Supplemental Information

Two novel RNA-binding proteins identification through computational prediction and experimental validation

Juan Xie[1], Xiaoli Zhang[1], Jinfang Zheng[1], Xu Hong[1], Xiaoxue Tong[1], Xudong Liu[1], Yaqiang Xue[2], Xuelian Wang[3], Yi Zhang[3] and Shiyong Liu[1]*

[1]School of Physics, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

[2]Laboratory for Genome Regulation and Human Health, ABLife Inc., Wuhan, Hubei 430075, China

[3]ABLife BioBigData Institute, Wuhan, Hubei 430075, China

*Correspondence: liushiyong@gmail.com

Figures

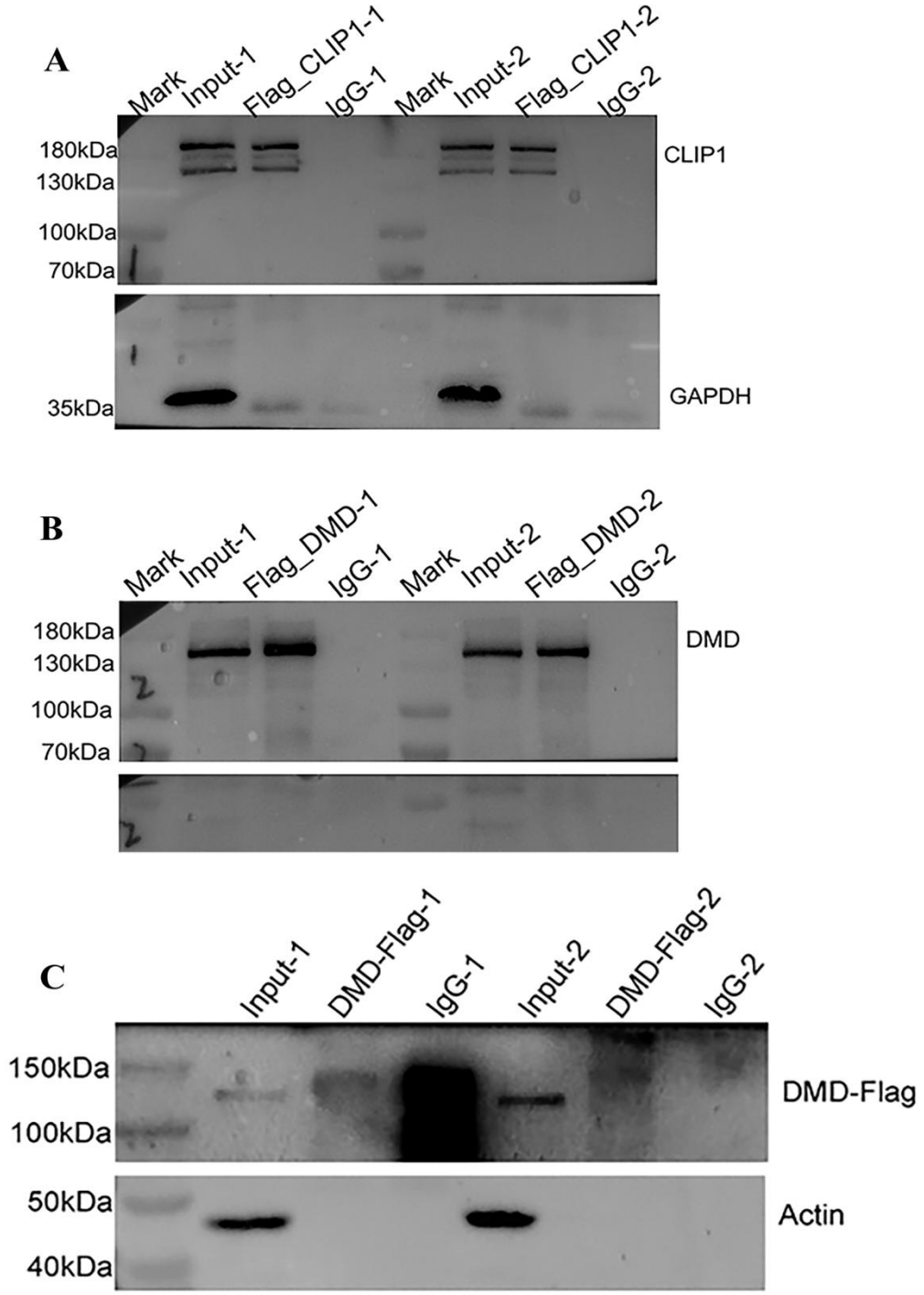

Figure S1. Western blot was used to confirm the immunoprecipitation efficacy of *CLIP1* (iRIP-seq) (A), *DMD* (iRIP-seq) (B) and *DMD* (CLIP-seq) (C) by Flag-antibody (Sigma, F1804) and control IgG-antibody.

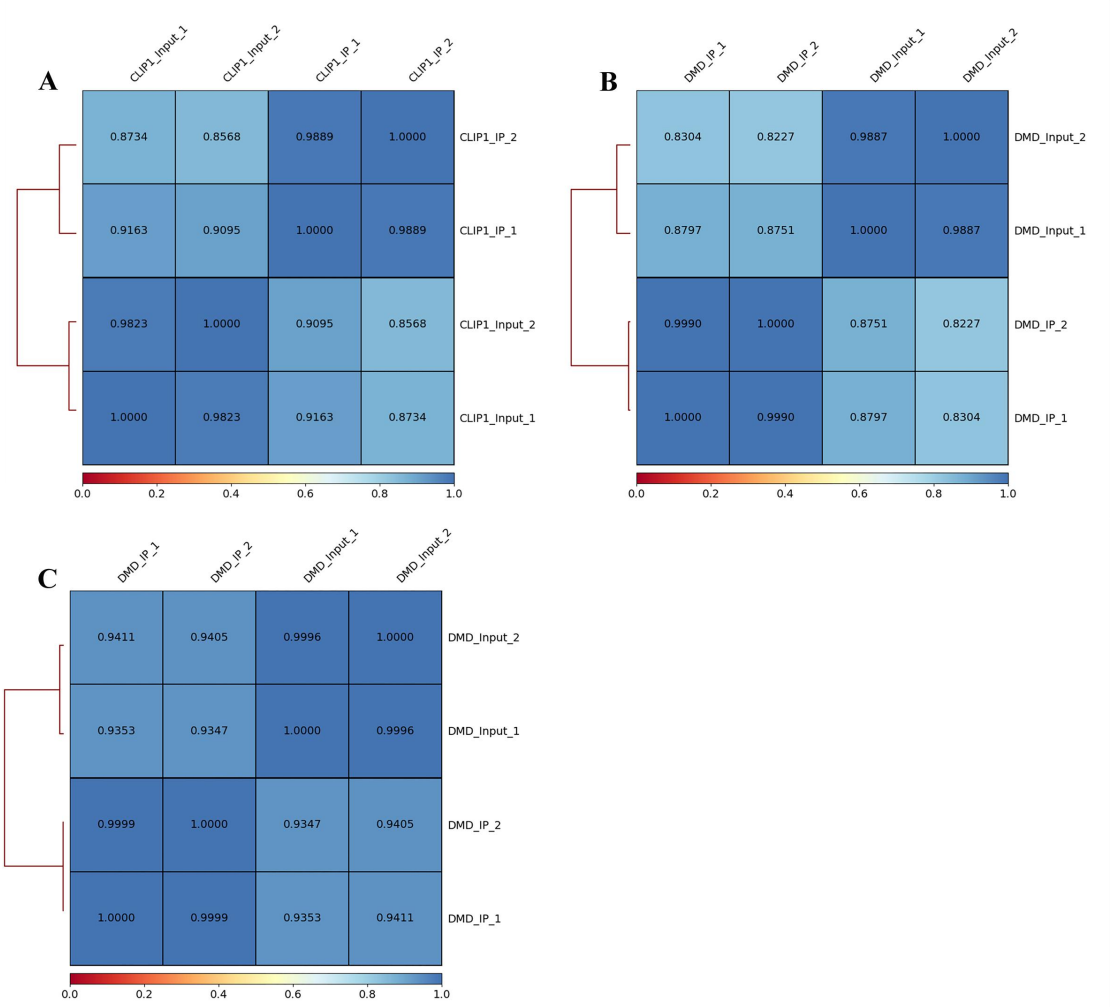

Figure S2. Correlation between IP and input samples. (A), (B) and (C) represent the correlation between the IP and input samples of the repeated experiments of *CLIP1* (iRIP-seq), *DMD* (iRIP-seq) and *DMD* (CLIP-seq), respectively. The results show that the correlation between the IP samples of CLIP1 and *DMD* is greater than 0.98, which fully illustrates the repetition of the experiments. The figure was plotted by deepTools[1].

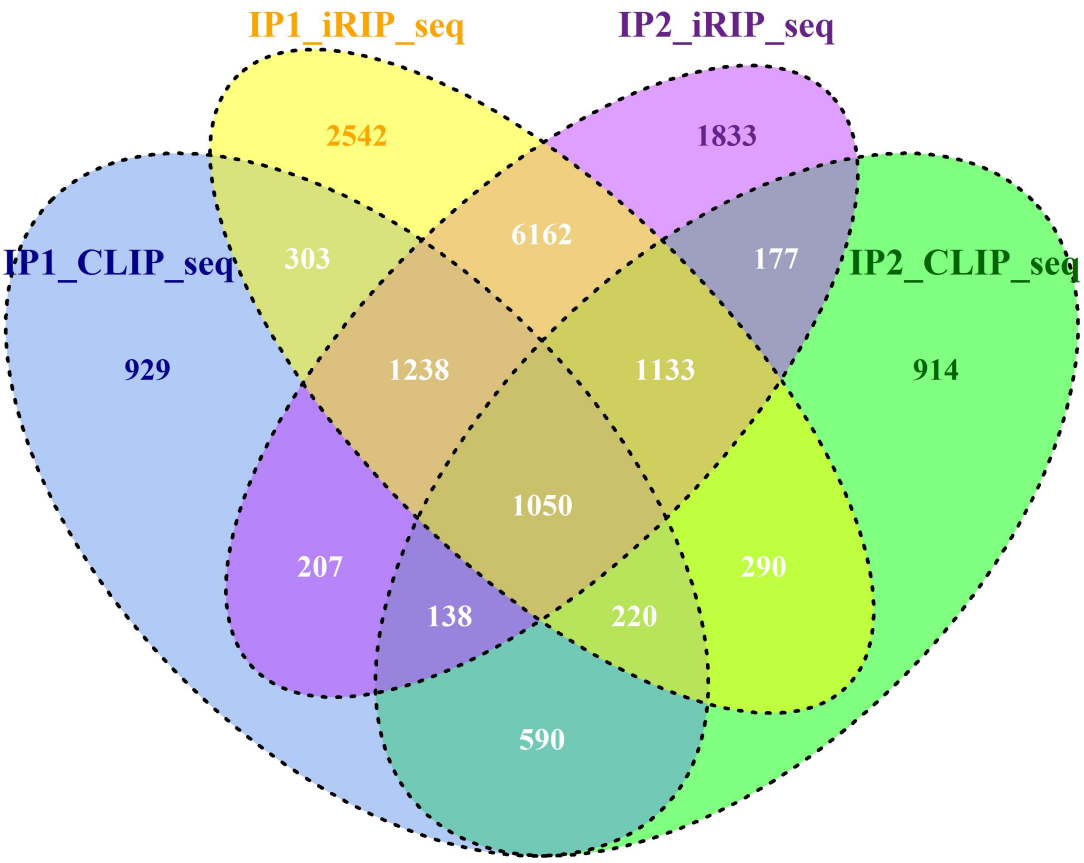

Figure S3. The overlap of genes corresponding to the binding peaks of different experiments. The CIMS results are shown.

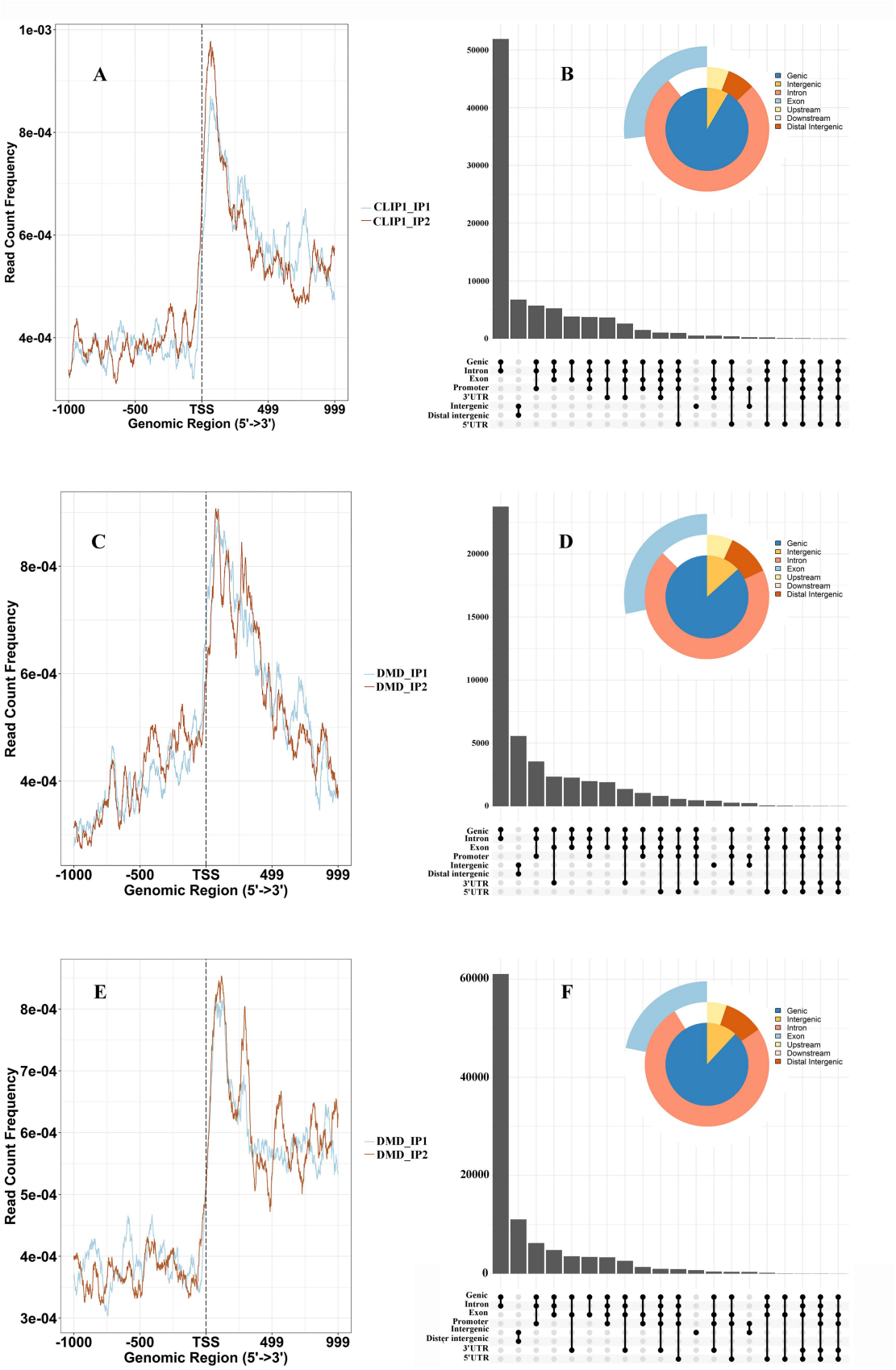

Figure S4. The distribution of peaks near TSS. The density of the peaks are shown in 1 kb upstream and downstream of the TSS for *CLIP1* (iRIP-seq)(A), *DMD*

(CLIP-seq) (C) and *DMD* (iRIP-seq)(E) respectively. (B), (D) and (F) describe the distribution of peaks of *CLIP1* (iRIP-seq), *DMD* (CLIP-seq) and *DMD* (iRIP-seq) respectively on the genome of 1 kb upstream and downstream of TSS. The pie charts in the upper right corner indicate the intersection of peaks on the genome. This corresponds to the black dotted line in the figure, and the number of intersections corresponds to the histogram portion in the figure. CIMS analysis results are shown.

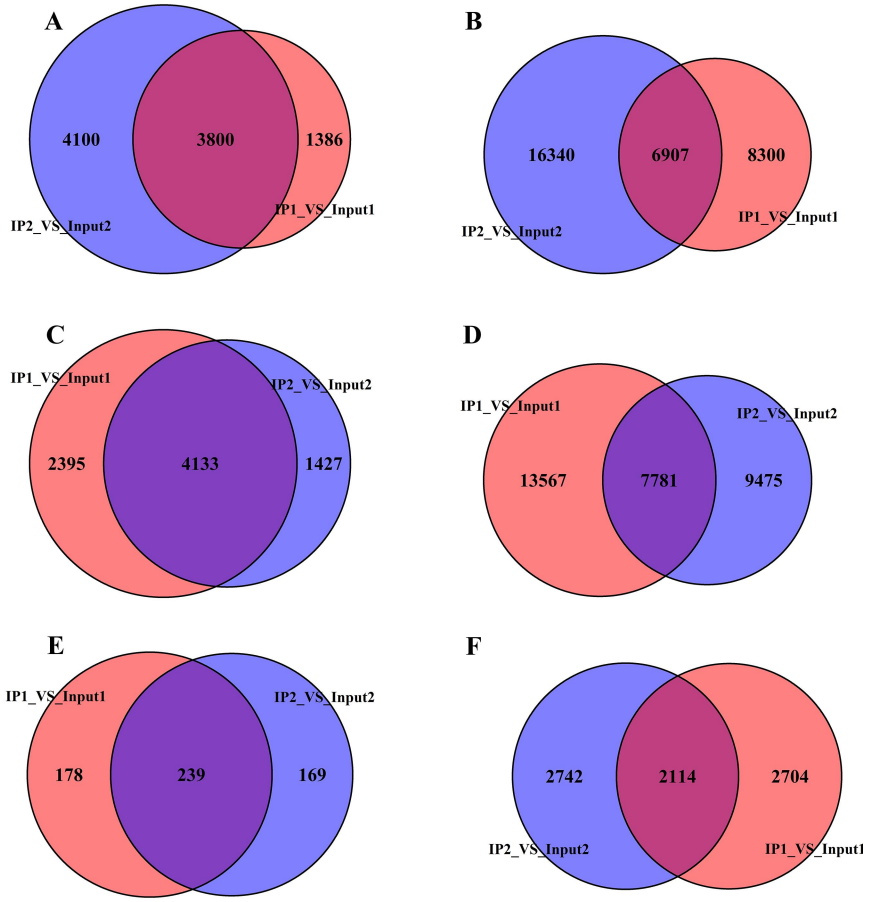

Figure S5. Overlapped binding peaks and related genes between repeated experiments. (A), (C) and (E) represent the number of genes present in two replicate experiments of *CLIP1* (iRIP-seq), *DMD* (iRIP-seq) and *DMD* (CLIP-seq), respectively. (B), (D) and (F) evaluated the consistency of peaks between two replicate samples of *CLIP1* (iRIP-seq), *DMD* (iRIP-seq) and *DMD* (CLIP-seq), which was calculated by BEDTools[2].

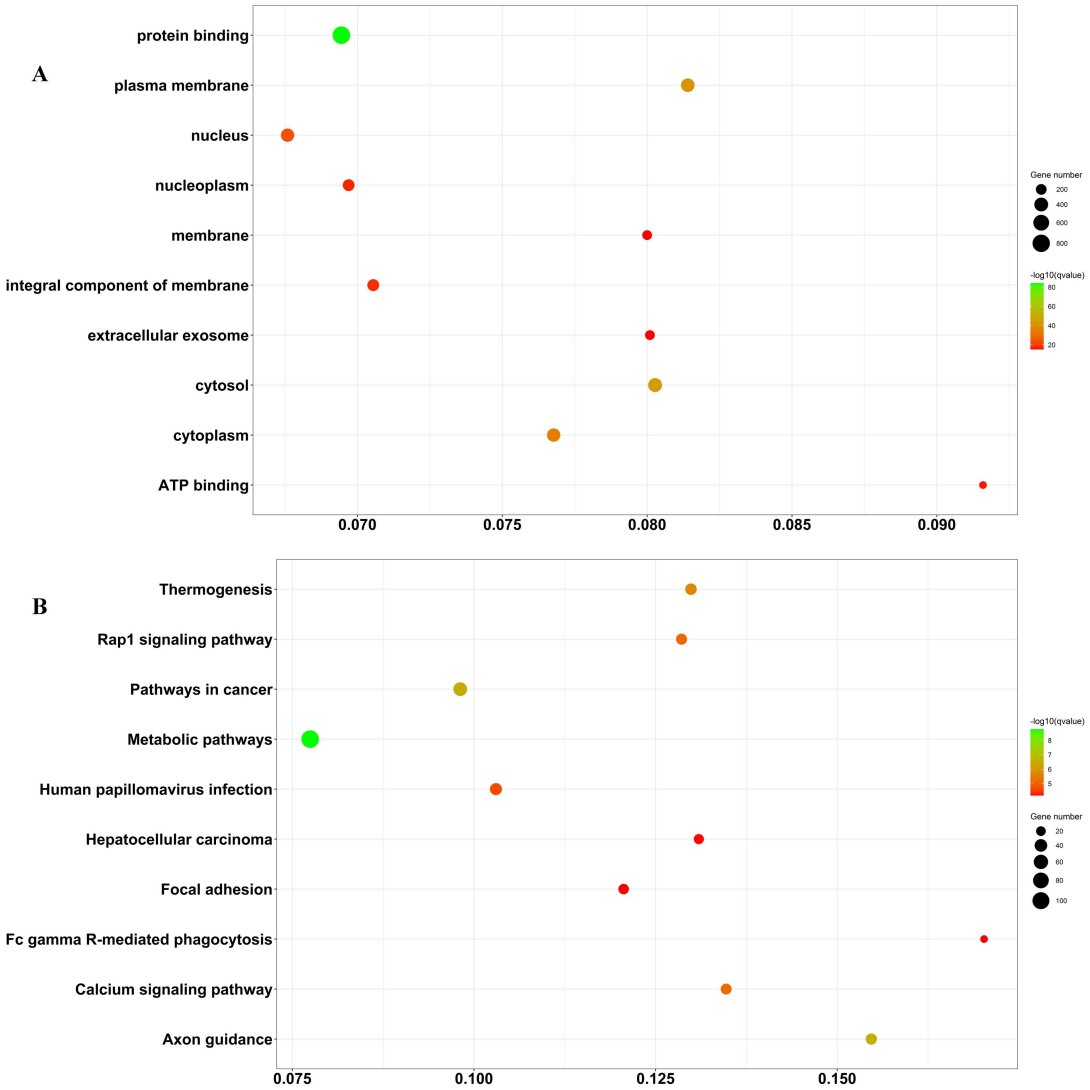

Figure S6. Computational GO and KEGG enrichment analysis for *DMD* (CLIP-seq). *DMD*-binding peak-related genes are used for GO (A) and KEGG (B) analysis, which are calculated by KOBAS 3.0[3]. The X axis represents the number of GO terms. The Y axis represents the type of GO or KEGG term. The genes are obtained by peak calling software CIMS, and the genes corresponding to peak obtained by Piranha peak caller are eliminated.

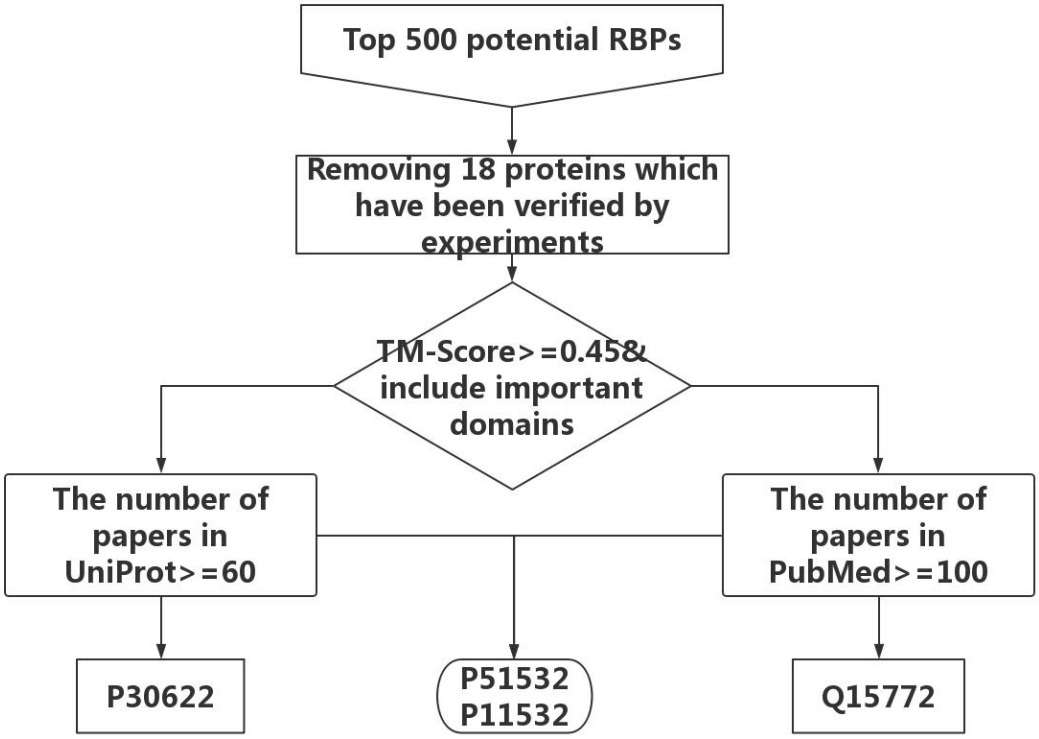

Figure S7. The flowchart describes the data preparation.

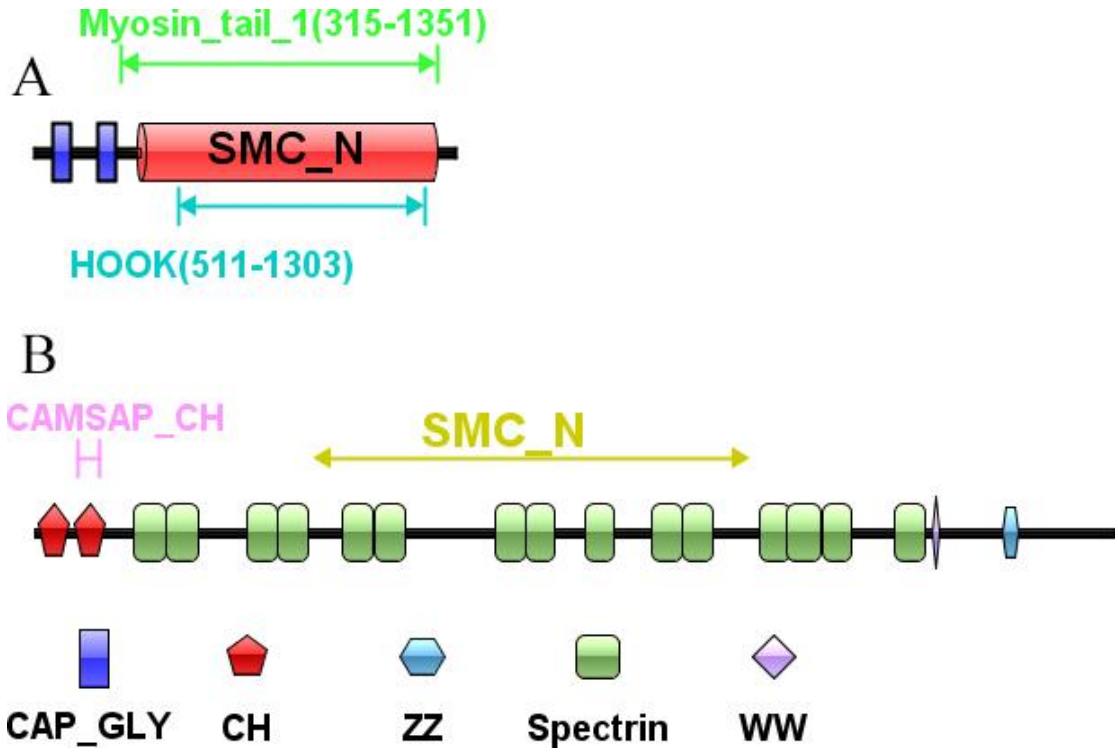

Figure S8. The domains pattern of potential target RBPs. RNA binding domains are shown. The domain pattern of *CLIP1* (P30622) (A) and *DMD* (P11532) (B).

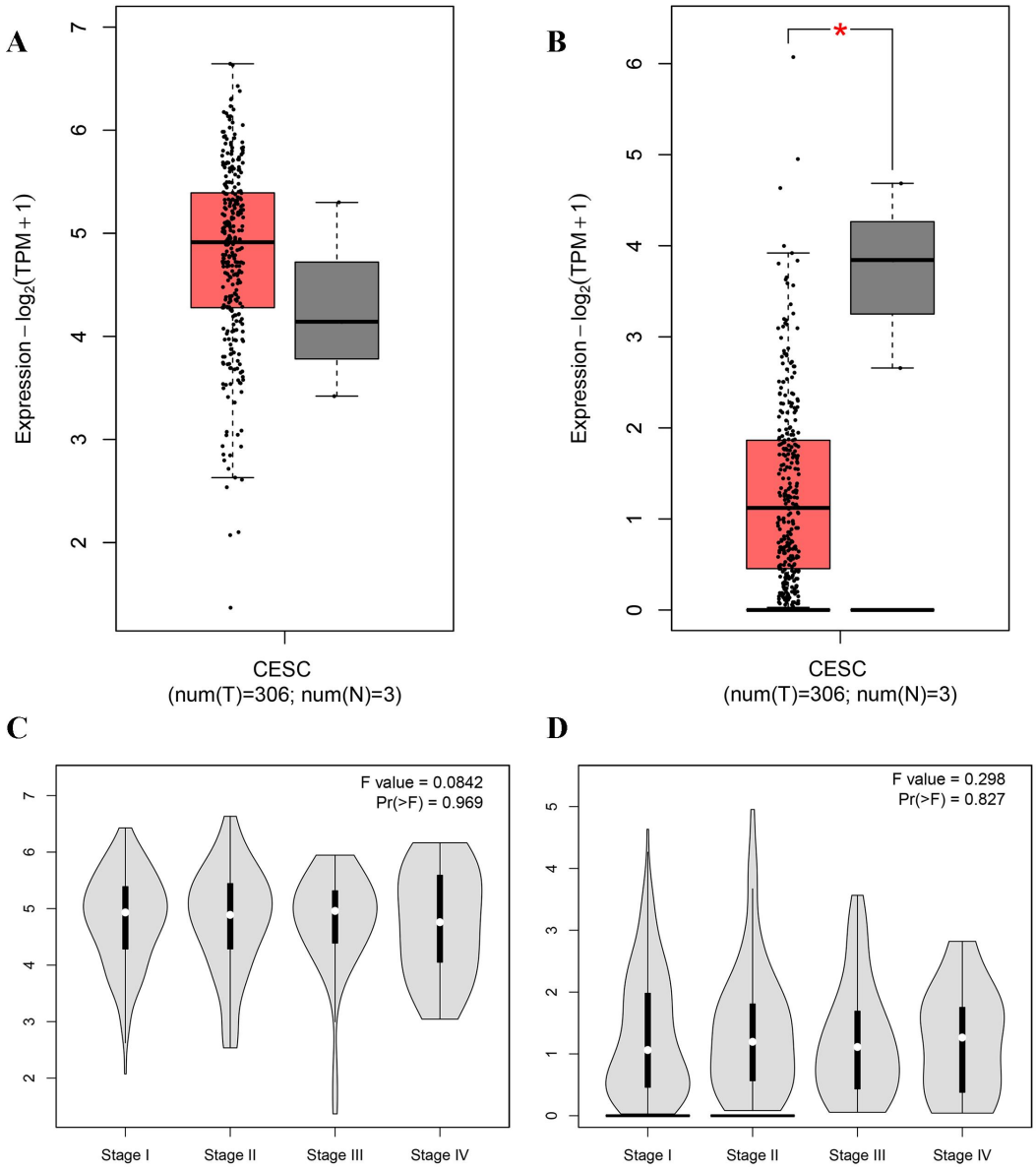

Figure S9. The expression of CLIP1 and DMD in cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC).

Reference

[1] F. Ramirez, D.P. Ryan, B. Gruning, V. Bhardwaj, F. Kilpert, A.S. Richter, S. Heyne, F. Dundar, T. Manke, deepTools2: a next generation web server for deep-sequencing data analysis, Nucleic Acids Res, 44 (2016) W160-165.

[2] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics, 26 (2010) 841-842.

[3] D. Bu, H. Luo, P. Huo, Z. Wang, S. Zhang, Z. He, Y. Wu, L. Zhao, J. Liu, J. Guo, S. Fang, W. Cao, L. Yi, Y. Zhao, L. Kong, KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis, Nucleic Acids Research, 49 (2021) W317-W325.