**SUPPLEMENTARY INFORMATION LEGENDS**

**Figure S1**. (A-F), (A) Length, (B) and (C) percentage of positive and negatively charged amino acids (or AAs),(D) PScore, (E) protein abundance and (F) phosphosite count were compared for the indicated whole proteome, nuclear proteome, and respective RLBP datasets. The exact p-values resulting from a Mann-Whitney-Wilcoxon test after Bonferroni correction for multiple comparisons are reported.

**Figure S2**. (A-I), (A) Hydrophobicity and hydrophilicity (GRAVY),(B) aliphatic index, (C) length, (D) protein disorder, (E) percentage of low-complexity regions, (F) PScore, (G) solubility, (H) phosphosite count, and (I) abundance were compared for the indicated Not Predicted (Not Pred) and Predicted (Pred) datasets of nucleic-acid binding proteins in both IP-MS and Prox-MS datasets. The exact p-values resulting from a Mann-Whitney-Wilcoxon test after Bonferroni correction for multiple comparisons are reported.

**Figure S3.** (A) Western blot of siRNA knockdowns presented in Figure 6C. (B) Quantification (*Left*) and representative images (*Right*) of PLA single antibody controls (blue = DAPI, red = PLA signal; scale bar = 20µm). Quantification of LIG1/S9.6 and FXR1 S9.6 are only shown for reference and presented in detail in Figures 6A-B. (C) Representative images of S9.6 staining from the quantification data presented in Figure 6B (siRNA knockdown; blue = DAPI, red = S9.6; scale bar = 20µm). (D) Representative images of S9.6 staining from the quantification data presented in Figure 6C (LIG1 inhibition with L82-G17; (blue = DAPI, red = S9.6; scale bar = 20µm). (E) Relative DRIP-qPCR signal values at *RPL13A*  genes in HeLa cells transfected with indicated siRNAs and treated with *in vitro* RNaseH pre-immunoprecipitation where indicated (mean +/- SEM, One-way ANOVA, n = 5).

**Table S1**. Proteins found to interact with R-loops. *Sheet 1*: R-loop binding proteins (RLBPs) found via IP-MS methods. *Sheet 3*: RLBPs found via Prox-MS methods. *Sheet 4*: Intersect of published RLBP datasets. *Sheet 4*: PFAM domain enrichment values for RLBPs.

**Table S2**. Protein features for the human proteome. *Sheet 1*: All features used for the random forest algorithm per protein. *Sheet 2*: PFAM domains per protein for the human proteome.

**Table S3**. Probabilities for each protein called by the easy ensemble random forest algorithms. *Sheet 1*: Probability for each protein in all 100 models developed with the IP-MS training dataset. *Sheet 2*: Probability for each protein in all 100 models developed with the Prox-MS training dataset. *Sheet 3*: Averaged probability for each protein. *Sheet 4*: Performance metrics on the test set for both models. *Sheet 5*: Candidate RLBPs with probabilities greater than a threshold of >=0.8, *Sheet 6*: RnaseH values

**Table S4.** Gene ontology analyses for significantly enriched biological processes (*Sheet 1-3*) and PFAM domain enrichment (*Sheet 4-6*) across our IP-MS RF and Prox-MS RF datasets and their overlap (DAVID v6.8; p<0.05, with FDR correction). Highlighted rows represent plotted values in **Fig. 4D** and **E**.

**Table S5**. Oligonucleotides used in this study.