Supplementary materials for MBSTAR : multiple instance learning for predicting specific functional binding sites in microRNA targets

Sanghamitra Bandyopadhyay[1,*], Dip Ghosh[1], Ramkrishna Mitra[2], Zhongming Zhao[2,3,4,5,*]

1 Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India
2 Department of Biomedical Informatics, Vanderbilt University, Nashville, Tennessee, 37232, USA
3 Department of Cancer Biology, Vanderbilt University, Nashville, Tennessee, 37232, USA
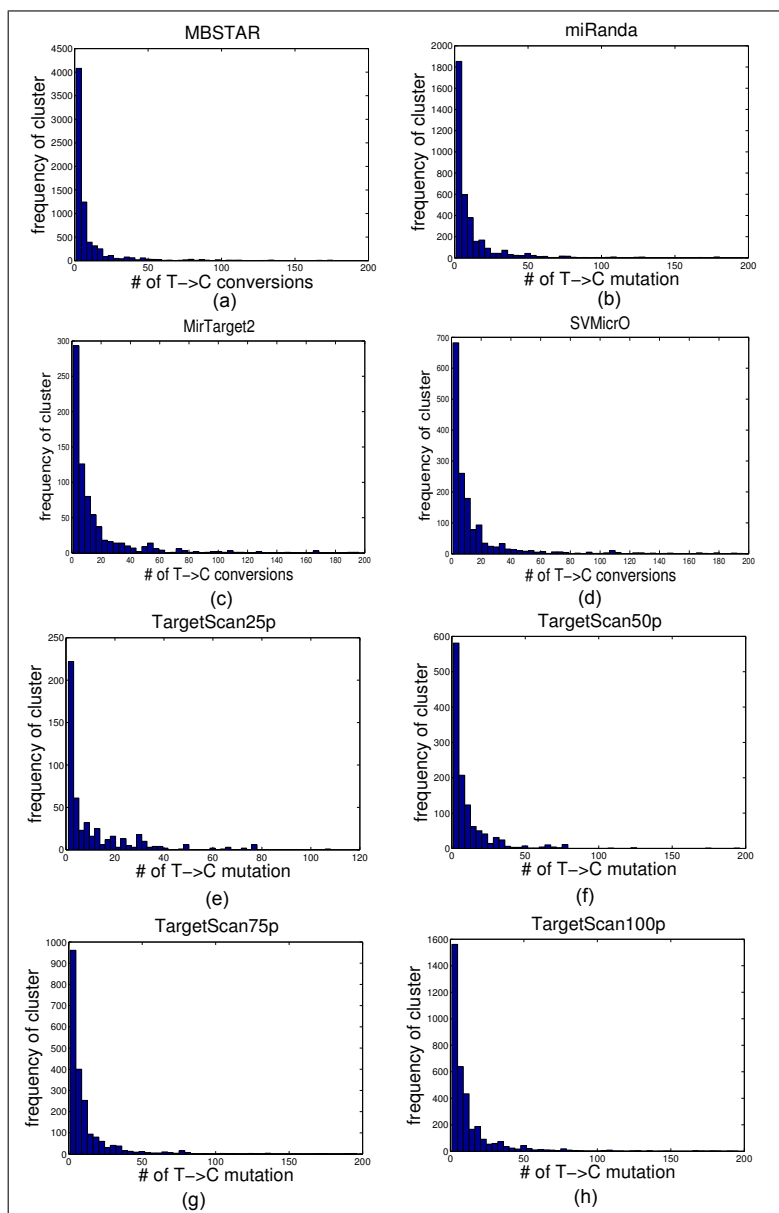4 Department of Psychiatry, Vanderbilt University, Nashville, Tennessee, 37232, USA
5 Center for Quantitative Sciences, Vanderbilt University, Nashville, Tennessee, 37232, USA

* Email Address:
SB: sanghami@isical.ac.in
ZZ: zhongming.zhao@vanderbilt.edu

Figure S1. Distribution of predicted PAR-CLIP gold standard clusters according to (a) Number of T→C conversion by MBSTAR, (b) Number of T→C conversion by miRanda, (c) Number of T→C conversion by MirTarget2, (d) Number of T→C conversion by SVMicrO, (e) Number of T→C conversion by TargetScan25p, (f) Number of T→C conversion by TargetScan50p, (g) Number of T→C conversion by TargetScan75p and (h) Number of T→C conversion by TargetScan100p.

Supplementary Methods

Here we describe that random guessing test to prove the superiority of MBSTAR. First all PBSs are extracted from 9531 validated miRNA-mRNA pairs using only seed site complementarity. We then randomly selected 1000 of such seed sites and test that how many of them overlap with PAR-CLIP clusters. We then randomly select 1000 of MB-STAR predicted FBSs and calculate the number of overlapped clusters with the same PAR-CLIP data set. This experiment is carried out 10,000 time and found that, for only 219 cases, random PBSs have more overlapping than predicted FBSs. So we conclude that MBSTAR can outperform the random approach with p-value less than 0.0219.

Table S1 Feature ranking based on Laplacian score derived from Laplacian score based feature selection method

Feature Index	Feature Description	1− Laplacian Score
3	Frequency of Nucleobase t in flanking site	0.915561414
363	maximum size of Interior Loop	0.898200106
2	Frequency of Nucleobase g in flanking site	0.893687327
4	Frequency of Nucleobase c in flanking site	0.893000146
1	Frequency of Nucleobase a in flanking site	0.887377625
341	MFE of RNA pair	0.880487537
15	Frequency of Nucleobase tt in flanking site	0.876769377
353	maximum size of Hairpin	0.865627287
5	Frequency of Nucleobase aa in flanking site	0.834235876
20	Frequency of Nucleobase cc in flanking site	0.830898806
10	Frequency of Nucleobase gg in flanking site	0.829856343
63	Frequency of Nucleobase ttt in flanking site	0.821761825
343	maximum size of Bulge	0.759142476
13	Frequency of Nucleobase at in flanking site	0.758825637
21	Frequency of Nucleobase aaa in flanking site	0.757034553
7	Frequency of Nucleobase ta in flanking site	0.739103589
255	Frequency of Nucleobase tttt in flanking site	0.725381025
42	Frequency of Nucleobase ggg in flanking site	0.710959477
12	Frequency of Nucleobase cg in flanking site	0.70295739
84	Frequency of Nucleobase ccc in flanking site	0.689674175
351	number of bulge with length 8 or greater	0.682303144
6	Frequency of Nucleobase ga in flanking site	0.665999366
19	Frequency of Nucleobase tc in flanking site	0.664192939
85	Frequency of Nucleobase aaaa in flanking site	0.644962661
9	Frequency of Nucleobase ag in flanking site	0.641590576
371	number of Interior loop with length 8 or greater	0.638535463
61	Frequency of Nucleobase att in flanking site	0.619012514
52	Frequency of Nucleobase ccg in flanking site	0.615575007
11	Frequency of Nucleobase tg in flanking site	0.613252133
14	Frequency of Nucleobase gt in flanking site	0.613241796
16	Frequency of Nucleobase ct in flanking site	0.604236154
17	Frequency of Nucleobase ac in flanking site	0.603085662
55	Frequency of Nucleobase tat in flanking site	0.602498918
44	Frequency of Nucleobase cgg in flanking site	0.594021664
31	Frequency of Nucleobase tta in flanking site	0.589109348
83	Frequency of Nucleobase tcc in flanking site	0.577646247

Table 1 – *Continued from previous page*

Feature Index	Feature Description	1− Laplacian Score
26	Frequency of Nucleobase gga in flanking site	0.567375701
23	Frequency of Nucleobase taa in flanking site	0.560677186
53	Frequency of Nucleobase aat in flanking site	0.557663507
8	Frequency of Nucleobase ca in flanking site	0.539738844
38	Frequency of Nucleobase gag in flanking site	0.531748313
59	Frequency of Nucleobase tgt in flanking site	0.531501302
253	Frequency of Nucleobase attt in flanking site	0.526975171
29	Frequency of Nucleobase ata in flanking site	0.526292668
127	Frequency of Nucleobase ttta in flanking site	0.523095294
41	Frequency of Nucleobase agg in flanking site	0.513393093
70	Frequency of Nucleobase gac in flanking site	0.510832228
18	Frequency of Nucleobase gc in flanking site	0.506982195
81	Frequency of Nucleobase acc in flanking site	0.501339427
80	Frequency of Nucleobase ctc in flanking site	0.501213518
172	Frequency of Nucleobase cggg in flanking site	0.498602254
37	Frequency of Nucleobase aag in flanking site	0.496192827
212	Frequency of Nucleobase cccg in flanking site	0.485067566
68	Frequency of Nucleobase cct in flanking site	0.484401333
340	Frequency of Nucleobase cccc in flanking site	0.484164215
361	number of Hairpin with length 8 or greater	0.478485471
213	Frequency of Nucleobase aaat in flanking site	0.478415283
58	Frequency of Nucleobase ggt in flanking site	0.477733019
46	Frequency of Nucleobase gtg in flanking site	0.475983857
67	Frequency of Nucleobase tct in flanking site	0.474215388
170	Frequency of Nucleobase gggg in flanking site	0.473449632
339	Frequency of Nucleobase tccc in flanking site	0.467316041
82	Frequency of Nucleobase gcc in flanking site	0.459745934
223	Frequency of Nucleobase ttat in flanking site	0.451717443
78	Frequency of Nucleobase gtc in flanking site	0.451621103
25	Frequency of Nucleobase aga in flanking site	0.451072015
64	Frequency of Nucleobase ctt in flanking site	0.449470288
79	Frequency of Nucleobase ttc in flanking site	0.448432561
36	Frequency of Nucleobase cca in flanking site	0.438928779
276	Frequency of Nucleobase ccct in flanking site	0.43868072
234	Frequency of Nucleobase gggt in flanking site	0.438434383
180	Frequency of Nucleobase ccgg in flanking site	0.436700677
247	Frequency of Nucleobase tatt in flanking site	0.435310952
62	Frequency of Nucleobase gtt in flanking site	0.433858046

Table 1 – *Continued from previous page*

Feature Index	Feature Description	1− Laplacian Score
106	Frequency of Nucleobase ggga in flanking site	0.432623855
47	Frequency of Nucleobase ttg in flanking site	0.431227982
87	Frequency of Nucleobase taaa in flanking site	0.431000278
72	Frequency of Nucleobase cac in flanking site	0.428639998
33	Frequency of Nucleobase aca in flanking site	0.426705292
22	Frequency of Nucleobase gaa in flanking site	0.426029529
28	Frequency of Nucleobase cga in flanking site	0.418504685
154	Frequency of Nucleobase ggag in flanking site	0.417844891
245	Frequency of Nucleobase aatt in flanking site	0.415687455
282	Frequency of Nucleobase ggac in flanking site	0.406929734
336	Frequency of Nucleobase ctcc in flanking site	0.401443411
324	Frequency of Nucleobase cctc in flanking site	0.397883898
166	Frequency of Nucleobase gagg in flanking site	0.397734672
326	Frequency of Nucleobase gacc in flanking site	0.395349065
74	Frequency of Nucleobase ggc in flanking site	0.395107569
292	Frequency of Nucleobase ccac in flanking site	0.393046702
51	Frequency of Nucleobase tcg in flanking site	0.392949051
169	Frequency of Nucleobase aggg in flanking site	0.38926036
362	number of Interior Loop	0.388808638
24	Frequency of Nucleobase caa in flanking site	0.386005791
330	Frequency of Nucleobase ggcc in flanking site	0.385176554
49	Frequency of Nucleobase acg in flanking site	0.382741065
93	Frequency of Nucleobase ataa in flanking site	0.379044653
319	Frequency of Nucleobase tttc in flanking site	0.37733167
60	Frequency of Nucleobase cgt in flanking site	0.376830617
45	Frequency of Nucleobase atg in flanking site	0.375840934
149	Frequency of Nucleobase aaag in flanking site	0.374169828
239	Frequency of Nucleobase ttgt in flanking site	0.372094503
256	Frequency of Nucleobase cttt in flanking site	0.369306634
221	Frequency of Nucleobase atat in flanking site	0.369078771
76	Frequency of Nucleobase cgc in flanking site	0.368044705
238	Frequency of Nucleobase gtgt in flanking site	0.367175494
352	number of Hairpin	0.364155665
43	Frequency of Nucleobase tgg in flanking site	0.361984266
69	Frequency of Nucleobase aac in flanking site	0.360236195
101	Frequency of Nucleobase aaga in flanking site	0.358324472
271	Frequency of Nucleobase ttct in flanking site	0.358291276
30	Frequency of Nucleobase gta in flanking site	0.358070462

Table 1 – *Continued from previous page*

Feature Index	Feature Description	1− Laplacian Score
191	Frequency of Nucleobase tttg in flanking site	0.356059473
119	Frequency of Nucleobase tata in flanking site	0.35581505
334	Frequency of Nucleobase gtcc in flanking site	0.355438539
198	Frequency of Nucleobase gacg in flanking site	0.355028267
314	Frequency of Nucleobase ggtc in flanking site	0.354717391
187	Frequency of Nucleobase tgtg in flanking site	0.354335047
284	Frequency of Nucleobase cgac in flanking site	0.35313783
265	Frequency of Nucleobase agct in flanking site	0.352370418
88	Frequency of Nucleobase caaa in flanking site	0.351851151
337	Frequency of Nucleobase accc in flanking site	0.351850663
254	Frequency of Nucleobase gttt in flanking site	0.351790602
86	Frequency of Nucleobase gaaa in flanking site	0.351641345
338	Frequency of Nucleobase gccc in flanking site	0.350235336
56	Frequency of Nucleobase cat in flanking site	0.349234924
57	Frequency of Nucleobase agt in flanking site	0.346893115
251	Frequency of Nucleobase tgtt in flanking site	0.344207301
97	Frequency of Nucleobase acaa in flanking site	0.343219044
117	Frequency of Nucleobase aata in flanking site	0.338737902
95	Frequency of Nucleobase ttaa in flanking site	0.336683681
35	Frequency of Nucleobase tca in flanking site	0.335286251
316	Frequency of Nucleobase cgtc in flanking site	0.333202835
295	Frequency of Nucleobase tagc in flanking site	0.328922182
323	Frequency of Nucleobase tctc in flanking site	0.327565675
259	Frequency of Nucleobase tctt in flanking site	0.325274796
206	Frequency of Nucleobase gtcg in flanking site	0.322921181
328	Frequency of Nucleobase cacc in flanking site	0.318595421
116	Frequency of Nucleobase ccga in flanking site	0.318266191
148	Frequency of Nucleobase ccca in flanking site	0.317167227
277	Frequency of Nucleobase aaac in flanking site	0.315942353
108	Frequency of Nucleobase cgga in flanking site	0.31520999
298	Frequency of Nucleobase gggc in flanking site	0.314128542
150	Frequency of Nucleobase gaag in flanking site	0.31293806
332	Frequency of Nucleobase cgcc in flanking site	0.310172492
123	Frequency of Nucleobase tgta in flanking site	0.307394972
342	number of bulge	0.305282409
183	Frequency of Nucleobase tatg in flanking site	0.30521312
40	Frequency of Nucleobase cag in flanking site	0.30374293
27	Frequency of Nucleobase tga in flanking site	0.302103065

Table 1 – *Continued from previous page*

Feature Index	Feature Description	1− Laplacian Score
113	Frequency of Nucleobase acga in flanking site	0.300549317
71	Frequency of Nucleobase tac in flanking site	0.30045886
209	Frequency of Nucleobase accg in flanking site	0.299789604
65	Frequency of Nucleobase act in flanking site	0.299751406
174	Frequency of Nucleobase gtgg in flanking site	0.299361123
210	Frequency of Nucleobase gccg in flanking site	0.299185544
39	Frequency of Nucleobase tag in flanking site	0.297555386
320	Frequency of Nucleobase cttc in flanking site	0.297123214
89	Frequency of Nucleobase agaa in flanking site	0.29708262
125	Frequency of Nucleobase atta in flanking site	0.295641929
105	Frequency of Nucleobase agga in flanking site	0.295217411
165	Frequency of Nucleobase aagg in flanking site	0.29475192
54	Frequency of Nucleobase gat in flanking site	0.292797514
50	Frequency of Nucleobase gcg in flanking site	0.289881846
133	Frequency of Nucleobase aaca in flanking site	0.289695801
237	Frequency of Nucleobase atgt in flanking site	0.28809256
94	Frequency of Nucleobase gtaa in flanking site	0.286449884
222	Frequency of Nucleobase gtat in flanking site	0.286043764
102	Frequency of Nucleobase gaga in flanking site	0.284769359
215	Frequency of Nucleobase taat in flanking site	0.28266858
229	Frequency of Nucleobase aagt in flanking site	0.280797772
177	Frequency of Nucleobase acgg in flanking site	0.279489317
211	Frequency of Nucleobase tccg in flanking site	0.279142631
145	Frequency of Nucleobase acca in flanking site	0.278566066
243	Frequency of Nucleobase tcgt in flanking site	0.278398018
308	Frequency of Nucleobase ccgc in flanking site	0.278106917
244	Frequency of Nucleobase ccgt in flanking site	0.275966417
281	Frequency of Nucleobase agac in flanking site	0.273145014
260	Frequency of Nucleobase cctt in flanking site	0.270413655
335	Frequency of Nucleobase ttcc in flanking site	0.268897807
257	Frequency of Nucleobase actt in flanking site	0.268848344
185	Frequency of Nucleobase agtg in flanking site	0.268538226
120	Frequency of Nucleobase cata in flanking site	0.26832259
178	Frequency of Nucleobase gcgg in flanking site	0.267872504
77	Frequency of Nucleobase atc in flanking site	0.267309056
136	Frequency of Nucleobase caca in flanking site	0.266395962
153	Frequency of Nucleobase agag in flanking site	0.265538086
186	Frequency of Nucleobase ggtg in flanking site	0.264493548

Table 1 – *Continued from previous page*

Feature Index	Feature Description	1− Laplacian Score
250	Frequency of Nucleobase ggtt in flanking site	0.263232344
130	Frequency of Nucleobase gcta in flanking site	0.262963113
139	Frequency of Nucleobase tgca in flanking site	0.262625623
270	Frequency of Nucleobase gtct in flanking site	0.262084469
233	Frequency of Nucleobase aggt in flanking site	0.261979586
115	Frequency of Nucleobase tcga in flanking site	0.261688329
73	Frequency of Nucleobase agc in flanking site	0.260749469
217	Frequency of Nucleobase agat in flanking site	0.260497146
208	Frequency of Nucleobase ctcg in flanking site	0.259778791
294	Frequency of Nucleobase gagc in flanking site	0.255048968
179	Frequency of Nucleobase tcgg in flanking site	0.25350705
272	Frequency of Nucleobase ctct in flanking site	0.253354408
143	Frequency of Nucleobase ttca in flanking site	0.252869258
171	Frequency of Nucleobase tggg in flanking site	0.252647818
168	Frequency of Nucleobase cagg in flanking site	0.252390097
230	Frequency of Nucleobase gagt in flanking site	0.249925912
204	Frequency of Nucleobase cgcg in flanking site	0.248858324
287	Frequency of Nucleobase ttac in flanking site	0.248314321
214	Frequency of Nucleobase gaat in flanking site	0.246990895
91	Frequency of Nucleobase tgaa in flanking site	0.246493725
289	Frequency of Nucleobase acac in flanking site	0.245559281
365	number of Interior Loop with length 2	0.243545307
32	Frequency of Nucleobase cta in flanking site	0.243427258
156	Frequency of Nucleobase cgag in flanking site	0.243263241
48	Frequency of Nucleobase ctg in flanking site	0.242688798
275	Frequency of Nucleobase tcct in flanking site	0.240779579
190	Frequency of Nucleobase gttg in flanking site	0.240752448
248	Frequency of Nucleobase catt in flanking site	0.240111792
263	Frequency of Nucleobase tact in flanking site	0.239239315
90	Frequency of Nucleobase ggaa in flanking site	0.238820203
66	Frequency of Nucleobase gct in flanking site	0.238514358
134	Frequency of Nucleobase gaca in flanking site	0.238009162
313	Frequency of Nucleobase agtc in flanking site	0.237805683
164	Frequency of Nucleobase ccag in flanking site	0.236691596
227	Frequency of Nucleobase tcat in flanking site	0.235943638
99	Frequency of Nucleobase tcaa in flanking site	0.234806139
207	Frequency of Nucleobase ttcg in flanking site	0.234585917
315	Frequency of Nucleobase tgtc in flanking site	0.234142928

Table 1 – *Continued from previous page*

Feature Index	Feature Description	1− Laplacian Score
98	Frequency of Nucleobase gcaa in flanking site	0.232963229
159	Frequency of Nucleobase ttag in flanking site	0.232783623
285	Frequency of Nucleobase atac in flanking site	0.23187766
261	Frequency of Nucleobase aact in flanking site	0.228202553
296	Frequency of Nucleobase cagc in flanking site	0.227965063
121	Frequency of Nucleobase agta in flanking site	0.22771681
131	Frequency of Nucleobase tcta in flanking site	0.226558295
157	Frequency of Nucleobase atag in flanking site	0.225610787
231	Frequency of Nucleobase tagt in flanking site	0.225519574
188	Frequency of Nucleobase cgtg in flanking site	0.225470175
100	Frequency of Nucleobase ccaa in flanking site	0.224175011
75	Frequency of Nucleobase tgc in flanking site	0.223928851
141	Frequency of Nucleobase atca in flanking site	0.221842483
142	Frequency of Nucleobase gtca in flanking site	0.219280065
135	Frequency of Nucleobase taca in flanking site	0.215569533
321	Frequency of Nucleobase actc in flanking site	0.215070486
155	Frequency of Nucleobase tgag in flanking site	0.214425186
228	Frequency of Nucleobase ccat in flanking site	0.213407483
151	Frequency of Nucleobase taag in flanking site	0.210421006
161	Frequency of Nucleobase acag in flanking site	0.209305256
303	Frequency of Nucleobase ttgc in flanking site	0.208958884
147	Frequency of Nucleobase tcca in flanking site	0.208503371
262	Frequency of Nucleobase gact in flanking site	0.208023769
109	Frequency of Nucleobase atga in flanking site	0.207868394
92	Frequency of Nucleobase cgaa in flanking site	0.207752445
235	Frequency of Nucleobase tggt in flanking site	0.207659405
126	Frequency of Nucleobase gtta in flanking site	0.206877773
246	Frequency of Nucleobase gatt in flanking site	0.20597805
300	Frequency of Nucleobase cggc in flanking site	0.205920581
252	Frequency of Nucleobase cgtt in flanking site	0.205707207
181	Frequency of Nucleobase aatg in flanking site	0.204993579
225	Frequency of Nucleobase acat in flanking site	0.204646703
103	Frequency of Nucleobase taga in flanking site	0.20380668
196	Frequency of Nucleobase cctg in flanking site	0.203627747
152	Frequency of Nucleobase caag in flanking site	0.203449802
107	Frequency of Nucleobase tgga in flanking site	0.203139141
163	Frequency of Nucleobase tcag in flanking site	0.203135409
317	Frequency of Nucleobase attc in flanking site	0.202917301

Table 1 – *Continued from previous page*

Feature Index	Feature Description	1− **Laplacian Score**
110	Frequency of Nucleobase gtga in flanking site	0.202093833
197	Frequency of Nucleobase aacg in flanking site	0.200472412
111	Frequency of Nucleobase ttga in flanking site	0.200045171
249	Frequency of Nucleobase agtt in flanking site	0.1993359
96	Frequency of Nucleobase ctaa in flanking site	0.196426912
175	Frequency of Nucleobase ttgg in flanking site	0.195347118
329	Frequency of Nucleobase agcc in flanking site	0.194204392
122	Frequency of Nucleobase ggta in flanking site	0.19409207
200	Frequency of Nucleobase cacg in flanking site	0.193887689
219	Frequency of Nucleobase tgat in flanking site	0.192564257
232	Frequency of Nucleobase cagt in flanking site	0.192201047
327	Frequency of Nucleobase tacc in flanking site	0.191839866
258	Frequency of Nucleobase gctt in flanking site	0.19121014
318	Frequency of Nucleobase gttc in flanking site	0.190344152
344	number of bulge with length 1	0.18854486
291	Frequency of Nucleobase tcac in flanking site	0.187274897
220	Frequency of Nucleobase cgat in flanking site	0.184025254
307	Frequency of Nucleobase tcgc in flanking site	0.183986871
241	Frequency of Nucleobase acgt in flanking site	0.182904126
158	Frequency of Nucleobase gtag in flanking site	0.18270943
312	Frequency of Nucleobase catc in flanking site	0.182423671
279	Frequency of Nucleobase taac in flanking site	0.182254663
236	Frequency of Nucleobase cggt in flanking site	0.181003598
240	Frequency of Nucleobase ctgt in flanking site	0.179609896
278	Frequency of Nucleobase gaac in flanking site	0.179243982
160	Frequency of Nucleobase ctag in flanking site	0.178770331
118	Frequency of Nucleobase gata in flanking site	0.178585827
325	Frequency of Nucleobase aacc in flanking site	0.178294791
264	Frequency of Nucleobase cact in flanking site	0.176386702
104	Frequency of Nucleobase caga in flanking site	0.174938885
34	Frequency of Nucleobase gca in flanking site	0.174840383
189	Frequency of Nucleobase attg in flanking site	0.17318187
309	Frequency of Nucleobase aatc in flanking site	0.17233501
112	Frequency of Nucleobase ctga in flanking site	0.17200947
333	Frequency of Nucleobase atcc in flanking site	0.171478909
192	Frequency of Nucleobase cttg in flanking site	0.170607584
269	Frequency of Nucleobase atct in flanking site	0.170402576
176	Frequency of Nucleobase ctgg in flanking site	0.170365203

Table 1 – *Continued from previous page*

Feature Index	Feature Description	1− Laplacian Score
280	Frequency of Nucleobase caac in flanking site	0.169628907
354	number of Hairpin with length 1	0.168126239
224	Frequency of Nucleobase ctat in flanking site	0.166755836
173	Frequency of Nucleobase atgg in flanking site	0.166318577
199	Frequency of Nucleobase tacg in flanking site	0.166145772
273	Frequency of Nucleobase acct in flanking site	0.165596062
202	Frequency of Nucleobase ggcg in flanking site	0.16481325
268	Frequency of Nucleobase cgct in flanking site	0.163347004
297	Frequency of Nucleobase aggc in flanking site	0.162162309
301	Frequency of Nucleobase atgc in flanking site	0.160959048
283	Frequency of Nucleobase tgac in flanking site	0.16083641
322	Frequency of Nucleobase gctc in flanking site	0.160206524
193	Frequency of Nucleobase actg in flanking site	0.159636645
124	Frequency of Nucleobase cgta in flanking site	0.158191639
195	Frequency of Nucleobase tctg in flanking site	0.157843538
167	Frequency of Nucleobase tagg in flanking site	0.157377616
184	Frequency of Nucleobase catg in flanking site	0.155519503
274	Frequency of Nucleobase gcct in flanking site	0.154853013
205	Frequency of Nucleobase atcg in flanking site	0.152687027
182	Frequency of Nucleobase gatg in flanking site	0.152261649
203	Frequency of Nucleobase tgcg in flanking site	0.152257951
267	Frequency of Nucleobase tgct in flanking site	0.151714488
286	Frequency of Nucleobase gtac in flanking site	0.151567604
331	Frequency of Nucleobase tgcc in flanking site	0.148856657
306	Frequency of Nucleobase gcgc in flanking site	0.147652792
226	Frequency of Nucleobase gcat in flanking site	0.146820638
288	Frequency of Nucleobase ctac in flanking site	0.146793845
242	Frequency of Nucleobase gcgt in flanking site	0.14612174
357	number of Hairpin with length 4	0.142580593
129	Frequency of Nucleobase acta in flanking site	0.142135549
290	Frequency of Nucleobase gcac in flanking site	0.141630904
356	number of Hairpin with length 3	0.138832052
128	Frequency of Nucleobase ctta in flanking site	0.135602996
216	Frequency of Nucleobase caat in flanking site	0.134179706
140	Frequency of Nucleobase cgca in flanking site	0.133749974
310	Frequency of Nucleobase gatc in flanking site	0.132684341
132	Frequency of Nucleobase ccta in flanking site	0.132482176
304	Frequency of Nucleobase ctgc in flanking site	0.132389078

Table 1 – *Continued from previous page*

Feature Index	Feature Description	1− Laplacian Score
360	number of Hairpin with length 7	0.131450831
146	Frequency of Nucleobase gcca in flanking site	0.128872297
302	Frequency of Nucleobase gtgc in flanking site	0.128592397
311	Frequency of Nucleobase tatc in flanking site	0.127715436
293	Frequency of Nucleobase aagc in flanking site	0.126253173
359	number of Hairpin with length 6	0.123689359
369	number of Interior Loop with length 6	0.122951423
367	number of Interior Loop with length 4	0.117976655
305	Frequency of Nucleobase acgc in flanking site	0.11339999
144	Frequency of Nucleobase ctca in flanking site	0.112834133
347	number of bulge with length 4	0.111954215
218	Frequency of Nucleobase ggat in flanking site	0.110246376
348	number of bulge with length 5	0.106258054
358	number of Hairpin with length 5	0.105702489
266	Frequency of Nucleobase ggct in flanking site	0.104079532
137	Frequency of Nucleobase agca in flanking site	0.10096554
201	Frequency of Nucleobase agcg in flanking site	0.096112259
346	number of bulge with length 3	0.092624579
366	number of Interior Loop with length 3	0.091690309
370	number of Interior Loop with length 7	0.086482095
345	number of bulge with length 2	0.083600697
194	Frequency of Nucleobase gctg in flanking site	0.075261849
368	number of Interior Loop with length 5	0.074094624
355	number of Hairpin with length 2	0.061156788
350	number of bulge with length 7	0.057720939
114	Frequency of Nucleobase gcga in flanking site	0.055005122
349	number of bulge with length 6	0.052970667
138	Frequency of Nucleobase ggca in flanking site	0.040897574
162	Frequency of Nucleobase gcag in flanking site	0.033499254
299	Frequency of Nucleobase tggc in flanking site	0.026614581
364	number of Interior Loop with length 1	0