Supplementary Figure 1: The scores based on inferred energy motif are highly correlated with the binding energy; x-axis: the scores based on inferred energy motif; y-axis: The predicted binding energy.

Supplementary Figure 2: The BPS motif from MM based on different start points and different training data sets; a. The start point is the energy motif with the same frequencies of four nucleotides at BP and the training data is enriched data; b. The start point is random and the training data is enriched data; c. The start point is the energy motif with the same frequencies of four nucleotides at BP and the training data is un-enriched data; d. The start point is random and the training data is un-enriched data.

Supplementary Figure 3: The PPT region includes a small set of U2AF65 preferred octanucleotides.
Supplementary Figure 4: The performances of BPP, SVM-BPfinder and HSF based on the experimentally verified BPs (a) The performance of three methods using their suggested way. Reg corresponds to the regular cases and the TNA corresponds to the cases with TNA structure (no error tolerant); (b) The ROC curves of three methods to predict the correct introns (no error tolerant); The suffix “A” corresponds to the TNA cases. (c) The ROC curves of the three methods to predict the correct BP sites for the regular cases; “rank” and “sc” respectively corresponds to rank based and score based ways; (d) The ROC curves of the three methods to predict the correct BP sites for the TNA cases; “rank” and “sc” respectively corresponds to rank based and score based ways.
Supplementary Figure 5: The performances of BPP, SVM-BPfinder and HSF based on the high-throughput sequencing data (a) The performance of three methods using their suggested way, Reg corresponds to the regular cases and the TNA corresponds the cases with TNA structure (1 nucleotide error tolerant); (b) The ROC curves of three methods to predict the correct introns; The suffix "A" corresponds to the TNA cases (1 nucleotide error tolerant); (c) The performance of three methods using their suggested way, Reg corresponds to the regular cases and the TNA corresponds the cases with TNA structure (no error tolerant); (d) The ROC curves of three methods to predict the correct introns; The suffix "A" corresponds to the TNA cases (no error tolerant); (e) The ROC curves of the three methods to predict the correct BP sites for the regular cases; "rank" and "sc" respectively corresponds to rank based and score based ways; (f) The ROC curves of the three methods to predict the correct BP sites for the TNA cases; "rank" and "sc" respectively corresponds to rank based and score based ways.
Supplementary Figure 6: The distribution of relative end frequencies along the BP distance to 3′SS