Legends for the supplementary materials

Supplementary material S1. (A) Different methods in APRICOT involved in the feature based scoring of the predicted domains (B) Different modules of APRICOT for the additional annotations of selected proteins.

Supplementary Figure S2. Resulting table obtained upon the analysis of the PTBP1 by APRICOT. The columns in the table have been explained in three parts: S2 (A) annotation of the protein retrieved from UniProt knowledgebase, S2 (B) annotation of the domains in the protein selected by APRICOT, and S2 (C) statistical output for each parameter associated with the domain predictions.

Supplementary Table S3. Description of different RBP and non-RBP datasets used in this study for training and evaluation of APRICOT pipeline.

Supplementary Table S4. Analysis of SwissProt-positive and SwissProt-negative protein datasets selected from SwissProt (UniProt knowledgebase) by APRICOT using different parameters. The values for the different assessment measures (SN, SP, ACC, MCC and F measure) calculated for SwissProt-positive and SwissProt-negative at systematically varying cut-offs for different parameters associated with domain predictions.

Supplementary Figure S5. The Receiver Operating Characteristic (ROC) curves and their Area Under Curves (AUC) for each parameter and combinations of parameters to assess their marginal contributions to the overall accuracy of APRICOT with which it identifies RNA-binding proteins.

Supplementary Table S6. Results obtained from APRICOT analysis of the human RBPs selected from the experimental studies.

Supplementary Table S7. Results obtained from APRICOT analysis of the RBPs from E. coli (strain K-12) selected from the experimental studies and extended by proteins selected from UniProt database using the Gene Ontology term for RNA-binding (GO:0003723).

Supplementary Table S8. Table showing an overview of the identification and characterization of kinase proteins from E. coli (strain K-12) using APRICOT analysis pipeline.

Supplementary Table S9. Table showing a synopsis of 3657 RBPs from the NBench study, which were analyzed by APRICOT to identify an overlap between the predicted RBD sites and the known RNA-binding residues were recorded. This analysis demonstrates that APRICOT can be complemented with the tools for the prediction of RNA-binding sites.