

Please refer to 'AARS-K562-hg38-CLIP.tsv' to name(RBP-cellLine-genome-CLIP.tsv) and upload the file. Please provide 15000 samples, including 5000 positive samples and 10000 negative samples.

The input data (.tsv format file, Table 1) for each RBP PrismNet model contains six columns. The first column indicates the sample type of data, "A" for training data and "B" for testing data. The second column is the transcript coordinate of the binding sites, split by "|". Users can add other information for samples in this column. The third and fourth columns contain the sequence and icSHAPE score (split by ","). The fifth column contains the binding scores, which are normalized scores of the peaks signal intensities in the CLIP data file. The binding scores in the negative samples are set to -1. Note that the binding scores are not used as the input of model, which is the sixth column, the label column that shows the sample label for the data, with 1 for a positive sample (a true binding site) and 0 for a negative sample (a false binding site).

Table1

Sample type	Transcript coordinate	Sequence	icSHAPE score	Binding score	Label
B	ENST00000389680 76 176	AC...A	0.585,0.113,...,0.63	0.98	1
A	ENST00000367142 244 344	GC...C	0.124,0.252,...,1.0	0.72	1
B	ENST00000372686 877 977	TT...A	0.592,0.476,...,0.41	-1	0
A	ENST00000400561 319 419	TT...G	0.372,0.124,...,0.0	-1	0