Supplementary Information

RPRD1B's direct interaction with phosphorylated RNA Polymerase II regulates polyadenylation of cell cycle genes and drives cancer progression

Rosamaria Y. Moreno [1], Svetlana B. Panina [1], and Y. Jessie Zhang[1]

[1] Department of Molecular Biosciences, University of Texas, Austin, Texas

[*] Correspondence should be addressed to Y. Jessie Zhang (jzhang@cm.utexas.edu)

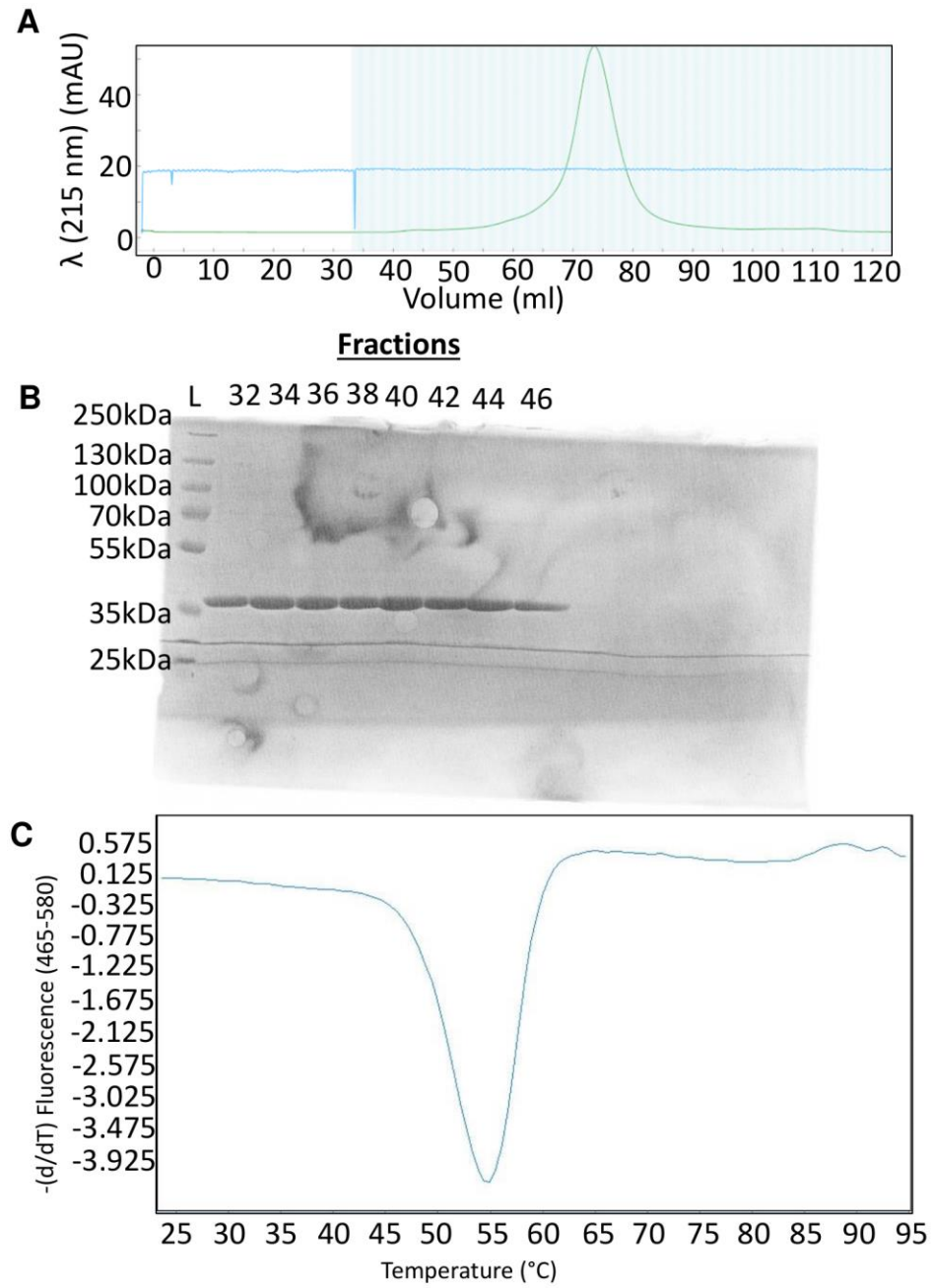


Figure S1. **Purification of RPRD1B CID**

(**A**) Size exclusion chromatography of purified RPRD1B CID protein (**B**) Uncropped SDS-PAGE gel of fractions isolated from size exclusion chromatography from RPRD1B-CID purification in *E. coli*. (**C**) Differential scanning fluorometry plot showing the melting temperature of the CID domain of RPRD1B

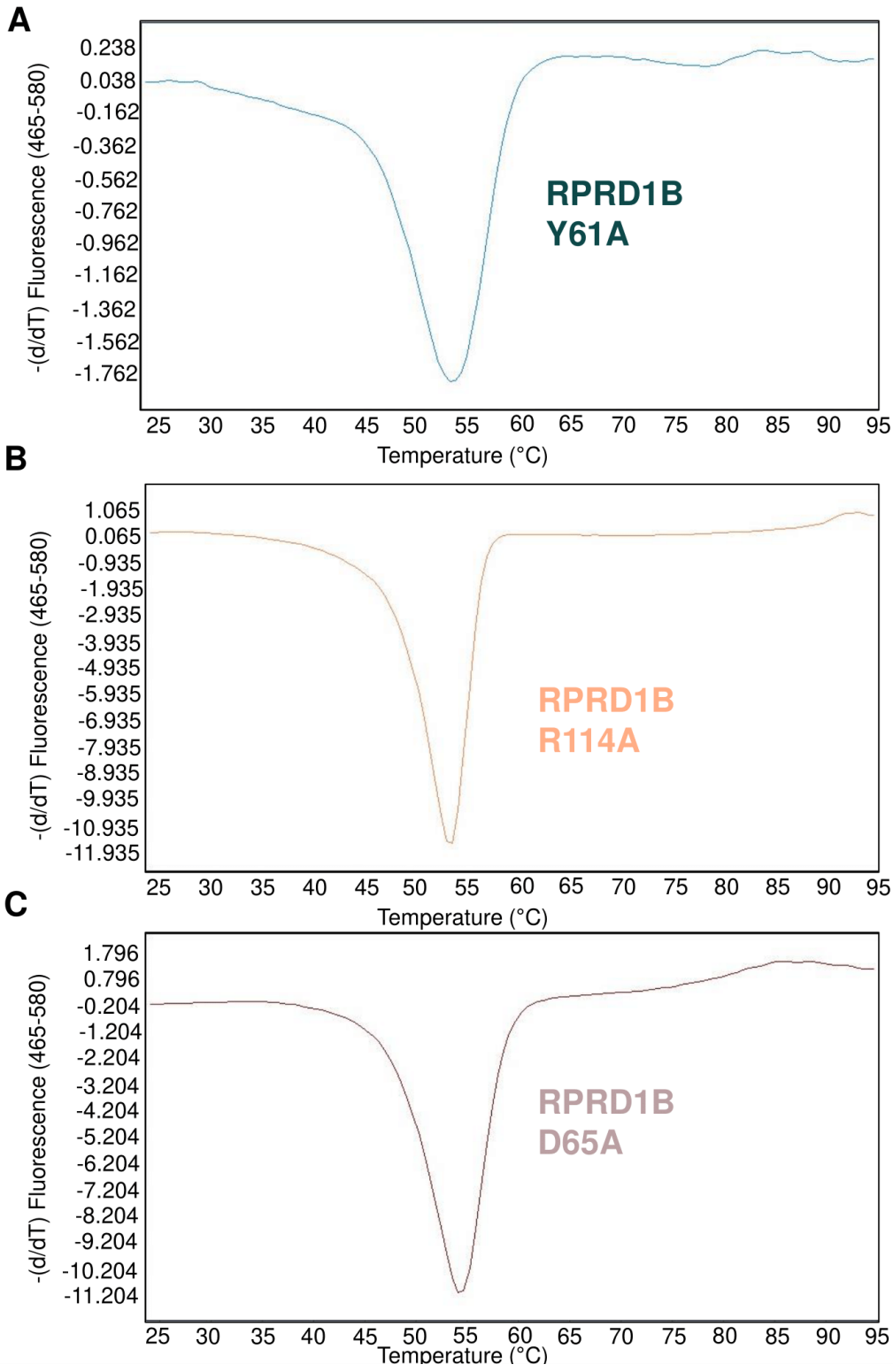
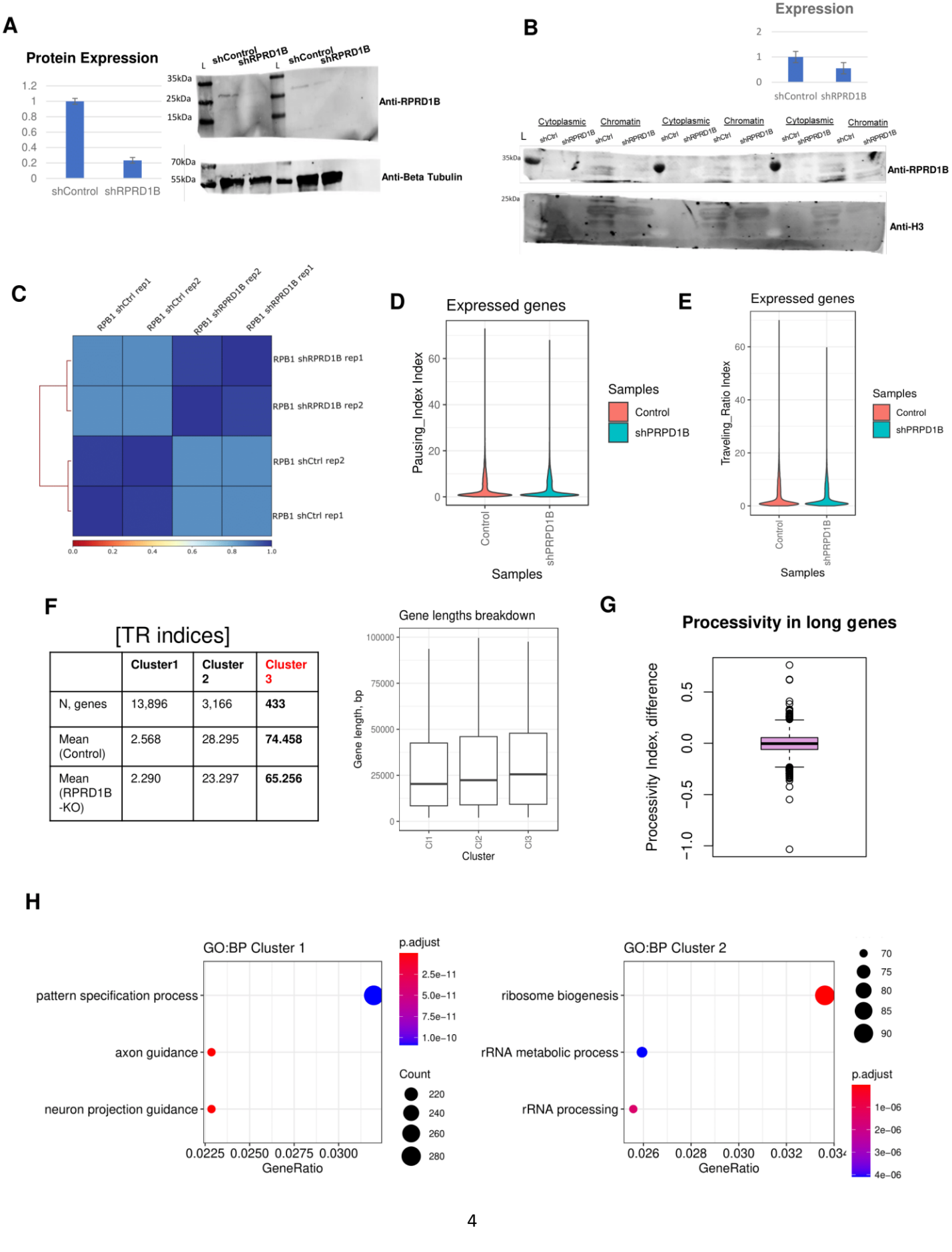


Figure S2. **Characterization of RPRD1B mutants.**

(**A-C**) Differential scanning fluorometry plots showing the melting temperature of mutant versions of the CID of RPRD1B.



	Cluster1	Cluster 2	Cluster 3
N, genes	13,896	3,166	**433**
Mean (Control)	2.568	28.295	**74.458**
Mean (RPRD1B -KO)	2.290	23.297	**65.256**

Figure S3. **Effects of RPRD1B knockdown on RPB1 genomic distribution**

(A) Western blot analysis of RPRD1B knockdown efficiency by shRNA where uncropped representative blot of shRPRD1B vs control (MISSION non-mammalian shRNA control plasmid) shown (50 µg each). Quantification of western blot was done with three biological replicates with error bars indicating standard error. **(B)** Western blot analysis of RPRD1B expression in cytoplasmic or chromatin fractions. Quantification of western blot was done with three biological replicates with error bars indicating standard error. **(C)** Heatmaps showing the Pearson correlation between ChIP-seq replicate datasets of RPB1-bound regions in shRPRD1B or shControl samples. **(D)** Pausing index, the ratio between read density of Pol II around TSS (-50 to +300) and gene body (TSS +3000 to TES) between Pol II ChIP-seq samples. Calculations were performed for all expressed genes with a length greater than 2000 bp. **(E)** Traveling Ratio, comparing read density in the promoter (TSS to TSS+300bp) to gene body (TSS+300bp to TES) from ChIP-seq samples. Calculations were performed for all expressed genes with a length greater than 2000 bp. (**F**) Traveling ratio indices were calculated for RPB1-RPRD1B KO and RPB1-Control ChIP-seq dataset GSE243457. Gene length breakdown in cluster 1, 2, and 3 is shown. **(G)** Processivity index, the ratio of the 5′/3′ PolII signal ratio in the gene body, was calculated. Long genes are identified as protein-coding genes with a length of > 75^{th} percentile of annotation distribution, e.g. > 38,808 bp (from gene start to end). Traveling ratio, Pausing index, and processivity index calculations were calculated as an average of sample replicates per condition. **(H)** Gene ontology (GO) analysis of enriched biological processes among genes within cluster 1 and 2.

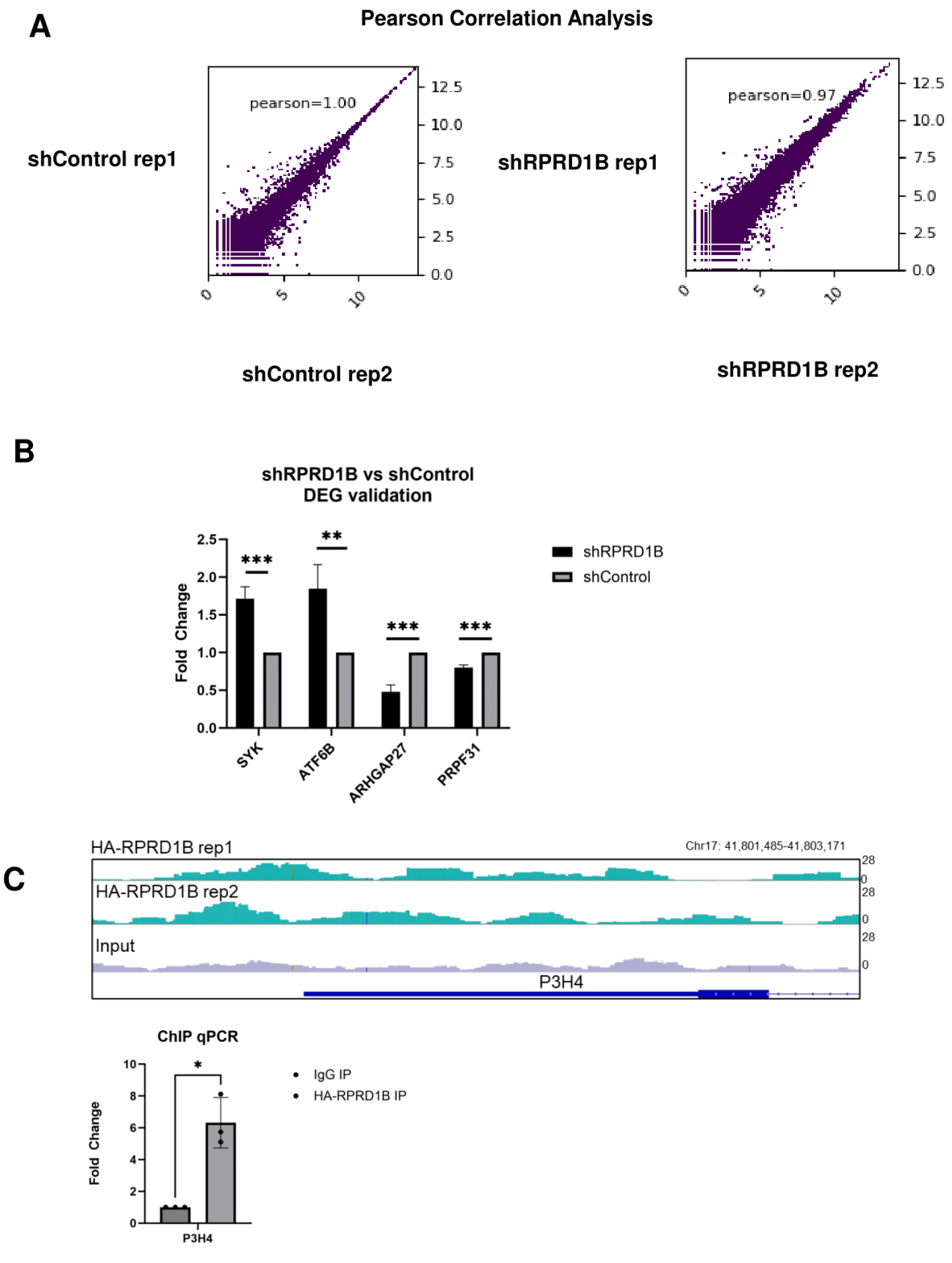


Figure S4. RNA-seq validation of shRPRD1B vs shControl samples

(A) Scatterplots showing the Pearson correlation between RNA-seq replicates for shRPRD1B and shControl samples. The genome was divided into bins of 10 kb and the number of mapped reads in the individual bins was calculated. (**B**) Relative transcript expression of down or upregulated genes from RNA-seq analysis of shRPRD1B. For each data point, $n = 3$, error bars indicate standard deviation of three biological replicates. (**C**) Relative transcript expression of HA-RPRD1B occupancy and IgG control samples on a representative gene. Fold enrichment was calculated by comparing the positive locus sequence in ChIP DNA over the negative IgG sample. For each data point, $n=3$, error bars indicate standard deviation of three biological replicates. $p \leq 0.001$ (**) and $p \leq 0.0001$ (***).

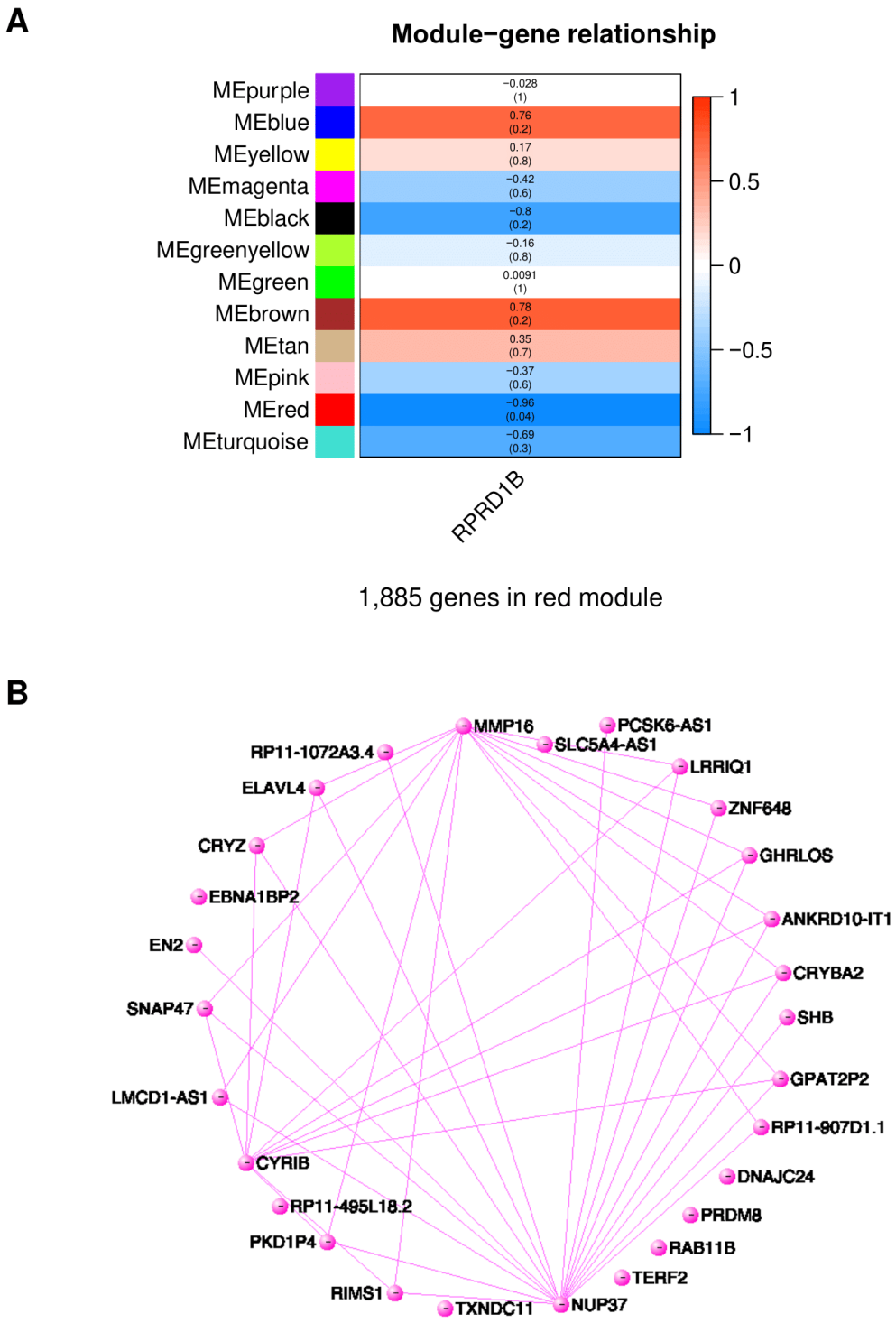


Figure S5. WGCNA analysis of shRPRD1B/Control RNA-Seq data.

(A) Pearson correlation coefficient matrix among module eigengenes (MEs), and RPRD1B status across the analyzed samples. Each cell reports the correlation coefficient with *p*-value. **(B)** Top 30 highly connected genes within the red module (p = 0.04).

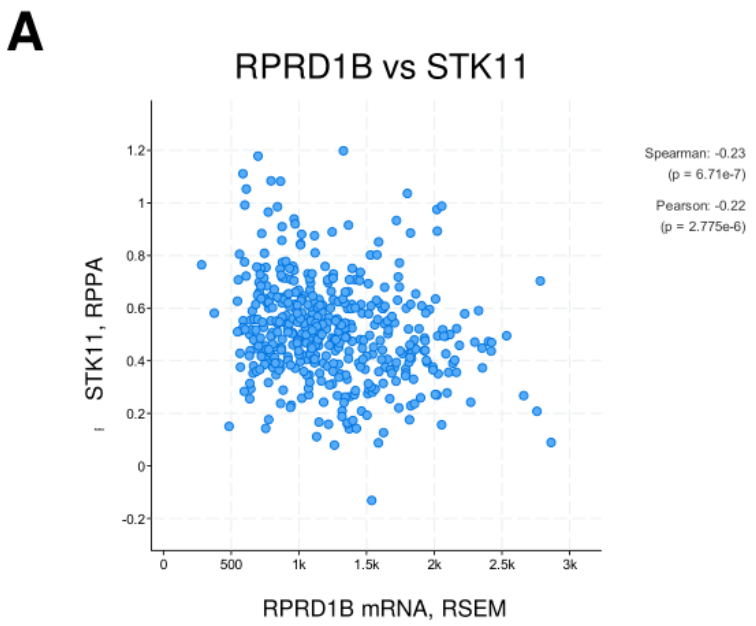


Figure S6. TCGA database analysis of RPRD1B alterations in colorectal cancer

(A) Correlation between *RPRD1B* mRNA level (RSEM, z-scores) and STK11 protein level (RPPA, z-scores) in colorectal adenocarcinoma samples (n = 464 samples).

Supplementary Table 6: Sequence information of qPCR primers

QPCR primers	
Gene	**Sequence**
SYK	FP: AAAGCCTGGCCACAGAAAGT RP: CTTCTCTCTGGGGGCCTTTG
ATF6B	FP: TGGGGTCTGCAGAACAGCAC RP: CTCAGAGGGGCTGACATCCA
ARHGAP27	FP: GAATGTGCTGGAGCTACGGA RP: GCCCTGAGCAATGGCCTTAT
PRPF31	FP: AGCTTCGGAGAGATCGAGGA RP: CTTGGTGGCCTCGTTTACCT
ACTB	FP:TCAAGATCATTGCTCCTCCTG RP:ACTCGTCATACTCCTGCTTG
STK11	FP: TGGGGTCACCCTCTACAACA RP: TTTCAGCAGGTCAGAGAGCG
P3H4	FP: GGTGAACGGTTGGACGTTAC RP: CTACCTGTGTCGACTGACCTG