-incubated B-cell extract with biotinylated ssRNA and dsDNA sharing the same sequence with *BAMBI* hybrid and depleted nonspecific proteins using streptavidin beads. We then repeated pull down and LC-MS/MS using *BAMBI* hybrid and showed that all hybrid-binding proteins were still specifically pulled down. Fold enrichment of proteins in pull down is calculated as the ratio of MS/MS counts of a detected protein from two samples. We set the threshold of fold enrichment ≥ 1.2 for a protein to be considered enriched.

RNA/DNA hybrid immunoprecipitation with qPCR (DRIP-PCR) and sequencing (DRIP-seq)

DRIP was adapted from a previous report (Skourti-Stathaki et al. 2011) with modification. Cultured B-cells (5×10^6) were lysed in 400 μ L cell lysis buffer (50 mM PIPES at pH 8.0, 100 mM KCl, 0.5% NP-40), and nuclei were collected by centrifugation. The nuclei pellet was resuspended in 200 μ L nuclear lysis buffer (25 mM Tris HCl at pH 8.0, 1% SDS, 5 mM EDTA). Genomic DNA containing R-loops was then extracted by phenol:chloroform and precipitated by ethanol. Purified material was resuspended in 200 μ L IP dilution buffer (16.7 mM Tris HCl at pH 8.0, 1 mM EDTA, 0.01% SDS, 1% Triton X-100, 167 mM NaCl) and sonicated at 4°C in Bioruptor (Diagenode) at Hi setting (30 sec on/30 sec off) for 5 min, three times, to fragments with an average size of 500 bp. Three micrograms of S9.6 monoclonal antibody (a gift from Dr. Stephen H. Leppla at NIH) or nonspecific mouse IgG (Santa Cruz, #sc-2025) was used for each immunoprecipitation. Input and precipitates were analyzed by quantitative PCR using specific primers (Supplemental Table S4).

DRIP-seq libraries were prepared from DRIP DNA and corresponding input DNA using Ovation Ultralow System (NuGen) and sequenced on an Illumina HiSeq 2500 platform. An average of 40 million 100-nt reads per sample was generated. Sequencing reads were preprocessed to remove adapter sequences from the end of reads using the program `fastx_clipper` from FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Low-quality sequences at the ends of reads represented by stretches of “#” in the quality score string in FASTQ file were also removed. Reads shorter than 35 nt after trimming were excluded from analysis. Sequencing reads

were aligned to human reference (hg19) using GSNAP (version 2013-10-28) (Wu and Nacu 2010) using the following parameters: mismatches % [(read length + 2)/12-2]; mapping score R 20; soft-clipping on (-trim-mismatch-score=-3). Reads with identical sequences were compressed into one unique sequence. R-loop peaks were identified using MACS2 (Zhang et al. 2008) and required to have \geq twofold enrichment in DRIP over input. We identified 2636 R-loop peaks, among which 1438 peaks reside in 743 genes.

SRSF1 PAR-CLIP

PAR-CLIP was performed as previously described with minor modifications (Hafner et al. 2010). Cultured human B-cells were treated with 1 mM 4-thiouridine (Sigma-Aldrich) and UV crosslinked at 312 nm for 5 min. The cells were collected, washed in 1 \times PBS, and fractionated into cytoplasmic and nuclear fractions. The lysate was treated with 1 unit/ μ L of RNase T1 for 15 min at 22°C. SRSF1 was immunoprecipitated with the anti-SRSF1 antibody (ABCAM, #ab38107). The beads with precipitates were washed three times with NP-40 lysis buffer and subsequently treated with 10 units/ μ L of RNase T1 for 15 min at 22°C. The beads were washed again three times in NP-40 lysis buffer and dephosphorylated with 0.5 unit/mL CIP alkaline phosphatase. The immunoprecipitated material was treated with 0.5 μ Ci γ -³²P-ATP and 1 unit/ μ L of T4 PNK kinase for 30 min at 37°C. Beads were washed five times with PNK wash buffer (50 mM Tris HCl at pH 7.5, 50 mM NaCl, 10 mM MgCl₂) and resuspended in 100 μ L 2 \times sample buffer and separated on a 4%–12% SDS-PAGE and transferred to a nitrocellulose membrane. The SRSF1 ribonucleoprotein complex was visualized by autoradiography, and the band corresponding to SRSF1 was isolated. RNA was extracted by Proteinase K digestion, purified by phenol-chloroform extraction, and precipitated with three volumes of ethanol. The purified RNA from each cellular fraction was ligated with a unique 3' adapter with Rnl2(1–249) K227Q ligase (NEB) overnight at 4°C. The RNA was loaded onto a 15% Urea-PAGE, and the ligated RNA cut out and extracted from the gel with 400 μ L 0.3M NaCl for 45 min at 60°C with vigorous shaking. The gel pieces were filtered away and RNA in the flow-through precipitated with three volumes of ethanol. The RNA pellet was dissolved in water and ligated with 5' adapter using Rnl1 ligase (NEB) for 1 h at 37°C. The RNA was loaded onto a 12% Urea-PAGE, and the ligated RNA cut out and extracted from the gel with 400 μ L 0.3 M NaCl for 45 min at 60°C with vigorous shaking. The gel pieces were filtered out and RNA in the flow-through precipitated with three volumes of ethanol. The RNA was reverse transcribed using SuperScript III reverse transcriptase (Thermo Fisher Scientific) with 3' RT primer for 2 h at 50°C, according to the manufacturer's instructions. Next, the generated cDNA was PCR amplified using Taq DNA polymerase (Thermo Fisher Scientific). The primers used for PAR-CLIP are listed in Supplemental Table S4. The PCR band corresponding to the correct size of amplification (143–153 bp) was purified using a 3% PippinPrep gel according to the manufacturer's instructions and quantified. PAR-CLIP cDNA libraries were sequenced on an Illumina HiSeq 3000 instrument. Clusters of overlapping reads uniquely mapped to the human genome hg19 were generated using the PARalyzer software (Corcoran et al. 2011), allowing for one mismatch and otherwise default settings. Clusters were annotated against the following GENCODE gtf file: GENCODE.v19.chr_patch_hapl_scaff.annotation.gtf (<http://www.genencodegenes.org>) (Harrow et al. 2012). The hg19 assembly was used. The main differences between hg19 and the more current GRCh38 is that GRCh38 contains more alternatively spliced sequences, centromeric regions, and the mitochondria genome. Because these sequences were not the focus of our study, the use of hg19 is unlikely to have affected our conclusions.

Reverse protein-RNA immunoprecipitation and sequencing

Reverse immunoprecipitation was carried out using Magna RNA-Binding Protein Immunoprecipitation Kit (Millipore) following the manufacturer's protocol. Briefly, for each immunoprecipitation reaction, 2 \times 10⁷ cultured human B-cells or primary skin fibroblasts were harvested and lysed in 100 μ L lysis Buffer with protease and RNase inhibitors. Five micrograms of anti-SRSF1 antibody (ABCAM, #ab38107), anti-FUS antibody (Novus, #NB100-561), and anti-DDX1 (ABCAM, #ab70252), negative control IgG (Millipore, #12-371 for mouse IgG; #12-370 for rabbit IgG) were conjugated to Magnetic Protein A/G beads. One hundred microliters of cell lysate was added into 900 μ L Immunoprecipitation Buffer with RNase inhibitor and incubated with 50 μ L beads-antibody complex overnight at 4°C. Bead-bound immunoprecipitates were then washed six times using cold Wash Buffer with RNase inhibitor and incubated with protease K in the presence of 1% SDS for 30 min at 55°C. RNA and DNA were then extracted from supernatants using phenol:chloroform:isoamyl alcohol and precipitated using ethanol. Precipitated RNA was digested by DNase I (DNA-free kit, Ambion). cDNA was synthesized using random hexamer primer by TaqMan Reverse Transcription Reagent kit (Applied Biosystems). Quantitative PCR was carried out to quantify cDNA and DNA with primers annealing to *BAMBI* and *DPP9* hybrids using Power SYBR Green PCR Master Mix (Thermo Fisher Scientific, #4367659). Primer sequences are listed in Supplemental Table S4. RNA from anti-SRSF1 immunoprecipitate and input RNA were prepared into RNA-seq libraries using Illumina TruSeq Stranded Total RNA Library Prep kit (Illumina, #20020596) and sequenced on HiSeq 2500. Sequencing reads were preprocessed and aligned as described above. Enrichment of transcripts in the immunoprecipitate was analyzed using the program described by Antanaviciute et al. (2017). Transcripts with fold enrichment >2 by anti-SRSF1 antibody are considered SRSF1-binding targets.

Biolayer interferometry

Analysis of dsDNA or RNA/DNA hybrid binding to candidate proteins was carried out using the Octet RED96 system (ForteBio) with sensor detection of the change in wavelength (nm shift). Purified candidate proteins DDX5 (Abnova, #TP300371), NONO (Abnova, #TP326567), SUPT5H (Abnova, #TP326321), and human RNase H1 (RNASEH1; a gift from Dr. Robert Crouch at the NIH) were evaluated. Biotinylated dsDNA or RNA/DNA hybrid at concentrations of 5 nM was immobilized onto a Streptavidin-SA biosensor. The biotinylated *DPP9* dsDNA 90-mer was generated as described above. dsDNA and RNA/DNA hybrid were loaded onto the sensors until saturation. The nucleotide-labeled sensors were then washed with buffer, followed by addition of DDX5 at concentrations of 2 and 8 μ g/mL, NONO at concentrations of 1 and 4 μ g/mL, SUPT5H at concentrations of 2 and 8 μ g/mL, and RNase H1 (RNASEH1) at 3.8 and 19 nM. All reactions were tested in TBS buffer (10 mM Tris at pH 7.4, 68 mM NaCl, 0.02% Tween-20). A reference sample of buffer and protein alone did not show any signal drift. Association and dissociation were monitored for 10 min each. All experiments were conducted in the Octet instrument with agitation at 1000 rpm.

Immunofluorescence

Fibroblasts were fixed with 4% paraformaldehyde for 15 min at room temperature, then washed three times with phosphate-buffered saline (PBS). Slides were then placed in blocking solution (5% normal goat serum, 0.3% Triton X-100 in PBS) for 1 h at room temperature. Primary antibody staining was done overnight at 4°C in PBS with 1% bovine serum albumin and 0.3% Triton

X-100 using 1:500 S9.6 antibody, 1:500 DDX18 antibody (ABCAM, #ab128197), or 1:500 nucleolin antibody (ABCAM, #ab22758). Slides were then washed three times with PBS for 5 min each, incubated with 1:500 secondary antibody (Invitrogen, #A-31572 for anti-rabbit and #A-11001 for anti-mouse) for 2 h at room temperature in the dark, and then washed three times with PBS for 5 min each before DAPI nuclear staining. Imaging was performed with a Leica DMI 6000CS laser confocal microscope with a Leica HCX PL APO 63× oil objective.

Intrinsically disordered region

IUPred (Dosztányi et al. 2005) was used to predict disordered regions in the protein. The 354 proteins that we included as containing disordered regions are those with at least 30% of the proteins with IUPred score >0.4 (Hentze et al. 2018). In “long” (global) disorder mode, a sequential neighborhood of 100 residues is considered in calculating the score, whereas in “short” (local) disorder mode, a sequential neighborhood of 25 residues is considered.

Data access

The PAR-CLIP, RNA-IP-seq, and DRIP-seq data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE117671. The mass spectrometry data from this study have been submitted to the PeptideAtlas database (<http://www.peptideatlas.org/>) with the identifier PASS01169.

Acknowledgments

We thank Dr. Robert Crouch for his advice and insightful discussions of R-loop; Dr. Stephan Leppa for the S9.6 antibody; and Dr. Grzegorz Piszczek for his assistance in performing the bilayer interferometry experiments. This work was supported by the Howard Hughes Medical Institute and the National Institute of Neurological Disorders and Stroke (NIH).

Author contributions: I.X.W., C.G., and V.G.C. designed the experiments and interpreted the results. I.X.W., C.G., J.F., N.R., and M.H. performed the experiments. J.B., Z.Z., and M.H. analyzed deep sequencing data. I.X.W., C.G., J.F., and V.G.C. wrote the manuscript.

References

Abdouni HS, King JJ, Ghorbani A, Fifield H, Berghuis L, Larjani M. 2018. DNA/RNA hybrid substrates modulate the catalytic activity of purified AID. *Mol Immunol* **93**: 94–106.

Antanaviciute A, Baquero-Perez B, Watson CM, Harrison SM, Lascelles C, Crinnion L, Markham AF, Bonthron DT, Whitehouse A, Carr IM. 2017. m6aViewer: software for the detection, analysis, and visualization of N⁶-methyladenosine peaks from m⁶A-seq/ME-RIP sequencing data. *RNA* **23**: 1493–1501.

Baker TA, Kornberg A. 1988. Transcriptional activation of initiation of replication from the *E. coli* chromosomal origin: an RNA-DNA hybrid near *oriC*. *Cell* **55**: 113–123.

Baltz AG, 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.

Benevides JM, Wang AH, Rich A, Kyogoku Y, van der Marel GA, van Boom JH, Thomas GJ. 1986. Raman spectra of single crystals of r(CGCG)(d(CGC) and d(CCCCGGGG) as models for A DNA, their structure transitions in aqueous solution, and comparison with double-helical poly(dG)-poly(dC). *Biochemistry (Mosc)* **25**: 41–50.

Bernecky C, Herzog F, Baumeister W, Plitzko JM, Cramer P. 2016. Structure of transcribing mammalian RNA polymerase II. *Nature* **529**: 551–554.

Bhattacharyya A, Murchie AI, Lilley DM. 1990. RNA bulges and the helical periodicity of double-stranded RNA. *Nature* **343**: 484–487.

Casas-Finet JR, Smith JD, Kumar A, Kim JG, Wilson SH, Karpel RL. 1993. Mammalian heterogeneous ribonucleoprotein A1 and its constituent domains: nucleic acid interaction, structural stability and self-association. *J Mol Biol* **229**: 873–889.

Cerritelli SM, Crouch RJ. 1995. The non-RNase H domain of *Saccharomyces cerevisiae* RNase H1 binds double-stranded RNA: Magnesium modulates the switch between double-stranded RNA binding and RNase H activity. *RNA* **1**: 246–259.

Cerritelli SM, Crouch RJ. 2009. Ribonuclease H: the enzymes in eukaryotes. *FEBS J* **276**: 1494–1505.

Chan YA, Aristizabal MJ, Lu PYT, Luo Z, Hamza A, Kobor MS, Stirling PC, Hieter P. 2014. Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. *PLoS Genet* **10**: e1004288.

Chatterjee S, Zagalbaum J, Savitsky P, Sturzenegger A, Huttner D, Janscak P, Hickson ID, Gileadi O, Rothenberg E. 2014. Mechanistic insight into the interaction of BLM helicase with intra-strand G-quadruplex structures. *Nat Commun* **5**: 5556.

Chen L, Chen JY, Zhang X, Gu Y, Xiao R, Shao C, Tang P, Qian H, Luo D, Li H, et al. 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.

Corcoran DL, Georgiev S, Mukherjee N, Gottwein E, Skalsky RL, Keene JD, Ohler U. 2011. PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol* **12**: R79.

Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**: 1367–1372.

Cristini A, Groh M, Kristiansen MS, 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.

De Vos M, El Ramy R, Quénet D, Wolf P, Spada F, Magroun N, Babbio F, Schreiber V, Leonhardt H, Bonapace IM, et al. 2014. Poly(ADP-ribose) polymerase 1 (PARP1) associates with E3 ubiquitin-protein ligase UHRF1 and modulates UHRF1 biological functions. *J Biol Chem* **289**: 16223–16238.

Dosztányi Z, Csizmok V, Tompa P, Simon I. 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**: 3433–3434.

El Hage A, Webb S, Kerr A, Tollervey D. 2014. Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genet* **10**: e1004716.

Elias JE, Gygi SP. 2010. Target-decoy search strategy for mass spectrometry-based proteomics. *Methods Mol Biol* **604**: 55–71.

Fedoroff OY, Salazar M, Reid BR. 1993. Structure of a DNA:RNA hybrid duplex: why RNase H does not cleave pure RNA. *J Mol Biol* **233**: 509–523.

Figiel M, Nowotny M. 2014. Crystal structure of RNase H3–substrate complex reveals parallel evolution of RNA/DNA hybrid recognition. *Nucleic Acids Res* **42**: 9285–9294.

Frey S, Richter RP, Görlich D. 2006. FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science* **314**: 815–817.

Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F. 2012. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell* **45**: 814–825.

González V, Guo K, Hurley L, Sun D. 2009. Identification and characterization of nucleolin as a *c-myc* G-quadruplex-binding protein. *J Biol Chem* **284**: 23622–23635.

Groh M, Lufino MMP, Wade-Martins R, Gromak N. 2014. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genet* **10**: e1004318.

Grunseich C, Wang IX, Watts JA, Burdick JT, Guber RD, Zhu Z, Bruzel A, Lanman T, Chen K, Schindler AB, et al. 2018. Senataxin mutation reveals how R-loops promote transcription by blocking DNA methylation at gene promoters. *Mol Cell* **69**: 426–437.e7.

Haeusler AR, Donnelly CJ, Periz G, Simko EAJ, Shaw PG, Kim M-S, Maragakis NJ, Troncoso JC, Pandey A, Sattler R, et al. 2014. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507**: 195–200.

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

Hanna MM, Meares CF. 1983. Topography of transcription: path of the leading end of nascent RNA through the *Escherichia coli* transcription complex. *Proc Natl Acad Sci* **80**: 4238–4242.

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, et al. 2012. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**: 1760–1774.

Hentze MW, Castello A, Schwarzl T, Preiss T. 2018. A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol* **19**: 327–341.

- Huertas P, Aguilera A. 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell* **12**: 711–721.
- Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, Nogales E, Doudna JA. 2016. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**: 867–871.
- Keller W, Crouch R. 1972. Degradation of DNA RNA hybrids by ribonuclease H and DNA polymerases of cellular and viral origin. *Proc Natl Acad Sci* **69**: 3360–3364.
- Krishnakumar R, Gamble MJ, Frizzell KM, Berrocal JG, Kininis M, Kraus WL. 2008. Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes. *Science* **319**: 819–821.
- Kuhn U, Nemeth A, Meyer S, Wahle E. 2003. The RNA binding domains of the nuclear poly(A)-binding protein. *J Biol Chem* **278**: 16916–16925.
- Li X, Manley JL. 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**: 365–378.
- Li JL, Harrison RJ, Reszka AP, Brosh RM, Bohr VA, Neidle S, Hickson ID. 2001. Inhibition of the Bloom's and Werner's syndrome helicases by G-quadruplex interacting ligands. *Biochemistry* **40**: 15194–15202.
- Melikishvili M, Chariker JH, Rouchka EC, Fondufe-Mittendorf YN. 2017. Transcriptome-wide identification of the RNA-binding landscape of the chromatin-associated protein PARP1 reveals functions in RNA biogenesis. *Cell Discov* **3**: 17043.
- Mischo HE, Gómez-González B, Grzechnik P, Rondón AG, Wei W, Steinmetz L, Aguilera A, Proudfoot NJ. 2011. Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol Cell* **41**: 21–32.
- Murthy KG, Manley JL. 1995. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes Dev* **9**: 2672–2683.
- Nguyen HD, Yadav T, Giri S, Saez B, Graubert TA, Zou L. 2017. Functions of replication protein A as a sensor of R loops and a regulator of RNaseH1. *Mol Cell* **65**: 832–847.e4.
- Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata S, Dohmae N, Ishitani R, Zhang F, Nureki O. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**: 935–949.
- Nudler E, Mustaev A, Lukhtanov E, Goldfarb A. 1997. The RNA–DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* **89**: 33–41.
- Oldfield CJ, Dunker AK. 2014. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem* **83**: 553–584.
- Panda AC, Abdelmohsen K, Yoon JH, Martindale JL, Yang X, Curtis J, Mercken EM, Chenette DM, Zhang Y, Schneider RJ, et al. 2014. RNA-binding protein AUF1 promotes myogenesis by regulating MEF2C expression levels. *Mol Cell Biol* **34**: 3106–3119.
- Pandit S, Zhou Y, Shiue L, Coutinho-Mansfield G, Li H, Qiu J, Huang J, Yeo GW, Ares M, Fu X-D. 2013. Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Mol Cell* **50**: 223–235.
- Passon DM, Lee M, Fox AH, Bond CS. 2011. Crystallization of a paraspeckle protein PSPC1–NONO heterodimer. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **67**: 1231–1234.
- Passon DM, Lee M, Rackham O, Stanley WA, Sadowska A, Filipovska A, Fox AH, Bond CS. 2012. Structure of the heterodimer of human NONO and paraspeckle protein component 1 and analysis of its role in subnuclear body formation. *Proc Natl Acad Sci* **109**: 4846–4850.
- Roberts RW, Crothers DM. 1992. Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science* **258**: 1463–1466.
- Roy D, Lieber MR. 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.
- Rychlik MP, Chon H, Cerritelli SM, Klimek P, Crouch RJ, Nowotny M. 2010. Crystal structures of RNase H2 in complex with nucleic acid reveal the mechanism of RNA–DNA junction recognition and cleavage. *Mol Cell* **40**: 658–670.
- Sarkar K, Han SS, Wen KK, Ochs HD, Dupré L, Seidman MM, Vyas YM. 2017. R-loops cause genomic instability in Wiskott–Aldrich syndrome T helper lymphocytes. *J Allergy Clin Immunol* **142**: 219–234.
- Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, Shinga J, Mizutani-Koseki Y, Toyoda T, Okamura K, et al. 2007. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450**: 908–912.
- Skourti-Stathaki K, Proudfoot NJ, Gromak N. 2011. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell* **42**: 794–805.
- Sollier J, Stork CT, García-Rubio ML, Paulsen RD, Aguilera A, Cimprich KA. 2014. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol Cell* **56**: 777–785.
- Stein H, 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.
- Takahama K, Takada A, Tada S, Shimizu M, Sayama K, Kurokawa R, Oyoshi T. 2013. Regulation of telomere length by G-quadruplex telomere DNA- and TERRA-binding protein TLS/FUS. *Chem Biol* **20**: 341–350.
- The UniProt Consortium. 2017. UniProt: the universal protein knowledge-base. *Nucleic Acids Res* **45**: D158–D169.
- Vohhodina J, Barros EM, Savage AL, Liberante FG, Manti L, Bankhead P, Cosgrove N, Madden AF, Harkin DP, Savage KI. 2017. The RNA processing factors THRAP3 and BCLAF1 promote the DNA damage response through selective mRNA splicing and nuclear export. *Nucleic Acids Res* **45**: 12816–12833.
- von Hacht A, Seifert O, Menger M, Schütze T, Arora A, Konthur Z, Neubauer P, Wagner A, Weise C, Kurreck J. 2014. Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic Acids Res* **42**: 6630–6644.
- Wahba L, Amon JD, Koshland D, Vuica-Ross M. 2011. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol Cell* **44**: 978–988.
- Wahba L, Costantino L, Tan FJ, Zimmer A, Koshland D. 2016. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev* **30**: 1327–1338.
- Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**: 873–881.
- Xu B, Clayton DA. 1996. RNA–DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA–DNA hybrids serving as primers. *EMBO J* **15**: 3135–3143.
- Yu K, Chedin F, Hsieh CL, Wilson TE, Lieber MR. 2003. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat Immunol* **4**: 442–451.
- Zhang Y, Liu T, Meyer CA, Eickhout J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.
- Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE, Borowsky M, Lee JT. 2010. Genome-wide identification of Polycomb-associated RNAs by RIP-seq. *Mol Cell* **40**: 939–953.
- Zuo Y, Deutscher MP. 2002. The physiological role of RNase T can be explained by its unusual substrate specificity. *J Biol Chem* **277**: 29654–29661.

Received March 20, 2018; accepted in revised form July 27, 2018.