Number of top k-mers	AUC ($h=3$, $AUC_{SVM}=0.80$)	AUC ($h=4$, $AUC_{SVM}=0.90$)	AUC ($h=5$, $AUC_{SVM}=0.93$)
100	0.65	0.68	0.79
200	0.68	0.72	0.80
1000	0.80	0.81	0.82

Table S3. Cross-validation of the performance of a predictor based on the difference between the number of occurrences of the top and bottom k-mers in a balanced set of 1000 positive and negative sequences, compared with the SVM performance.