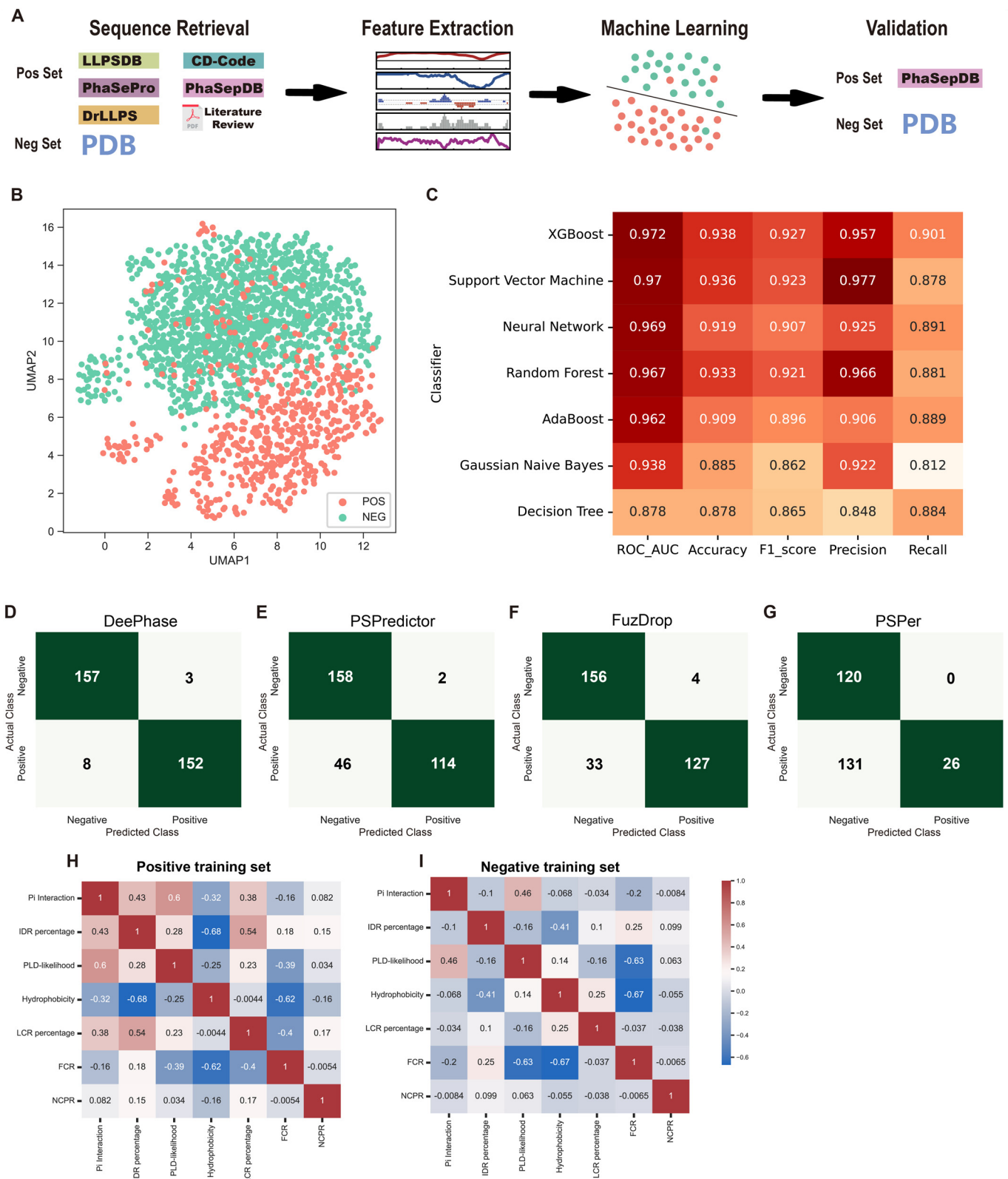


Expanded View Figures

Figure EV1. Phase separation predictor performance.

(A) Workflow for constructing the phase separation predictor. (B) 2D vector projection of training datasets prior to one-sided selection undersampling by UMAP. (C) Efficacy of seven clustering algorithms on a consistent training set. (D–G) Confusion matrices for external dataset predictions by (D) DeePhase, (E) PSPredictor, (F) FuzDrop, and (G) PSPer. FuzDrop's cutoff is 0.6, while the others are 0.5, as suggested by their respective studies. PSPer could not process 5 sequences for unspecified reasons, which are excluded from the matrix. Deep green indicates true outcomes, while light green indicates false outcomes. (H, I) Correlation coefficient of seven features in (H) positive training set and (I) negative training set. Source data are available online for this figure.



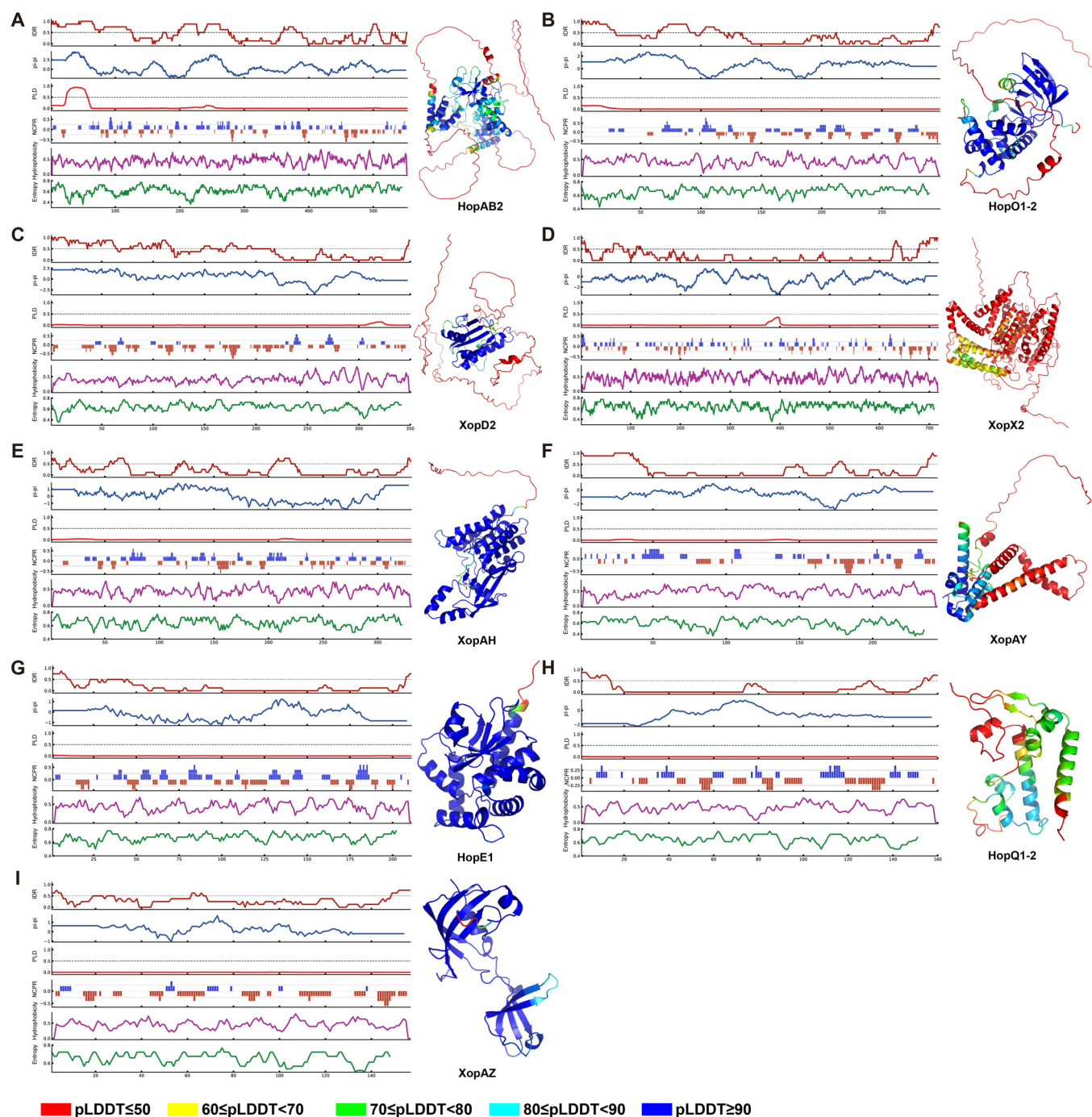


Figure EV2. Effector protein feature extraction and structural prediction.

(A-I) Feature extraction and AlphaFold2 structural prediction for four positive candidates: (A) HopAB2, (B) HopO1-2, (C) XopD2, (D) XopX2, (E) XopAH, (F) XopAY, (G) HopE1, (H) HopQ1-2 and (I) XopAZ. Features, from top to bottom, include IDR, pi interaction, prion-like domain likelihood, net charge per residues, hydrophobicity, and Shannon entropy. AlphaFold2-predicted structures are color-coded by predicted local-distance difference test (pLDDT) shown at the bottom of image. Structure image sizes are not to scale. This relates to Fig. 3. Source data are available online for this figure.

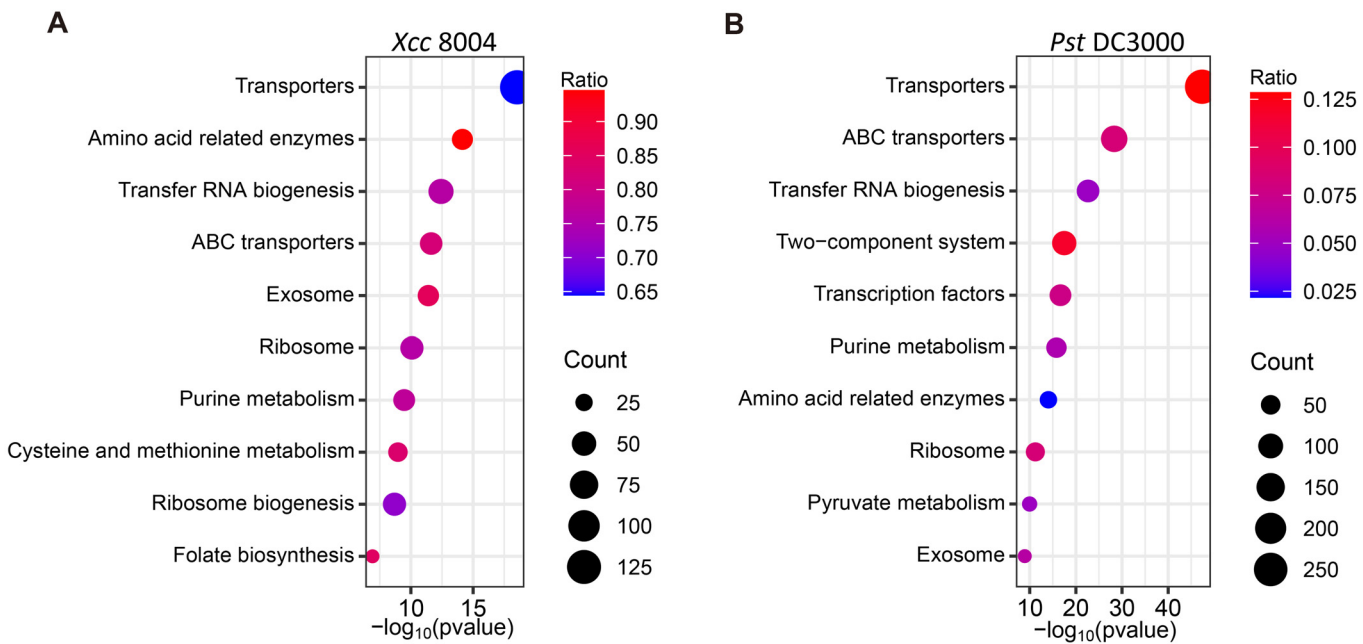
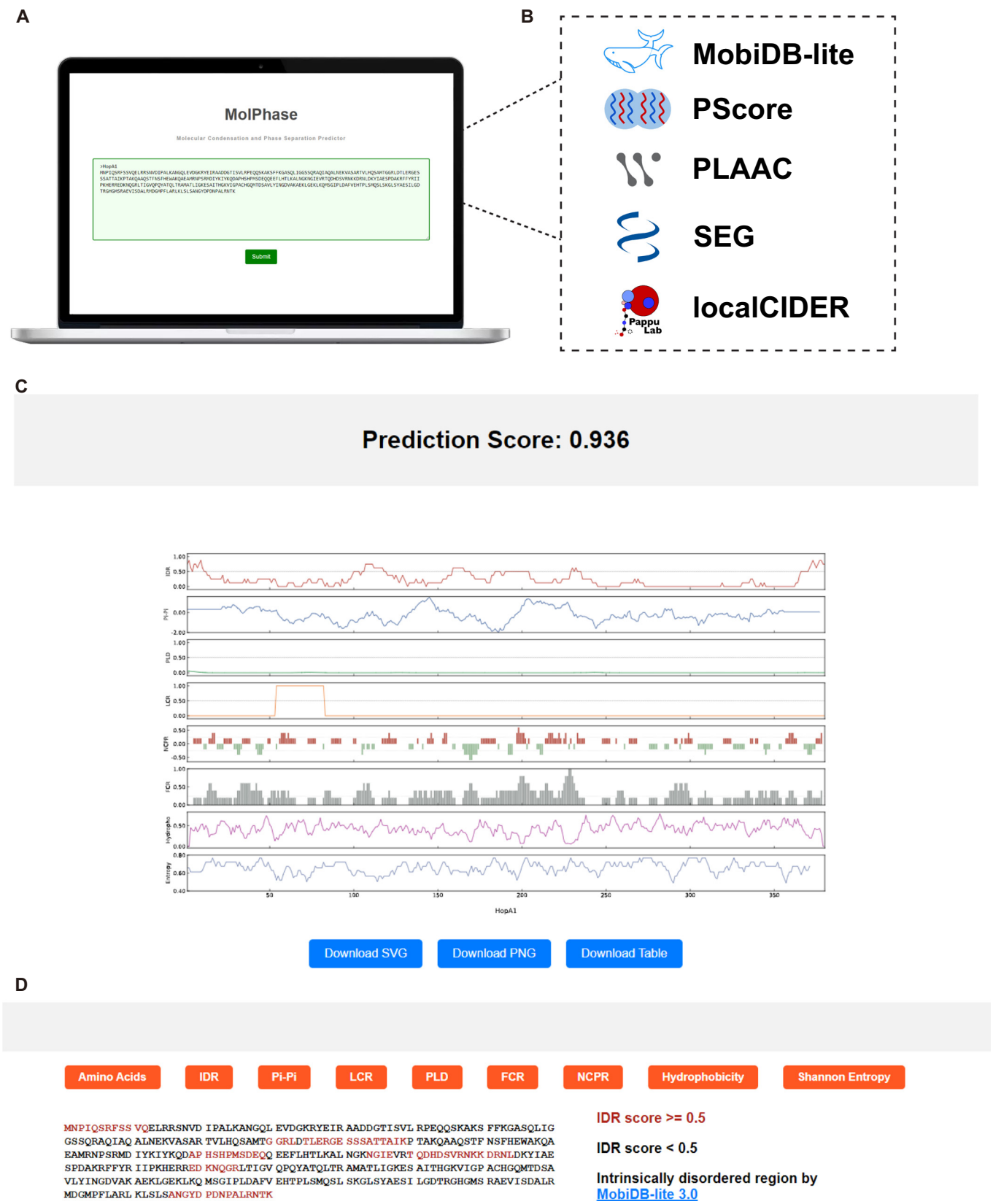


Figure EV3. KEGG analysis for phase separation negative proteins.

(A, B) KEGG enrichment analysis for proteins with phase separation prediction scores below 0.1 in (A) *Xcc* 8004 and (B) *Pst* DC3000. The x axis shows the $-\log_{10} P$ value for the Fisher exact test, the y axis lists enrichment items, the ratio indicates the percentage of the entire pathway enriched, and the count shows the number of enriched items. Source data are available online for this figure.



**Figure EV4. MolPhase online predictor interface.**

(A) A screenshot of the MolPhase predictor, displaying the input sequence for the effector HopA1. (B) Tools utilized for illustrating features in (C). IDR was determined by Mobidb-lite, pi-pi contacts by PScore, LCR by SEG, and PLD by PLAAC. NCPR, FCR, hydrophobicity, and Shannon Entropy were assessed by localCIDER. The displayed score is the phase separation prediction score, ranging from 0 to 1. (C) Specific features aiding in predicting potential phase separation proteins, using HopA1 as an example. Features, from top to bottom, are IDR, pi interaction, PLD, LCR, NCPR, FCR, hydrophobicity, and Shannon Entropy. (D) Features illustrate in the amino acid sequence view, and feature higher than the threshold score will be highlighted.