Supplemental Material

HNRNPH1-dependent splicing of a fusion oncogene reveals a targetable RNA G-quadruplex interaction

Carla Neckles[1], Robert Boer[2], Nicholas Aboreden[1], Allison M. Cross[1], Robert L. Walker[3], Bong-Hyun Kim[4], Suntae Kim[1], John S. Schneekloth Jr.[2], Natasha J. Caplen[1*]

[1] Functional Genetics Section, Genetics Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA

[2] Chemical Biology Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702, USA

[3] Molecular Genetics Section, Genetics Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA

[4] CCR Collaborative Bioinformatics Resource, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick, MD, 21702, USA

* To whom correspondence should be addressed. Tel: 240-760-7366; Fax: 240-541-4481; Email: ncaplen@mail.nih.gov

Present Address:

Bong-Hyun Kim, Bioinformatics Division, Kite, a Gilead Company, Santa Monica, CA 90404, USA

Suntae Kim, Clinical Operations, KCRN Research, LLC, Germantown, MD 20874, USA

METHODS

Protein expression and purification by IMAC

Briefly, a glycerol stock of BL21(DE3) Star pRare *E. coli* cells transformed with appropriate plasmid was used to inoculate 3 ml of MDAG-135 medium in a 15-ml snap-cap polypropylene tube (Studier 2005). The seed culture was grown for 16 hours, shaking at 250 rpm and 37°C. A 500-ml baffled shake flask containing 100 ml Dynamite broth was inoculated with 2 ml of the seed (1:50 dilution of the production culture volume). The culture was grown in a 37°C shaker at 250 rpm to an OD_{600} of 6.0-7.0 before shifting the temperature to 16°C. IPTG was added to a final concentration of 0.5 mM and the cultures were grown overnight for a total growth time of 24 hours after inoculation. The culture was centrifuged at 4,000 x *g* (Beckman JS-5.3 rotor) for 20 minutes at 4°C and the pellets were frozen at -80°C.

E. coli cell pellet was resuspended in a lysis buffer of 25 mM HEPES, pH 7.3, 375 mM NaCl, 1.25 mM TCEP, and Complete protease inhibitor (Sigma-Aldrich, St. Louis, MO) as per the manufacturer's instructions. The volume of lysis buffer was calculated based on achieving a final concentration of 100 OD units/ml. Each sample was lysed by a microfluidizer (Microfluidics, Newton, MA) and clarified at 100,000 x g for 30 minutes. The clarified supernatant was frozen at -80°C until purification. The clarified lysate was thawed at room temperature, filtered (0.45 micron), and adjusted to contain 50 mM imidazole. Using a Bio-Rad NGC FPLC, the sample was applied to a 2-mL HisTrap column (GE Healthcare, Chicago, IL) that had been pre-equilibrated in 20 mM HEPES, pH 7.3, 300 mM NaCl, 1 mM TCEP, 50 mM imidazole. The column was washed with the equilibration buffer to baseline. Proteins were eluted with a linear elution gradient to 500 mM imidazole over 20 column volumes, and the fractions analyzed by SDS-PAGE. Pooled fractions were dialyzed twice for at least 4 hours at 4°C against at least 20 sample volumes of 20 mM HEPES, pH 7.3, 300 mM NaCl, 1 mM TCEP using 3.5 kDa MWCO Snakeskin dialysis membrane (Pierce).

CRISPR-Cas9-mediated mCherry tagging of *NR0B1* in TC32 cells

Design of sgRNA and donor plasmid. A sgRNA targeting the C-terminus of *NR0B1* (ACTGCACTACTGCACTTGTG) was designed using Benchling software (San Francisco, CA) and cloned into the Cas9-containing plasmid PX459 (plasmid #62988, Addgene, Cambridge, MA). Forward and reverse oligonucleotides (**Table S3**) containing the sgRNA sequence were

phosphorylated using T4 PNK (New England Biolabs, Ipswich, MA), annealed, and ligated into a *BbsI* digested PX459 plasmid using Quick Ligase (New England Biolabs).

The NR0B1-T2A-mCherry donor vector was generated by insertion of the left homology arm (X:30322699-30323598), right homology arm (X:30321796-30322695), and T2A-mCherry cassette into the pUC19 vector (plasmid #50005, Addgene) using Gibson Assembly® (New England Biolabs). Briefly, for generation of the left and right homology arms, TC32 genomic DNA was isolated using the Maxwell® 16 Cell LEV DNA purification kit (Promega, Madison, WI) and 900 bp fragments upstream and downstream of the stop codon in *NR0B1* exon 2 were amplified using two sequential PCR reactions. To allow for linearization of the donor vector after co-transfection with the CRISPR-Cas9 plasmid (Zhang et al. 2017), the primers used in the initial round of amplification contained adapters for the sgRNA sequence at the 5' end of the left homology arm and the 3' end of the right homology arm (**Table S3**). Primers containing 30 bp of the overlapping sequence to *HincII*-digested pUC19 were used in the second round of amplification and are listed in **Table S3**. Similarly, the T2A-mCherry cassette was amplified from the Cas9-T2A-mCherry plasmid (plasmid #64324, Addgene) using primers containing 30 bp overlap to the adjacent homology arm and are listed in **Table S3**. All PCR fragments were amplified using Platinum™ PCR SuperMix High Fidelity (Invitrogen) and gel purified (Qiagen). Plasmids were sequenced using the Sanger method (CCR Genomics Core, CCR, NCI).

Generation of TC32 cells expressing mCherry tagged NR0B1. TC32 cells (6 x 10^5) were resuspended with 4 µg donor plasmid and 6 µg of sgRNA-containing PX459 plasmid in Nucleofector solution R (Lonza, Walkersville, MD) and electroporated using the D-032 program of the Amaxa Nucleofector II (Lonza). Following electroporation, cells were treated with a non-homologous end joining inhibitor, 1 µM SCR7 pyrazine (Sigma-Aldrich), for 24 hours to enhance the efficiency of CRISPR-Cas9 genome editing. We expanded the electroporated cells, and bulk sorted for mCherry expression on a BD FACSaria™ fusion flow cytometer (BD Biosciences, San Jose, CA).

TABLES AND FIGURES

Table S1. Sequences of RNAs used in this study

Vendor	Application	RNA samples	Sequence (5' – 3')	Modifications
Integrated DNA Technologies	EMSA, antibody-based RNA-binding assay, CD scans, Thioflavin T assays, and transfection studies	rG1	ACCGGGGCAGGGGAAGAGGGGGAUU	3'-biotinylation
		rG1 mt 1	ACCGAAGCAGAAGAAGAGAAGGAUU	
		rG2a	AGCAGAGGUGGGCGGGGAGGAGGAC	
		rGa mt 1	AGCAGAGGUAAGCGAAGAAAAGGAC	
		rG2b	GUGGGCGGGGAGGAGGACGCGGUGGAAU	
		rG2b mt 1	GUAAGCGGAAAGGAAAACGCAAUGGAAU	
		Linker	GAUCGUGGAGGCAUGAGCAGAGGUG	
Dharmacon	Fluorescent intensity assay	rG1	ACCGGGGCAGGGGAAGAGGGGG	5′-Alexa Fluor 647
	Surface plasmon resonance studies and CD melt curves	rG1	ACCGGGGCAGGGGAAGAGGGGG	5'-biotinylation
		rG1 mt 1	ACCGAAGCAGAAGAAGAGAAGGAUU	

Table S2. siRNA sequences

siRNA	Vendor	Vendor ID	Sequence (5' – 3')
siNegative	Qiagen	SI03650318	Proprietary
siFLI1	Ambion	s5266	CAAACGAUCAGUAAGAAUAtt
siHNRNPH1	Ambion	s6730	GGAUUUGGGUCAGAUAGAUtt
siHNRNPH2	Ambion	145362	GCACUAAAUAGCUACUCCAtt
siHNRNPF	Qiagen	SI00300461	AAGCGTTCGTGCAGTTTGCCT

Table S3. qPCR primers – EWS-FLI1, RNAi targets, Reference genes

Gene target	Gene symbol	Primer	Primer sequence (5' – 3')	Amplicon length (bp)
RNAi targets	*HNRNPH1*	Hs_HNRNPH1 F	TACACATGCGGGGATTACCTT	120 (ENST00000639490.1, ENST00000644339.1, ENST00000393432.8, ENST00000356731.9, ENST00000639162.1, ENST00000329433.10, ENST00000643310.1, ENST00000523921.1, ENST00000519033.5, ENST00000442819.6, ENST00000638505.1, ENST00000521173.5, ENST00000510411.5)
		Hs_HNRNPH1 R	CTTCACCAGTTACTCTGCCATC	
	HNRNPH2	Hs_HNRNPH2 F	GAAGCATACAGGTCCGAATAGC	159 (ENST00000316594.5)
		Hs_HNRNPH2 R	CGCCCCTGAAAGTCCACTG	
	HNRNPF	Hs_HNRNPF_1_SG QuantiTect Primer Assay	NM_001098204, NM_001098205, NM_001098206, NM_001098207, NM_001098208, NM_004966	150 (all variants)
EWS-FLI1 7/6 and 7/5 fusions	*EWS-FLI1* (EF7-7)	EWSR1 exon 7 F	ATCCTACAGCCAAGCTCCAA	type 1 (7/6): 177 type 2 (7/5): 243
		FLI1 exon 7 R	GGCCGTTGCTCTGTATTCTTAC	
Protein-coding EWSR1 variants (EWSR1-001)	*EWSR1* (EE7-9/10)	EWSR1 exon 7 F	ATCCTACAGCCAAGCTCCAA	279 (NM_013986, and NM_001163285) 282 (NM_001163286, and NM_005243)
		EWSR1 exon 9/10 R	GGTCCACCAGGCTTATTGAA	
Reference genes	*ACTB*	Hs_ACTB F	GGCACCCAGCACAATGAAG	66 (ENST00000464611.1, ENST00000493945.6, ENST00000646664.1)
		Hs_ACTB R	CCGATCCACACGGAGTACTTG	
	RPL27	Hs_RPL27 F	ATCGCCAAGAGATCAAAGATAA	123 (ENST00000253788.11, ENST00000589913.6, ENST00000589037.5, ENST00000586277.5)
		Hs_RPL27 R	TCTGAAGACATCCTTATTGACG	
	NACA	Hs_NACA_F	TGGAACAGAATCTGACAGTGATG	82 (ENST00000356769.7, ENST00000546392.5, ENST00000549855.5, ENST00000546862.5, ENST00000550920.5, ENST00000552055.5, ENST00000550952.5, ENST00000454682.5, ENST00000549259.5, ENST00000393891.8, ENST00000552540.5, ENST00000548563.5)
		Hs_NACA_R	CTGGGCTTGTTGTGTGGTTG	

EWS-FLI1 target genes	*NR0B1*	Hs_NR0B1 F	AGCACAAATCAAGCGCAGG	150 (ENST00000378970.4)
		Hs_NR0B1 R	GAAGCGCAGCGTCTTCAAC	
	PRKCB	Hs_PRKCB F	AGCCCCACGTTTTGTGACC	117 (ENST00000643927.1, ENST00000321728.12)
		Hs_PRKCB R	GCTGGGAACATTCATCACGC	
	VRK1	Hs_VRK1 F	TGCACCTTGTGTTGTAAAAGTGG	81 (ENST00000216639.7)
		Hs_VRK1 R	TTTGCAGCTCGTTGGTAGAAC	
	GRK5	Hs_GRK5 F	CCAACACGGTCTTGCTGAAAG	133 (ENST00000392870.2)
		Hs_GRK5 R	TCTCTGTCTATGGTCCTTCGG	
	FCGRT	Hs_FCGRT F	CAACAAGGAGCTCACCTTCC	92 (ENST00000593381.5, ENST00000426395.7, ENST00000221466.9)
		Hs_FCGRT R	GCTCCTTCCACTCCAGGTTT	
	NKX2.2	Hs_NKX2.2 F	TCTACGACAGCAGCGACAAC	74 (ENST00000377142.4)
		Hs_NKX2.2 R	ACCGTGCAGGGAGTACTGAA	
	EZH2	Hs_EZH2 F	GGACCACAGTGTTACCAGCAT	79 (ENST00000350995.6, ENST00000483967.5, ENST00000320356.6, ENST00000460911.5, ENST00000478654.5, ENST00000476773.5)
		Hs_EZH2 R	GTGGGGTCTTTATCCGCTCAG	
	CDKN1A	Hs_CDKN1A F	TGTCCGTCAGAACCCATGC	139 (ENST00000405375.5, ENST00000615513.4, ENST00000244741.9, ENST00000373711.3, ENST00000448526.6)
		Hs_CDKN1A R	AAAGTCGAAGTTCCATCGCTC	
	PHLDA1	Hs_PHLDA1 F	GGAGATCGACTTTCGGTGCC	85 (ENST00000602540.5, ENST00000266671.9, ENST00000619060.1)
		Hs_PHLDA1 R	GGCCTGACGATTCTTGTACTG	
	FOXO1	Hs_FOXO1 F	GGATGTGCATTCTATGGTGTACC	86 (ENST00000379561.5)
		Hs_FOXO1 R	TTTCGGGATTGCTTATCTCAGAC	
	FOSL2	Hs_FOSL2 F	CAGAAATTCCGGGTAGATATGCC	130 (ENST00000379619.5, ENST00000436647.1, ENST00000264716.8)
		Hs_FOSL2 R	GGTATGGGTTGGACATGGAGG	

Table S4. PCR primers – Primers for splicing assay, sequencing and the design of sgRNA and donor plasmids

	Application	Primer	Primer sequence (5' – 3')	PCR product size (bp)
	Sequencing	T3 TAG	ATTAACCCTCACTAAAGGG	N/A
		T7 TAG	AATACGACTCACTATAGG	
EWS-FLI1 (EF7-7)	Splicing assay	EWSR1 exon 7 F	ATCCTACAGCCAAGCTCCAA	177
		FLI1 exon 7 R	GGCCGTTGCTCTGTATTCTTAC	
EWS-FLI1 (EF7-8)	Splicing assay and sequencing	EWSR1 exon 7 F	TATTCCTCTACACAGCCGACTAGTTATGAT	377
		FLI1 exon 7 R	GCTAGGCGACTGCTGGTC	
EWS-FLI1 (EF7-8)	Sequencing	T7- EWSR1 exon 7 F	AATACGACTCACTATAGGTATTCCTCTACAC AGCCGACTAGTTATGAT	414
		T3-FLI1 exon 7 R	ATTAACCCTCACTAAAGGGGCTAGGCGACTG CTGGTC	
sgRNA targeting the C-terminus of *NR0B1*	sgRNA plasmid	sgRNA F	CACCGACTGCACTACTGCACTTGTG	N/A
		sgRNA R	AAACCACAAGTGCAGTAGTGCAGTC	
Left homology arm (LHA)	Donor plasmid	LHA_first round F	ACTGCACTACTGCACTTGTGTGGTTTTGAAC ACTCATTAGAATCATTCTT	First round PCR: 953
		LHA_second round F	GATTACGCCAAGCTTGCATGCCTGCAGGTCA CTGCACTACTGCACTTGTGTGG	Second round PCR: 983
		LHA R	ACCGCATGTTAGCAGACTTCCTCTGCCCTCT ATCTTTGTACAGAGCATTTCCA	
Right homology arm (RHA)		RHA_F	ACCGGCGGCATGGACGAGCTGTACAAGTAAA GTCATGTGGGGCACACA	First round PCR: 953
		RHA_first round R	ACTGCACTACTGCACTTGTGTGGATAGACAT ATAGTAGCCTTGATTTT	Second round PCR: 983
		RHA_second round R	GAGCTCGGTACCCGGGGATCCTCTAGAGTCA CTGCACTACTGCACTTGTGTGG	
mCherry cassette		T2A-mCherry cassette F	ATGATGCTGGAAATGCTCTGTACAAAGATAG AGGGCAGAGGAAGTCTG	831
		T2A-mCherry cassette R	CACTACTGCACTTGTGTGCCCCACATGACTT TACTTGTACAGCTCGTCCAT	
CRISPR plasmid sgRNA	Sequencing	Human U6 promoter F	GACTATCATATGCTTACCGT	N/A
pUC19 vector	Sequencing	pUC19_Addgene# 50005_M13 F	AGCGGATAACAATTTCACACAGG	N/A
		pUC19_Addgene# 50005_M13 R	CCCAGTCACGACGTTGTAAAACG	
Donor plasmid	Sequencing	LHA-T2A mCherry_junction primer	GGGAACTCAGCAAATACTCAGTG	N/A
		mCherry_RHA_junction primer	GACTACTTGAAGCTGTCCTTCC	

Table S5. Sequences of DNA used to express truncated HNRNPH1 proteins
*Numbering based on wild-type HNRNPH1 protein sequence

Domain	Sequence (5' to 3')	Encoded protein	Residues*
qRRM1	ATGAGATCTGGATCCCACCATCACCATCAC CATAGATCTGATATCACAAGTTTGTACAAA AAAGTTGGCGAAAACCTGTACTTCCAAGGC ATGTTGGGCACGGAAGGTGGAGAGGGATTC GTGGTGAAGGTCCGGGGCTTGCCCTGGTCT TGCTCGGCCGATGAAGTGCAGAGGTTTTTT TCTGACTGCAAAATTCAAAATGGGGCTCAA GGTATTCGTTTCATCTACACCAGAGAAGGC AGACCAAGTGGCGAGGCTTTTGTTGAACTT GAATCAGAAGATGAAGTCAAATTGGCCCTG AAAAAAGACAGAGAAACTATGGGACACAGA TATGTTGAAGTATTCAAGTCAAACAACGTT GAAATGGATTGGGTGTTGAAGCATACTGGT CCATAA	MRSGSHHHHHHRSDITSLYKKVGE NLYFQGMLGTEGGEGFVVKVRGLP WSCSADEVQRFFSDCKIQNGAQGI RFIYTREGRPSGEAFVELESEDEV KLALKKDRETMGHRYVEVFKSNNV EMDWVLKHTGP	1-102
qRRM2	ATGAGATCTGGATCCCACCATCACCATCAC CATAGATCTGATATCACAAGTTTGTACAAA AAAGTTGGCGAAAACCTGTACTTCCAAGGC AATAGTCCTGACACGGCCAATGATGGCTTT GTACGGCTTAGAGGACTTCCCTTTGGATGT AGCAAGGAAGAAATTGTTCAGTTCTTCTCA GGGTTGGAAATCGTGCCAAATGGGATAACA TTGCCGGTGGACTTCCAGGGGAGGAGTACG GGGGAGGCCTTCGTGCAGTTTGCTTCACAG GAAATAGCTGAAAAGGCTCTAAAGAAACAC AAGGAAAGAATAGGGCACAGGTATATTGAA ATCTTTAAGAGCAGTAGAGCTGAAGTTAGA ACTCATTAA	MRSGSHHHHHHRSDITSLYKKVGE NLYFQGNSPDTANDGFVRLRGLPF GCSKEEIVQFFSGLEIVPNGITLP VDFQGRSTGEAFVQFASQEIAEKA LKKHKERIGHRYIEIFKSSRAEVR TH	103-194
qRRM3	ATGAGATCTGGATCCCACCATCACCATCAC CATAGATCTGATATCACAAGTTTGTACAAA AAAGTTGGCGAAAACCTGTACTTCCAAGGC GGGGATGGTGGCTCTACTTTCCAGAGCACA ACAGGACACTGTGTACACATGCGGGGATTA CCTTACAGAGCTACTGAGAATGACATTTAT AATTTTTTTTCACCGCTCAACCCTGTGAGA GTACACATTGAAATTGGTCCTGATGGCAGA GTAACTGGTGAAGCAGATGTCGAGTTCGCA ACTCATGAAGATGCTGTGGCAGCTATGTCA AAAGACAAAGCAAATATGCAACACAGATAT GTAGAACTCTTCTTGAATTCTACAGCAGGA GCAAGCGGTGGTGCTTACGAACACAGATAT GTAGAACTCTTCTTGTAA	MRSGSHHHHHHRSDITSLYKKVGE NLYFQGGDGGSTFQSTTGHCVHMR GLPYRATENDIYNFFSPLNPVRVH IEIGPDGRVTGEADVEFATHEDAV AAMSKDKANMQHRYVELFLNSTAG ASGGAYEHRYVELFL	277-381
qRRM1-2	ATGAGATCTGGATCCCACCATCACCATCAC CATAGATCTGATATCACAAGTTTGTACAAA AAAGTTGGCGAAAACCTGTACTTCCAAGGC ATGTTGGGCACGGAAGGTGGAGAGGGATTC GTGGTGAAGGTCCGGGGCTTGCCCTGGTCT TGCTCGGCCGATGAAGTGCAGAGGTTTTTT TCTGACTGCAAAATTCAAAATGGGGCTCAA GGTATTCGTTTCATCTACACCAGAGAAGGC AGACCAAGTGGCGAGGCTTTTGTTGAACTT GAATCAGAAGATGAAGTCAAATTGGCCCTG AAAAAAGACAGAGAAACTATGGGACACAGA TATGTTGAAGTATTCAAGTCAAACAACGTT	MRSGSHHHHHHRSDITSLYKKVGE NLYFQGMLGTEGGEGFVVKVRGLP WSCSADEVQRFFSDCKIQNGAQGI RFIYTREGRPSGEAFVELESEDEV KLALKKDRETMGHRYVEVFKSNNV EMDWVLKHTGPNSPDTANDGFVRL RGLPFGCSKEEIVQFFSGLEIVPN GITLPVDFQGRSTGEAFVQFASQE IAEKALKKHKERIGHRYIEIFKSS RAEVRTH	1-194

	GAAATGGATTGGGTGTTGAAGCATACTGGT CCAAATAGTCCTGACACGGCCAATGATGGC TTTGTACGGCTTAGAGGACTTCCCTTTGGA TGTAGCAAGGAAGAAATTGTTCAGTTCTTC TCAGGGTTGGAAATCGTGCCAAATGGGATA ACATTGCCGGTGGACTTCCAGGGGAGGAGT ACGGGGGAGGCCTTCGTGCAGTTTGCTTCA CAGGAAATAGCTGAAAAGGCTCTAAAGAAA CACAAGGAAAGAATAGGGCACAGGTATATT GAAATCTTTAAGAGCAGTAGAGCTGAAGTT AGAACTCATTAA		
qRRM2-3	ATGAACTCTCCGGATACTGCTAATGATGGT TTCGTACGTCTGCGTGGTCTGCCATTCGGT TGTTCTAAAGAGGAGATCGTTCAGTTCTTC TCTGGCCTGGAGATCGTACCGAACGGCATC ACCCTGCCGGTGGACTTTCAGGGTCGTTCT ACTGGCGAAGCATTCGTTCAGTTCGCTTCC CAGGAAATCGCAGAGAAAGCCCTGAAGAAA CACAAAGAGCGTATCGGTCATCGTTACATT GAAATCTTCAAATCTTCCCGTGCTGAGGTG CGTACCCATTACGATCCACCACGTAAACTG ATGGCTATGCAACGTCCGGGTCCGTACGAC CGTCCGGGTGCCGGTCGCGGTTACAACTCT ATCGGTCGCGGTGCTGGCTTCGAACGTATG CGTCGTGGTGCGTATGGTGGTGGTTACGGT GGTTACGATGACTATAACGGCTACAACGAT GGCTATGGCTTCGGCTCTGACCGCTTCGGT CGCGACCTGAACTACTGCTTCTCTGGTATG TCCGATCACCGTTACGGTGATGGTGGCTCC ACCTTTCAGAGCACCACTGGTCACTGTGTT CACATGCGTGGCCTGCCGTACCGTGCAACT GAGAACGACATCTATAACTTCTTCTCTCCG CTGAACCCGGTTCGTGTTCATATTGAGATC GGTCCGGATGGCCGTGTGACTGGTGAAGCC GACGTTGAGTTCGCAACCCATGAGGACGCT GTAGCTGCGATGTCCAAAGATAAAGCTAAC ATGCAGCATCGCTATGTTGAGCTGTTCCTG AACAGCACTGCTGGCGCTTCTGGTGGTGCT TACGAACATCGCTATGTAGAACTGTTCCTG CACCATCACCATCACCACTAA	MNSPDTANDGFVRLRGLPFGCSKE EIVQFFSGLEIVPNGITLPVDFQG RSTGEAFVQFASQEIAEKALKKHK ERIGHRYIEIFKSSRAEVRTHYDP PRKLMAMQRPGPYDRPGAGRGYNS IGRGAGFERMRRGAYGGGYGGYDD YNGYNDGYGFGSDRFGRDLNYCFS GMSDHRYGDGGSTFQSTTGHCVHM RGLPYRATENDIYNFFSPLNPVRV HIEIGPDGRVTGEADVEFATHEDA VAAMSKDKANMQHRYVELFLNSTA GASGGAYEHRYVELFLHHHHHH	103-381
Δ1-103	HNRNPH1-13871H (Creative BioMart)	MGSSHHHHHHSSGLVPRGSHMASM TGGQQMGRGSSPDTANDGFVRLRG LPFGCSKEEIVQFFSGLEIVPNGI TLPVDFQGRSTGEAFVQFASQEIA EKALKKHKERIGHRYIEIFKSSRA EVRTHYDPPRKLMAMQRPGPYDRP GAGRGYNSIGRGAGFERMRRGAYG GGYGGYDDYNGYNDGYGFGSDRFG RDLNYCFSGMSDHRYGDGGSTFQS TTGHCVHMRGLPYRATENDIYNFF SPLNPVRVHIEIGPDGRVTGEADV EFATHEDAVAAMSKDKANMQHRYV ELFLNSTAGASGGAYEHRYVELFL NSTAGASGGAYGSQMMGGMGLSNQ SSYGGPASQQLSGGYGGGYGGQSS MSGYDQVLQENSSDFQSNIA	104-449

Table S6. Homo sapiens *EWSR1* transcripts from the GRCh37 assembly - ENSG00000182944

Name	Transcript ID	bp	Protein	Biotype	CCDS	UniProt	RefSeq
EWSR1-001	ENST00000397938.2	2654	656aa	Protein coding	CCDS13851	Q01844	NM_001163285 NM_001163286 NM_005243 NM_013986 NP_001156757 NP_001156758 NP_005234 NP_053733
EWSR1-002	ENST00000406548.1	2207	655aa	Protein coding	CCDS54513	Q01844	-
EWSR1-021	ENST00000414183.2	2189	661aa	Protein coding	CCDS13852	Q01844	-
EWSR1-008	ENST00000332035.6	1989	600aa	Protein coding	CCDS54514	B0QYJ7 Q01844	-
EWSR1-004	ENST00000333395.6	1265	354aa	Protein coding	CCDS54512	Q01844	NM_001163287 NP_001156759
EWSR1-201	ENST00000332050.6	2443	583aa	Protein coding	-	C9JGE3	-
EWSR1-007	ENST00000331029.7	2267	618aa	Protein coding	-	B0QYK0	-
EWSR1-019	ENST00000360091.3	1289	308aa	Protein coding	-	-	-
EWSR1-013	ENST00000437155.2	933	292aa	Protein coding	-	F8WC90	-
EWSR1-018	ENST00000455726.1	710	233aa	Protein coding	-	B0QYJ7	-
EWSR1-017	ENST00000415761.1	587	180aa	Protein coding	-	B0QYJ6	-
EWSR1-014	ENST00000447973.1	582	166aa	Protein coding	-	B0QYJ4	-
EWSR1-016	ENST00000444626.1	560	77aa	Protein coding	-	B0QYJ5	-
EWSR1-012	ENST00000436425.1	553	156aa	Protein coding	-	B0QYJ3	-
EWSR1-005	ENST00000479135.1	7285	No protein	Retained intron	-	-	-
EWSR1-006	ENST00000483415.1	3231	No protein	Retained intron	-	-	-
EWSR1-009	ENST00000469669.1	2630	No protein	Retained intron	-	-	-
EWSR1-011	ENST00000490315.1	2301	No protein	Retained intron	-	-	-
EWSR1-010	ENST00000483629.1	1818	No protein	Retained intron	-	-	-
EWSR1-015	ENST00000493426.1	581	No protein	Retained intron	-	-	-
EWSR1-003	ENST00000485037.1	479	No protein	Retained intron	-	-	-

Table S7 qPCR additional experimental details (see separate pdf)

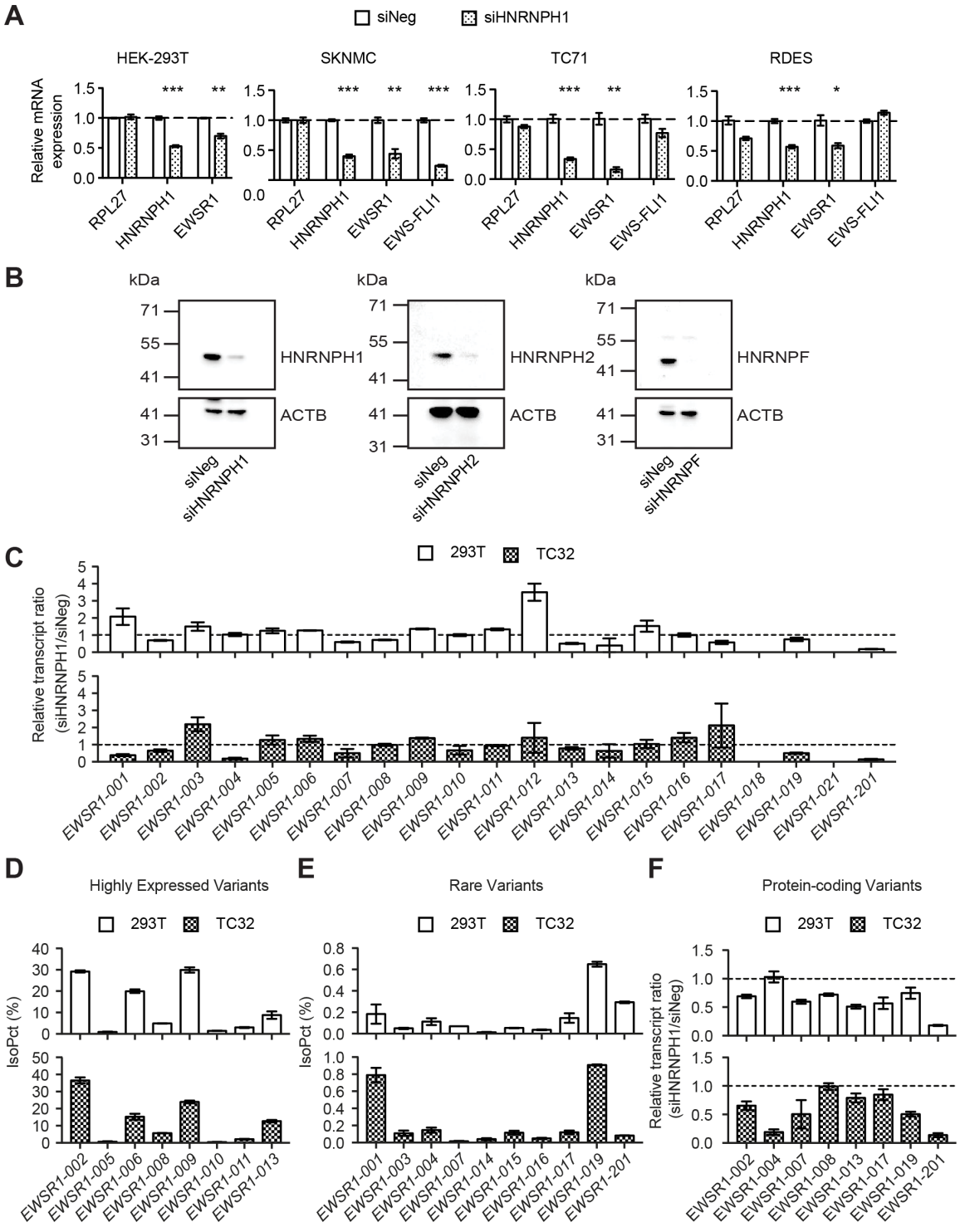

Figure S1 related to Figure 1. HNRNPH1-mediated processing of *EWSR1* exon 8 containing pre-mRNAs.

(A) qPCR assessment of *RPL27* (control), *HNRNPH1*, *EWS-FLI1* and *EWSR1* mRNA expression in *HNRNPH1*-silenced EWS cell lines (SKMNC, TC71 and RD-ES) and HEK-293T cells. Data

normalized to *ACTB* are expressed relative to siNeg-transfected cells (48 hours, mean ± SEM, n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to siNeg-transfected cells. For additional experimental details, refer to **Table S7.**

(B) Immunoblot analysis of whole cell lysates prepared from TC32 EWS cells 48 hours post transfection of *HNRNPH1*, *HNRNPH2*, and *HNRNPF* siRNAs using antibodies against the proteins indicated.

(C) Expression of all *EWSR1* transcript variants from targeted RNA-seq data analyzed using RSEM for *HNRNPH1*-transfected HEK-293T cells (top) and TC32 EWS cells (bottom). Data are expressed relative to siNeg-transfected cells (48 hours, mean ± SEM, n=3).

(D) Expression of highly expressed *EWSR1* transcript variants from targeted RNA-seq data analyzed using RSEM for *HNRNPH1*-transfected HEK-293T cells (top) and TC32 EWS cells (bottom). Data are expressed relative to siNeg-transfected cells (48 hours, mean ± SEM, n=3).

(E) Expression of rare *EWSR1* transcript variants from targeted RNA-seq data analyzed using RSEM for *HNRNPH1*-transfected HEK-293T cells (top) and TC32 EWS cells (bottom). Data are expressed relative to siNeg-transfected cells (48 hours, mean ± SEM, n=3).

(F) Expression of protein-coding *EWSR1* transcript variants from targeted RNA-seq data analyzed using RSEM for *HNRNPH1*-transfected HEK-293T cells (top) and TC32 EWS cells (bottom). Data are expressed relative to siNeg-transfected cells (48 hours, mean ± SEM, n=3).

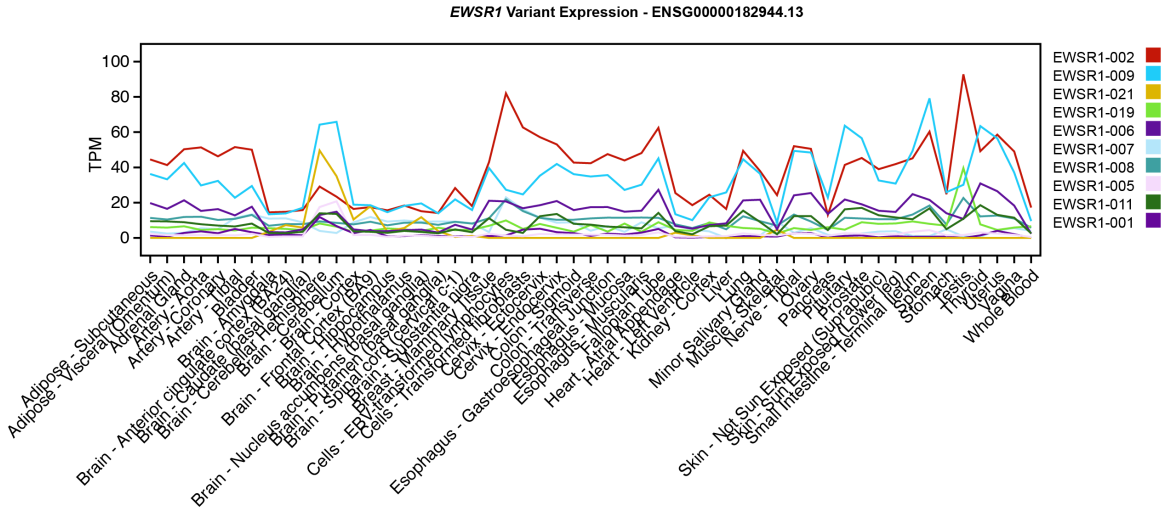

Figure S2 related to Figure 1. Expression of *EWSR1* transcript variants vary across tissues. Data obtained from the Genotype-Tissue Expression (GTEx) portal on April 17, 2018.

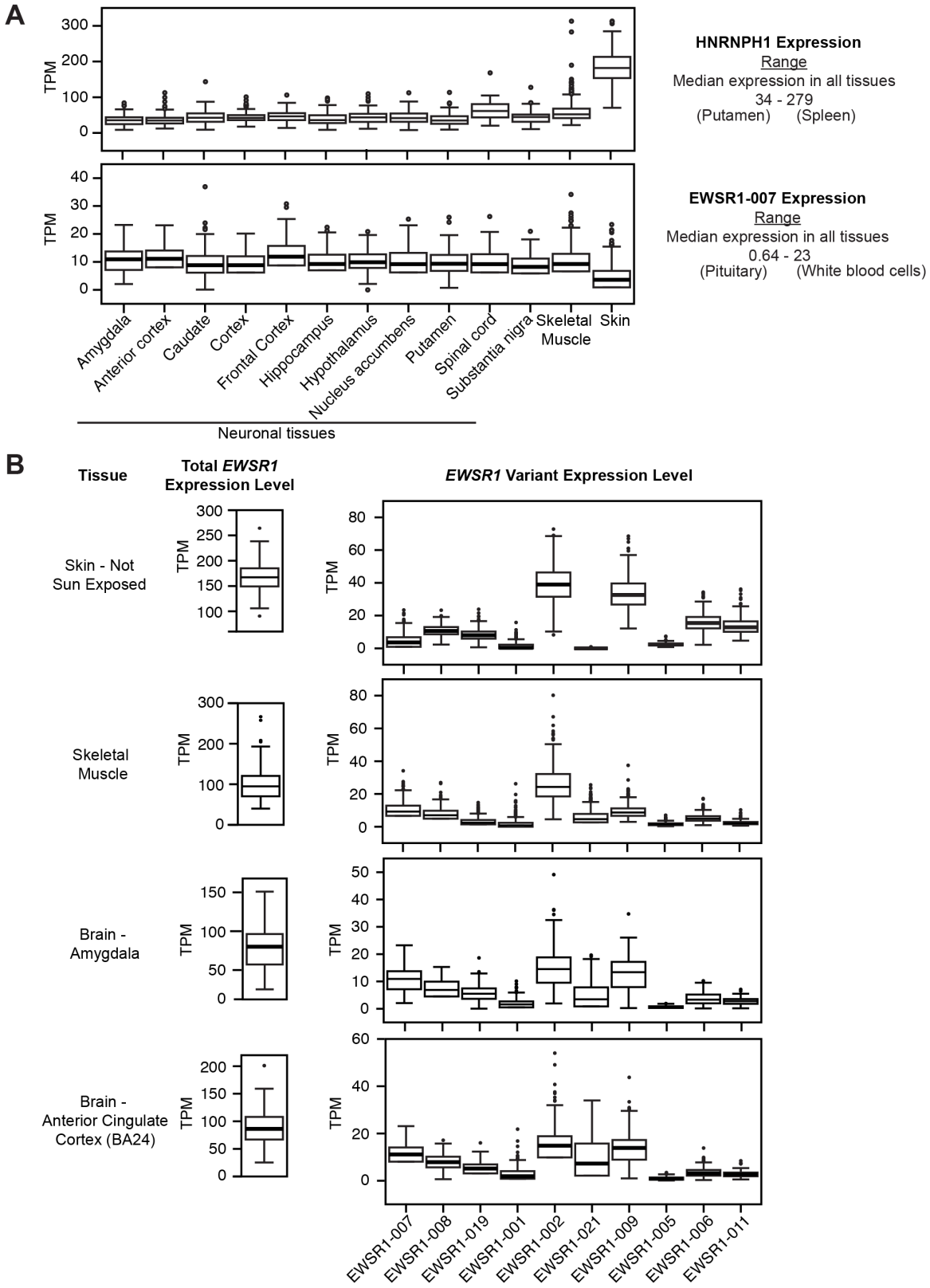

Figure S3 related to Figure 1. *EWSR1-007* transcript variant is more readily expressed in tissues containing lower HNRNPH1 expression levels.

(A) Expression of total *HNRNPH1* (top) and *EWSR1-007* variant (bottom) transcripts in neuronal tissues, skeletal muscle, and skin.
(B) Comparison of total *EWSR1* transcript expression level and its transcript variants in indicated tissues.
Data obtained from the Genotype-Tissue Expression (GTEx) portal on May 5, 2018.

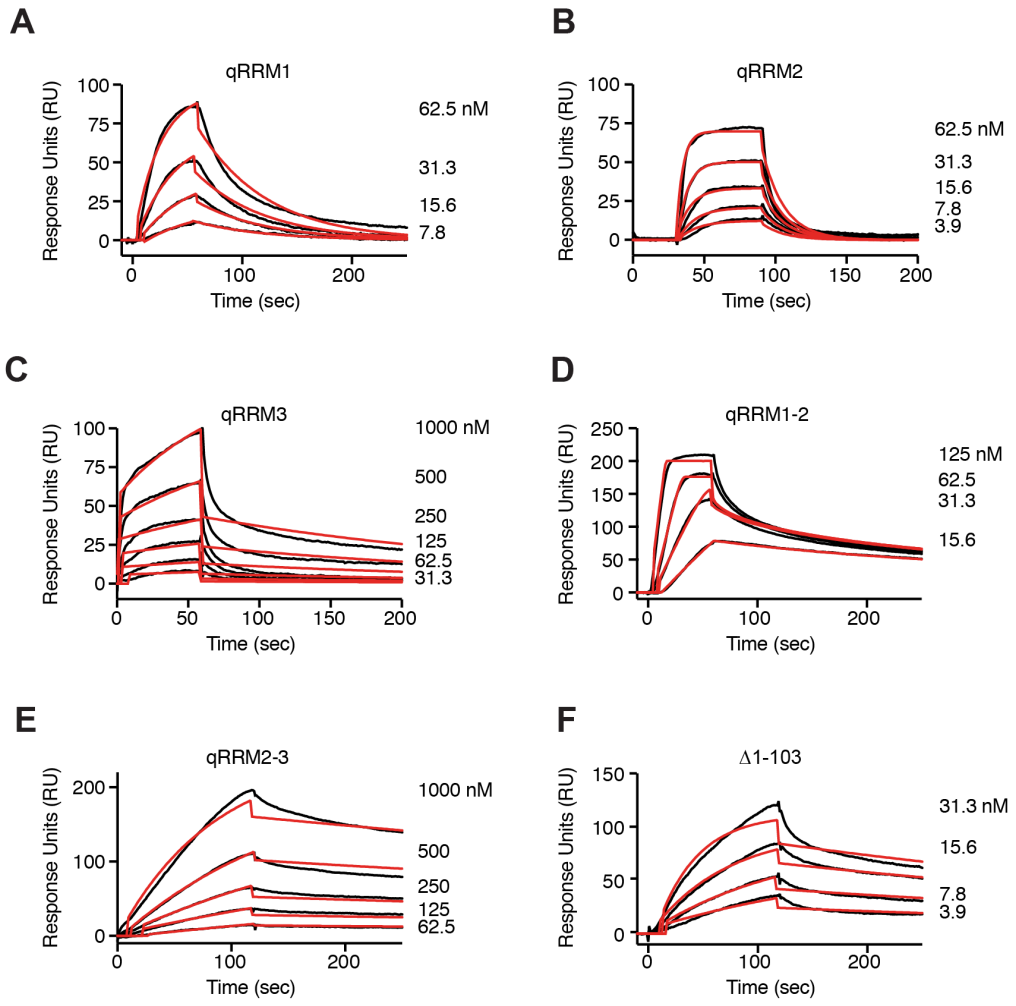

Figure S4 related to Figure 3. Kinetic analysis of different HNRNPH1 domains and rG1 interactions by SPR. The binding of different HNRNPH1 domains to a rG1 RNA target surface is shown. Black lines represent the binding responses for each protein at the indicated concentrations on each sensorgram. HNRNPH1 was exposed to the surface for 2 minutes (association phase) followed by 10 minutes flow of running buffer (dissociation phase). Data for qRRM1-2 were fit globally to a 1:1 binding model with mass transfer and all other data were fitted using a 1:1 binding model. The respective fits are shown in red. The resulting parameter values are provided in Table 1.

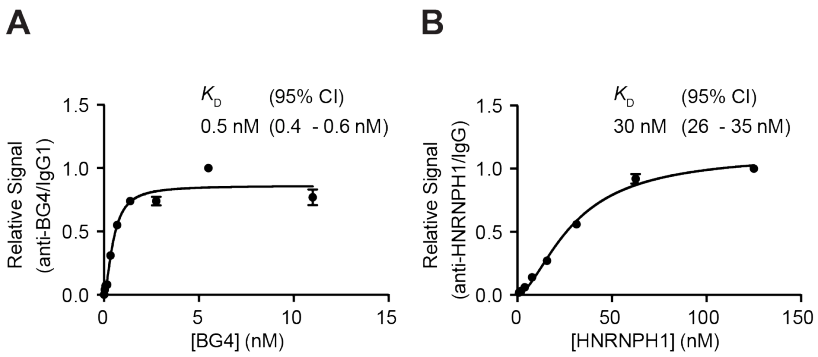

Figure S5 related to Figure 4. HNRNPH1 directly binds to unstructured G-tracts and RNA G-quadruplex structures within *EWSR1* exon 8 rG1.
(A) Binding curve for rG1-BG4 immunocomplex determined by the antibody-based RNA binding assay using BG4 antibody. Apparent dissociation constants (K_d^{app}) for each immunocomplex are indicated. Data are shown as the mean ± SEM of 3 technical replicates.
(B) Binding curve for rG1-HNRNPH1 immunocomplex determined by the antibody-based RNA binding assay using HNRNPH1 antibody. Apparent dissociation constants (K_d^{app}) for each immunocomplex are indicated. Data are shown as the mean ± SEM of 3 technical replicates.

CI = confidence interval.

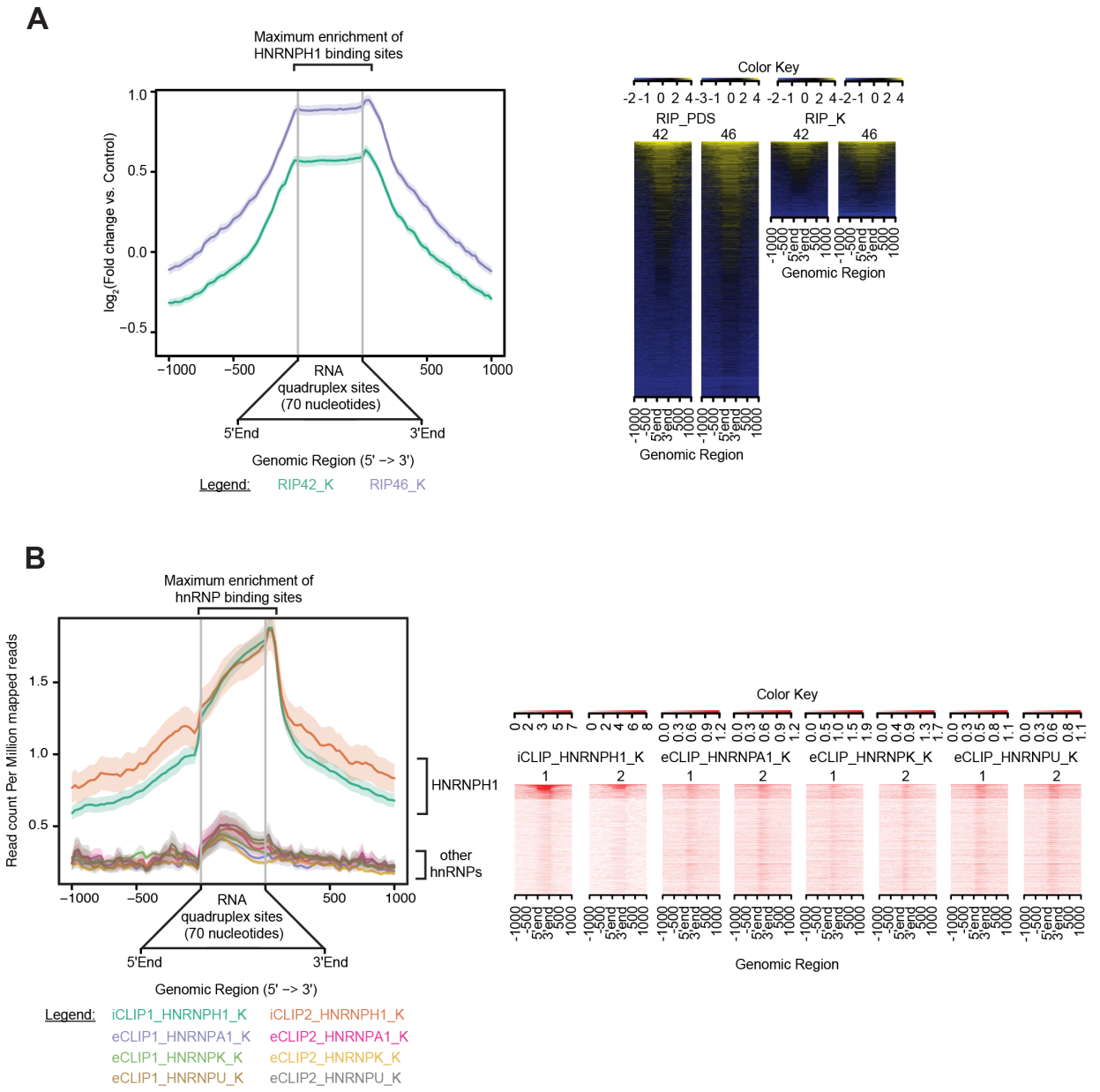

Figure S6 related to Figure 5. A subgroup of RNA guanine quadruplex regions are enriched near HNRNPH1 binding sites in the human genome.

(A) Coverage plots (left) and read density heat maps (right) represent binding read count for HNRNPH1 near RNA guanine quadruplex sites (± 1 Kb). The RNA targets and binding sites for HNRNPH1 were obtained from RIP seq dataset (Uren et al. 2016). The RNA guanine quadruplex sites were obtained from the sequencing dataset GSE77282 (Kwok et al. 2016).

(B) Coverage plots (left) and read density heat maps (right) represent binding read count for the indicated hnRNPs near RNA guanine quadruplex sites (± 1 Kb). The RNA targets and binding sites for each hnRNP were obtained from iCLIP or eCLIP datasets (Uren et al. 2016, Van Nostrand et al. 2016). The RNA guanine quadruplex sites were obtained from the sequencing dataset GSE83617 (Guo and Bartel 2016).

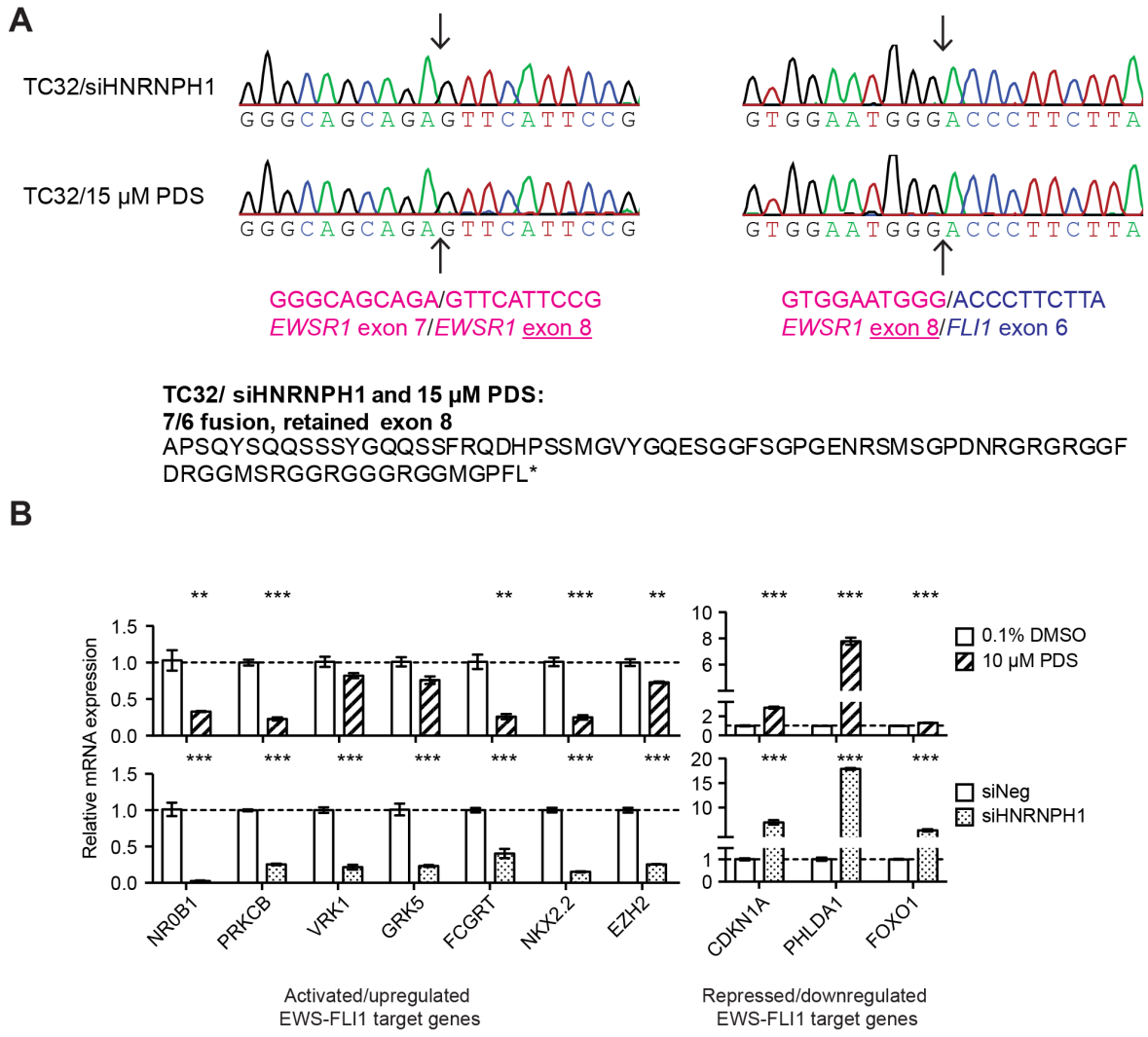

Figure S7 related to Figure 7. In TC32 cells, PDS modulates *EWS-FLI1* mRNA processing, decreases EWS-FLI1 protein levels, and restores mRNA expression of EWS-FLI1 deregulated targets.

(A) Representative sequence chromatograms show retention of *EWSR1* exon 8 for the additional PCR product amplified from siHNRNPH1-transfected and PDS-treated TC32 cells. The exon junctions are indicated by arrows. The predicted translation of a fusion transcript that retains *EWSR1* exon 8 results in the introduction of a premature stop codon.

(B) qPCR assessment of EWS-FLI1-regulated genes in PDS-treated (top) or *HNRNPH1*-silenced (bottom) TC32 EWS cells (48 hours). Data normalized to *ACTB* are expressed relative to control, 0.1% DMSO or siNeg-transfected cells (mean ± SEM, n=3). ** p<0.01; *** p<0.001 compared to control. For additional experimental details, refer to **Table S7.**

SUPPLEMENTAL REFERENCES

Guo, J.U. and Bartel, D.P. 2016. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* 353: aaf5371.

Kwok, C.K., Marsico, G., Sahakyan, A.B., Chambers, V.S. and Balasubramanian, S. 2016. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat Methods* 13: 841-844.

Studier, F.W. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41: 207-234.

Uren, P.J., Bahrami-Samani, E., de Araujo, P.R., Vogel, C., Qiao, M., Burns, S.C., Smith, A.D. and Penalva, L.O. 2016. High-throughput analyses of hnRNP H1 dissects its multi-functional aspect. *Rna Biol* 13: 400-411.

Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K. et al. 2016. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* 13: 508-514.

Zhang, J.P., Li, X.L., Li, G.H., Chen, W., Arakaki, C., Botimer, G.D., Baylink, D., Zhang, L., Wen, W., Fu, Y.W. et al. 2017. Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biol* 18: 35.