

SUPPLEMENTAL MATERIAL 8

for the paper “Position dependencies in transcription factor binding sites”

A. Tomovic and E. J. Oakeley

Table Sup8-1. The number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) results for the new and old scoring functions using different threshold values.

THRESHOLD	new scoring function S'_{new}				old scoring function S'_{old}			
	TP	TN	FP	FN	TP	TN	FP	FN
0.7	47	669,315	138	0	47	640,730	28,723	0
0.8	46	669,362	91	1	46	666,112	3,341	1
0.9	46	669,435	18	1	32	669,326	127	15

For this experiment, we used the dataset available at:

www.fmi.ch/members/andrija.tomovic/expTPTN-supp8.zip and
matrixID: MA0041

Table Sup8-2. Average false positive rate per nucleotide and per TF for different methods.

METHOD	Average FP per nucleotide	Average FP per TF
S'_{new} (threshold = 0.7)	0.04	1.32
S'_{new} (threshold = 0.8)	0.013	0.466
S'_{old} (threshold = 0.7)	0.27	9.453
S'_{old} (threshold = 0.8)	0.05	1.733
ConSite (threshold = 0.7)	0.029	6.33
ConSite (threshold = 0.8)	0.03	1.04
PATSER	0.184	6.453

For this experiment, we used the dataset available at:

www.fmi.ch/members/andrija.tomovic/exp-supp8.zip and
matrixIDs: MA0006, MA0041, MA0048, MA0052, MA0054, MA0065, MA0066,
MA0083, MA0086, MA0091, MA0097, MA0114, MA0116, MA0121, and MA0123

Table Sup8-3. Comparison of the predictions from the new scoring function and the ConSite with experimentally verified data.

Gene RefSq ID	New scoring function S'_{new}		ConSite	
	TP	FP	TP	FP
NM_184041	MA0052, MA0055, MA0056, MA0057, MA0079	78	MA0052, MA0055, MA0056, MA0057, MA0079	81
NM_001927	MA0052, MA0055, MA0056, MA0057, MA0079	74	MA0052, MA0055, MA0056, MA0057, MA0079	80
NM_002479	MA0052, MA0056, MA0055, MA0057, MA0079	84	MA0052, MA0056, MA0055, MA0057, MA0079	87
NM_079422	MA0052, MA0055, MA0056, MA0057, MA0079	79	MA0052, MA0055, MA0056, MA0057, MA0079, MA0083	86
NM_003281	MA0052, MA0055, MA0056, MA0057, MA0079	78	MA0052, MA0055, MA0056, MA0057, MA0079	81
NM_000257	MA0055, MA0056, MA0057, MA0079	75	MA0052, MA0055, MA0056, MA0057, MA0079	77
NM_002471	MA0052, MA0055, MA0056, MA0057, MA0079	78	MA0052, MA0055, MA0056, MA0057, MA0079	82
NM_001100	MA0052, MA0055, MA0056, MA0057, MA0079	77	MA0052, MA0055, MA0056, MA0057, MA0079, MA0083	80
NM_005159	MA0052, MA0055, MA0056, MA0057, MA0079	77	MA0052, MA0055, MA0056, MA0057, MA0079	84

Note: MA0055 is not available in our database, because there is no sequence data in the JASPAR database (September, 2006), only profile. However, searching using profile it is possible to find it.