TableS1 The top 1000 features in the optimal feature set o

Order	Feature	F-score	kmer
1	8mer	0.10931	TCCCAAAG
2	8mer	0.107445	GCGGGGCG
3	8mer	0.106471	AACGGCCC
4	8mer	0.105616	CGGTGACG
5	8mer	0.098194	CGGGTCAC
6	8mer	0.095685	TCACGGGA
7	8mer	0.09567	TGCGGTTC
8	8mer	0.09254	GGGGGACC
9	8mer	0.091784	CCACGACC
10	8mer	0.091036	CGCCCCGG
11	8mer	0.09051	TCAACGTC
12	8mer	0.088855	CGAAGTCG
13	8mer	0.088667	CGGGGGAC
14	8mer	0.08771	ATGGGGCG
15	8mer	0.087693	TCGGCCCC
16	8mer	0.087595	AGCGGAGC
17	8mer	0.087322	AATCGACG
18	8mer	0.085901	GCTCACGG
19	8mer	0.084716	GTAGGGCG
20	8mer	0.084641	CGCAGTTC
21	8mer	0.083424	CCGCAGTT
22	8mer	0.082921	GCGGAGCC
23	8mer	0.081369	CATCGGCT
24	8mer	0.079854	AGTAGGAC
25	8mer	0.079681	CCCAAAGT
26	8mer	0.079471	AAGGGCTG
27	8mer	0.078261	GGGATGCG
28	8mer	0.077741	GATCCCGA
29	8mer	0.077287	GCGTGAGC
30	8mer	0.077177	CGGGCGGG
31	8mer	0.077057	GCGGGGGA
32	8mer	0.076646	AGACGGCG
33	8mer	0.075641	CTTGTATC
34	8mer	0.07546	CTTAAACC
35	8mer	0.075416	GAGCCCGA
36	8mer	0.07508	CGTAGCTT
37	8mer	0.074414	CCCCCGAA
38	8mer	0.073205	GGCGAGGT
39	8mer	0.073205	CCGCACCG
40	8mer	0.072737	CGCAGCGT
41	8mer	0.072567	CATCGCGG
42	8mer	0.072132	TGGGCCCG
43	8mer	0.072036	CCCGTTTC
44	8mer	0.072034	GTGCGTGA
45	8mer	0.071804	GGATGCGG
46	8mer	0.070518	AAGGGGCG
47	8mer	0.07038	CCGAAGGC
48	8mer	0.069523	CCTGTGTT
49	8mer	0.069495	GTGGAAGA
50	8mer	0.068895	ATCCCGAA
51	8mer	0.068761	CTTGAACC
52	8mer	0.068296	ACGAGCCC
53	8mer	0.068213	AGACCCTC
54	8mer	0.068069	ACGTGCGG
55	8mer	0.068033	GTGCGACT

Order	Feature	F-score
351	8mer	0.045957
352	8mer	0.045842
353	8mer	0.045821
354	8mer	0.045807
355	8mer	0.04574
356	8mer	0.045598
357	8mer	0.045566
358	8mer	0.045552
359	8mer	0.045544
360	8mer	0.045541
361	Triplet	0.045475
362	8mer	0.045446
363	8mer	0.045446
364	8mer	0.045399
365	8mer	0.045291
366	8mer	0.045257
367	8mer	0.045235
368	8mer	0.045227
369	8mer	0.045213
370	8mer	0.045208
371	8mer	0.045197
372	8mer	0.045184
373	8mer	0.045154
374	8mer	0.04515
375	8mer	0.045117
376	8mer	0.045071
377	8mer	0.045064
378	8mer	0.045057
379	8mer	0.044993
380	8mer	0.044947
381	8mer	0.044928
382	8mer	0.044916
383	8mer	0.044909
384	8mer	0.044842
385	8mer	0.044812
386	8mer	0.044798
387	8mer	0.044713
388	SC-PseDNC	0.04466
389	8mer	0.044572
390	8mer	0.044498
391	8mer	0.044442
392	8mer	0.044404
393	8mer	0.044402
394	8mer	0.044323
395	8mer	0.044257
396	8mer	0.044228
397	8mer	0.044215
398	8mer	0.044191
399	8mer	0.044183
400	8mer	0.044145
401	8mer	0.044099
402	8mer	0.044081
403	8mer	0.044067
404	8mer	0.04403
405	8mer	0.043915

56	8mer	0.067987	GACCATCG
57	8mer	0.067781	ACGAATCA
58	8mer	0.067781	TCGTCGAC
59	8mer	0.067781	TAACCGCC
60	8mer	0.067781	GTCGACCG
61	8mer	0.067781	CGAGTCGG
62	8mer	0.067781	TATGCGGG
63	8mer	0.067781	ATGCGGGG
64	8mer	0.067692	AGCCACCG
65	8mer	0.067483	GCGTGCTC
66	8mer	0.067415	TCCGGGCG
67	8mer	0.067239	GGAATCCG
68	8mer	0.067199	TGACGAGC
69	8mer	0.067076	CGCACAGT
70	8mer	0.066653	GCGCGTCC
71	8mer	0.066595	TCCGCCGG
72	8mer	0.066489	GGCCCGTG
73	8mer	0.066467	GCGGATCT
74	8mer	0.06636	GGGCCCCT
75	8mer	0.06628	CCTGGGCC
76	8mer	0.066234	GACGAATC
77	8mer	0.066039	AACGTCGG
78	8mer	0.06574	TTATGCGG
79	8mer	0.06574	CGACGAAT
80	8mer	0.065703	GCGGTAAC
81	8mer	0.065606	CCTGGGCG
82	8mer	0.065523	TTGTATCC
83	8mer	0.065484	TGGTCTCG
84	8mer	0.065447	CCACTGCA
85	8mer	0.065364	CGACAGGA
86	8mer	0.065247	CATGACAC
87	8mer	0.065246	ATCCGCCG
88	8mer	0.064857	TGTCGGTA
89	8mer	0.064394	CTCATTGC
90	8mer	0.06426	GTAACCGC
91	8mer	0.064055	CCGGTTGA
92	8mer	0.063797	TGTGCGCC
93	8mer	0.063633	TGCGCCGT
94	8mer	0.063586	GCACCGAA
95	8mer	0.063417	AGGAGGGC
96	8mer	0.063183	AACCGCGG
97	8mer	0.063161	CGGTACTT
98	8mer	0.063124	CCCCAGCG
99	8mer	0.062931	ATACCGCC
100	8mer	0.062913	TCCAGGCG
101	8mer	0.062893	TCCCGCAC
102	8mer	0.062656	ACCGGACT
103	8mer	0.06263	AGGTCGAG
104	8mer	0.062472	GAAACCCG
105	8mer	0.062448	TTGCCGGT
106	8mer	0.06233	TCAAGCAC
107	8mer	0.062312	CCAAGATC
108	8mer	0.062211	ATTGCCGG
109	8mer	0.061952	TCACGGTG
110	8mer	0.06195	GATGCGGA
111	8mer	0.061771	GGCCCCTG
112	8mer	0.061704	CGGGGCGA
113	8mer	0.061609	TCGTCCAA

406	8mer	0.043904
407	8mer	0.043888
408	8mer	0.043869
409	8mer	0.043817
410	8mer	0.043812
411	8mer	0.043806
412	8mer	0.043794
413	8mer	0.043789
414	8mer	0.043774
415	8mer	0.043767
416	8mer	0.04375
417	8mer	0.043725
418	8mer	0.043715
419	8mer	0.043713
420	8mer	0.043673
421	8mer	0.043639
422	8mer	0.043625
423	8mer	0.043623
424	8mer	0.043575
425	8mer	0.043522
426	8mer	0.043511
427	8mer	0.043503
428	8mer	0.04347
429	8mer	0.043468
430	8mer	0.043459
431	8mer	0.043415
432	8mer	0.043357
433	8mer	0.043339
434	8mer	0.04333
435	8mer	0.04332
436	8mer	0.043304
437	8mer	0.043299
438	8mer	0.043284
439	8mer	0.043258
440	8mer	0.043251
441	8mer	0.04325
442	8mer	0.043201
443	8mer	0.043186
444	8mer	0.043135
445	8mer	0.043132
446	8mer	0.043127
447	8mer	0.0431
448	8mer	0.043092
449	8mer	0.043052
450	8mer	0.043045
451	8mer	0.043035
452	8mer	0.042965
453	8mer	0.042959
454	8mer	0.042959
455	8mer	0.04295
456	8mer	0.042895
457	8mer	0.042872
458	8mer	0.042863
459	8mer	0.042848
460	8mer	0.04278
461	8mer	0.042751
462	8mer	0.042736
463	8mer	0.042727

114	8mer	0.061468	CGGGGCCT
115	8mer	0.061418	CGATGATA
116	8mer	0.06132	GGGCGAGG
117	8mer	0.061146	TCGGTACT
118	8mer	0.061133	CGGTGTCG
119	8mer	0.061067	GGGGTCGA
120	8mer	0.060888	TCAAGCGA
121	8mer	0.060681	GGTCGTGG
122	8mer	0.06066	GGAGGGGT
123	8mer	0.060619	CCTTCCGC
124	8mer	0.060596	ACTCCCCG
125	8mer	0.060568	CAAGCGAT
126	8mer	0.060416	AGGGGATA
127	8mer	0.060304	CCTCGGCC
128	8mer	0.060137	GCATTCGG
129	8mer	0.060047	GGGCGCCC
130	8mer	0.059996	TTGGGCGC
131	8mer	0.059859	ATGCGGAT
132	8mer	0.059776	GGGGCCTT
133	8mer	0.059758	GCTTCGCC
134	8mer	0.059737	GTGTTACT
135	8mer	0.059509	CCAACATG
136	8mer	0.059376	GACCGCAA
137	8mer	0.059293	TGCGGATC
138	8mer	0.059292	CACGACCC
139	8mer	0.059268	CAACATGG
140	8mer	0.059228	GGGCGCGT
141	8mer	0.0589	CGTAGGGC
142	8mer	0.058785	CATCGGTA
143	8mer	0.058759	GGATCAGC
144	8mer	0.058752	AAACGTCG
145	8mer	0.058738	AGATCCCG
146	8mer	0.058588	TCCGTAGC
147	8mer	0.05853	GGCGCAGC
148	8mer	0.058336	ACGCGGAG
149	8mer	0.058097	TAGGACAA
150	8mer	0.057951	GCCTTGGG
151	8mer	0.057646	AATCCCCG
152	8mer	0.057561	CGCCGTTA
153	8mer	0.057561	GTAAACCG
154	8mer	0.057514	GGCCTATG
155	8mer	0.057493	CGGACGTG
156	8mer	0.057406	GCCGGAAG
157	8mer	0.057352	TTCCCTAA
158	8mer	0.057325	CGCGGGAA
159	8mer	0.057279	GGAATATC
160	8mer	0.057262	GGCGCGTC
161	8mer	0.057246	GGTAACCG
162	8mer	0.057244	AATCCGCC
163	8mer	0.057216	ACATCGCG
164	8mer	0.057143	GAGGGCAG
165	8mer	0.057128	TGATCCGG
166	8mer	0.057127	CTGATCCG
167	8mer	0.057105	GCCGCAGT
168	8mer	0.057074	GACGAGCC
169	8mer	0.05707	GTCGGTAC
170	8mer	0.05707	TCGTAGGG
171	8mer	0.056971	CGGGCATC

464	8mer	0.042692
465	8mer	0.042663
466	8mer	0.042634
467	8mer	0.042609
468	8mer	0.042582
469	8mer	0.042576
470	8mer	0.042567
471	8mer	0.042554
472	8mer	0.042547
473	8mer	0.042476
474	8mer	0.042462
475	Triplet	0.042416
476	8mer	0.04237
477	8mer	0.04234
478	8mer	0.042305
479	8mer	0.042265
480	SC-PseDNC	0.042253
481	8mer	0.042235
482	8mer	0.042223
483	8mer	0.042168
484	8mer	0.042149
485	8mer	0.042135
486	8mer	0.042134
487	8mer	0.042068
488	8mer	0.042059
489	8mer	0.042056
490	8mer	0.042056
491	8mer	0.04204
492	8mer	0.042039
493	8mer	0.041942
494	8mer	0.041938
495	8mer	0.041901
496	8mer	0.041879
497	8mer	0.04183
498	8mer	0.04182
499	8mer	0.041809
500	8mer	0.041808
501	8mer	0.041726
502	8mer	0.041704
503	8mer	0.041699
504	Triplet	0.04159
505	8mer	0.041576
506	8mer	0.041533
507	8mer	0.041448
508	8mer	0.041447
509	8mer	0.041437
510	8mer	0.041436
511	8mer	0.041417
512	8mer	0.041416
513	8mer	0.041376
514	8mer	0.041342
515	8mer	0.041297
516	8mer	0.041287
517	8mer	0.041237
518	8mer	0.041175
519	8mer	0.041136
520	8mer	0.04113
521	8mer	0.041106

172	8mer	0.056836	CGCCGTGT
173	8mer	0.056612	TGCGGTAA
174	8mer	0.056504	GGAGCCGG
175	8mer	0.056504	ATAGCTAA
176	8mer	0.056455	CGGGATTC
177	8mer	0.056438	GAGACGCC
178	8mer	0.056382	CGTGCTCG
179	8mer	0.056377	CGGGCCCC
180	8mer	0.056307	CGGAGCCC
181	8mer	0.056099	AGCTACGC
182	8mer	0.056053	ATCGCGGG
183	8mer	0.055956	TCGCGCCG
184	8mer	0.055933	GATTCAAC
185	8mer	0.055575	GGAATCGT
186	8mer	0.055533	AAAAATTA
187	8mer	0.055492	CGGATCTA
188	8mer	0.05548	AGACCCGC
189	8mer	0.055231	ACCTGTAC
190	8mer	0.0552	GCCCTGGT
191	8mer	0.055116	GTCCAATG
192	8mer	0.055039	CGGCGATG
193	8mer	0.054965	TGAGTGTC
194	8mer	0.054878	TTCAAGCG
195	8mer	0.0548	GAGAATCG
196	8mer	0.054507	CAAAACTC
197	8mer	0.054459	AATATACC
198	8mer	0.05445	TATACCAG
199	8mer	0.054389	ACCCGCAG
200	8mer	0.054301	CGGTAACC
201	8mer	0.054267	TAGAGGCG
202	8mer	0.054134	CGGACTTT
203	8mer	0.053955	GCCCCGTG
204	8mer	0.053902	AAAATTAG
205	8mer	0.053867	CGAGCCCG
206	8mer	0.053538	GACCCTCC
207	8mer	0.053492	ATGCAACG
208	8mer	0.053471	TCCAATGC
209	8mer	0.05329	CCTCAGCC
210	8mer	0.053201	CCGCATTC
211	8mer	0.052943	ACGGAGGG
212	8mer	0.052902	GCTGCTCG
213	8mer	0.052845	GCGATGGG
214	8mer	0.052795	CTCGCGTC
215	8mer	0.052789	AATCCCAA
216	8mer	0.052597	GAATATTG
217	8mer	0.052512	GACCTAGC
218	8mer	0.052453	TGGCCGTC
219	8mer	0.052419	CCAGGTCC
220	8mer	0.052418	GGGCTGAC
221	8mer	0.052379	TTTGCGGA
222	8mer	0.052338	CCGCCCCC
223	8mer	0.052314	CTGACGAG
224	8mer	0.052189	GGGGCCTA
225	8mer	0.052127	GCAGAGGG
226	8mer	0.052092	CGTCGGCG
227	8mer	0.051893	CCCGAAGT
228	8mer	0.051883	GCGGTGTC
229	8mer	0.051652	CGTGAGCC

522	8mer	0.041081
523	8mer	0.041078
524	8mer	0.041011
525	8mer	0.040951
526	8mer	0.040833
527	8mer	0.040822
528	8mer	0.040799
529	8mer	0.040795
530	8mer	0.040795
531	8mer	0.040795
532	8mer	0.040795
533	8mer	0.040795
534	8mer	0.040795
535	8mer	0.040786
536	8mer	0.040726
537	8mer	0.040705
538	8mer	0.040688
539	8mer	0.040643
540	8mer	0.04057
541	8mer	0.040562
542	8mer	0.040556
543	8mer	0.040475
544	8mer	0.040464
545	8mer	0.040462
546	8mer	0.040398
547	8mer	0.040398
548	8mer	0.040395
549	8mer	0.040386
550	8mer	0.040378
551	8mer	0.040371
552	8mer	0.040365
553	8mer	0.040326
554	8mer	0.040326
555	8mer	0.040314
556	8mer	0.040306
557	8mer	0.040293
558	8mer	0.040258
559	8mer	0.040252
560	8mer	0.040247
561	8mer	0.04024
562	8mer	0.040219
563	8mer	0.040205
564	8mer	0.040192
565	8mer	0.040188
566	8mer	0.040173
567	8mer	0.040165
568	8mer	0.040044
569	8mer	0.040027
570	8mer	0.039989
571	8mer	0.039986
572	8mer	0.039906
573	8mer	0.039904
574	8mer	0.039873
575	8mer	0.039868
576	8mer	0.039852
577	8mer	0.039832
578	8mer	0.039812
579	8mer	0.039805

230	8mer	0.05164	GCCGCATT
231	8mer	0.051615	GGGTCGAG
232	8mer	0.051532	TCGAGTCG
233	8mer	0.051301	TCCCAGCT
234	8mer	0.051286	GAGCGCCA
235	8mer	0.051239	ATTGCCTG
236	8mer	0.051176	CAAGATCG
237	8mer	0.051123	GCGTTCAG
238	8mer	0.051068	GCTCCTCG
239	8mer	0.050908	ACGGGCAG
240	8mer	0.050867	AGATCGCG
241	8mer	0.050733	GATATTGA
242	8mer	0.050698	AGCCCGAA
243	8mer	0.050688	TTTTTAGT
244	8mer	0.050647	CTGGGCCC
245	8mer	0.050597	CCGCCAGC
246	8mer	0.050593	CTCCGTAG
247	8mer	0.050576	CGGAGGTC
248	8mer	0.050492	CGTCCAAT
249	8mer	0.050489	TGAATGCA
250	8mer	0.050383	GCCTGGCC
251	8mer	0.050332	GGGCGGGC
252	8mer	0.050237	TGCCGGTC
253	8mer	0.0502	TAATACCG
254	8mer	0.050185	ATTACTGC
255	8mer	0.050111	CCTGAGCA
256	8mer	0.050078	GACTGTAT
257	8mer	0.050071	GGCGTGCT
258	8mer	0.049934	GGAAACGG
259	8mer	0.049918	GCGGTTCC
260	8mer	0.049887	CCTCCGTA
261	8mer	0.049846	GATCGCGC
262	8mer	0.049818	TTGGTGCT
263	8mer	0.049787	CGGAGGCG
264	8mer	0.049757	CGAGATCG
265	8mer	0.049721	TGAACCCG
266	8mer	0.049702	AATAGCCA
267	8mer	0.049633	AGCGGGGC
268	8mer	0.049554	GGTGAACC
269	8mer	0.049369	ATAAATGG
270	8mer	0.049266	CACTGCAC
271	8mer	0.049199	TGCCTGCG
272	8mer	0.049125	CAAGGAGT
273	8mer	0.049102	CATTGCCG
274	8mer	0.048983	GCCAGAGT
275	8mer	0.0489	GGCGCCCA
276	8mer	0.048871	AGCCGCAT
277	8mer	0.048848	CTCACGGG
278	8mer	0.048843	CTTCCCTA
279	8mer	0.048756	GATCGTCG
280	8mer	0.048723	GGCTATTG
281	8mer	0.04869	GGGACTGA
282	8mer	0.048651	TGAACACG
283	8mer	0.048619	GGGCGGGG
284	8mer	0.048591	ACTGTGGG
285	8mer	0.048586	GACCCGCA
286	8mer	0.048551	AGACCTGG
287	8mer	0.048515	GACTCGCG

580	8mer	0.039756
581	8mer	0.039632
582	8mer	0.039601
583	8mer	0.039546
584	8mer	0.039545
585	8mer	0.039509
586	8mer	0.039489
587	8mer	0.039408
588	8mer	0.039396
589	8mer	0.039384
590	8mer	0.039362
591	8mer	0.039358
592	8mer	0.039265
593	8mer	0.039222
594	8mer	0.039208
595	8mer	0.039192
596	8mer	0.039183
597	8mer	0.039178
598	8mer	0.039146
599	8mer	0.039133
600	8mer	0.039092
601	8mer	0.03906
602	8mer	0.039038
603	8mer	0.039002
604	8mer	0.039
605	8mer	0.038999
606	8mer	0.038984
607	8mer	0.038973
608	8mer	0.038944
609	8mer	0.038939
610	8mer	0.03893
611	8mer	0.03885
612	8mer	0.038789
613	8mer	0.038745
614	8mer	0.038735
615	8mer	0.03869
616	8mer	0.038687
617	8mer	0.038675
618	8mer	0.038664
619	8mer	0.03865
620	8mer	0.038649
621	8mer	0.038618
622	8mer	0.038616
623	8mer	0.038607
624	8mer	0.0386
625	8mer	0.038583
626	8mer	0.038573
627	8mer	0.038571
628	8mer	0.038566
629	8mer	0.038562
630	8mer	0.038544
631	8mer	0.038543
632	8mer	0.038531
633	8mer	0.038508
634	8mer	0.038481
635	8mer	0.038451
636	8mer	0.038446
637	8mer	0.038361

288	8mer	0.048481	CGCCGCCG
289	8mer	0.048465	CGGGAGCG
290	8mer	0.048459	CGCCCGGC
291	8mer	0.048409	GGCGCGGC
292	8mer	0.048382	ATCGTCGT
293	8mer	0.048319	GTTGTACG
294	8mer	0.048104	GAACGGCC
295	8mer	0.048021	GTGGGAGG
296	8mer	0.048009	AGGTCCCT
297	8mer	0.047939	CGGCCATC
298	8mer	0.047913	CGAGACTG
299	8mer	0.047904	CCCGAAGG
300	8mer	0.047877	GTTCTGCA
301	8mer	0.047858	CCTGTCGG
302	8mer	0.047854	CGGCGGGG
303	8mer	0.047816	ATCGTAGA
304	8mer	0.047784	CTCGTCAA
305	Triplet	0.047765	Triplet
306	8mer	0.047735	AGTCGGAG
307	8mer	0.047729	CCTCTCAT
308	8mer	0.047708	TGCCCCGA
309	8mer	0.047676	ACTTCGAG
310	8mer	0.04766	TGCGGGGG
311	8mer	0.047656	GCCCGAAC
312	8mer	0.047521	TGGGCGCC
313	8mer	0.047505	ACATGGCA
314	Triplet	0.047503	Triplet
315	8mer	0.047497	TGCACGCA
316	8mer	0.047407	CGCATTCC
317	8mer	0.047393	CACCGGAC
318	8mer	0.047279	CGCTTGAA
319	8mer	0.047215	GAGATCCC
320	8mer	0.047169	GGTCGAGA
321	8mer	0.047111	AAGTCGGG
322	8mer	0.04707	CGACCCGG
323	8mer	0.047052	GCGCCGTG
324	8mer	0.046985	TGCGTCAA
325	8mer	0.04697	TGGACAGC
326	8mer	0.046957	TCGTCAAC
327	8mer	0.0469	CACGGGAG
328	8mer	0.046795	ACGACCCG
329	8mer	0.046768	TATCCAGT
330	8mer	0.046766	GGTCATCG
331	8mer	0.046758	ACGCAGGG
332	8mer	0.046711	CGCGCCGT
333	8mer	0.046694	CTCCCAAA
334	8mer	0.046656	CGGGGCGG
335	8mer	0.046639	TTATGTCG
336	8mer	0.046637	GCGCAGCA
337	8mer	0.046537	GACCGCGA
338	8mer	0.046483	AGAAGCGG
339	8mer	0.046434	ATCGCCGA
340	8mer	0.04641	TGTCCGGT
341	8mer	0.046361	AGCAGTCC
342	8mer	0.046282	AGCTCGGC
343	8mer	0.046234	TCTGGTAT
344	8mer	0.046221	CGGGCACT
345	8mer	0.046201	GTTAATTT

638	8mer	0.03836
639	8mer	0.038343
640	8mer	0.038332
641	8mer	0.038297
642	8mer	0.038275
643	8mer	0.03827
644	8mer	0.038252
645	8mer	0.038251
646	8mer	0.038249
647	8mer	0.038244
648	8mer	0.038228
649	8mer	0.038226
650	8mer	0.038193
651	8mer	0.038168
652	8mer	0.038135
653	8mer	0.038124
654	8mer	0.03805
655	8mer	0.038048
656	8mer	0.038015
657	8mer	0.037947
658	8mer	0.037943
659	8mer	0.037942
660	8mer	0.037932
661	8mer	0.037889
662	8mer	0.037866
663	8mer	0.037866
664	8mer	0.037866
665	8mer	0.037866
666	8mer	0.037866
667	8mer	0.037861
668	8mer	0.037833
669	8mer	0.037833
670	8mer	0.037822
671	8mer	0.037764
672	8mer	0.037759
673	8mer	0.037755
674	8mer	0.03775
675	8mer	0.037746
676	8mer	0.037736
677	8mer	0.037682
678	8mer	0.037677
679	8mer	0.037644
680	8mer	0.03761
681	8mer	0.037582
682	8mer	0.037572
683	8mer	0.037572
684	8mer	0.037561
685	8mer	0.037556
686	8mer	0.037538
687	8mer	0.037517
688	8mer	0.037512
689	8mer	0.037508
690	8mer	0.037503
691	8mer	0.037495
692	8mer	0.03748
693	8mer	0.03746
694	8mer	0.037429
695	8mer	0.037404

346	8mer	0.046144	CAGGACGA
347	8mer	0.046108	GGTCGGGC
348	8mer	0.04606	CGGGGCCA
349	8mer	0.046033	CGCCCGAG
350	8mer	0.045988	TCGCCCTC

696	8mer	0.037394
697	8mer	0.037359
698	8mer	0.037355
699	8mer	0.037317
700	8mer	0.037314

TableS2 The top 1000 features in the optimal feature set of

Order	Feature	F-score	kmer
1	5mer	0.118544	GGGCG
2	8mer	0.10931	TCCCAAAG
3	8mer	0.107445	GCGGGGCG
4	8mer	0.106471	AACGGCCC
5	8mer	0.105616	CGGTGACG
6	8mer	0.098194	CGGGTCAC
7	8mer	0.095685	TCACGGGA
8	8mer	0.09567	TGCGGTTC
9	8mer	0.09254	GGGGGACC
10	8mer	0.091784	CCACGACC
11	8mer	0.091036	CGCCCCGG
12	5mer	0.090816	GCCGG
13	8mer	0.09051	TCAACGTC
14	8mer	0.088855	CGAAGTCG
15	8mer	0.088667	CGGGGGAC
16	8mer	0.08771	ATGGGGCG
17	8mer	0.087693	TCGGCCCC
18	8mer	0.087595	AGCGGAGC
19	8mer	0.087322	AATCGACG
20	8mer	0.085901	GCTCACGG
21	8mer	0.084716	GTAGGGCG
22	8mer	0.084641	CGCAGTTC
23	8mer	0.083424	CCGCAGTT
24	8mer	0.082921	GCGGAGCC
25	8mer	0.081369	CATCGGCT
26	8mer	0.079854	AGTAGGAC
27	8mer	0.079681	CCCAAAGT
28	8mer	0.079471	AAGGGCTG
29	5mer	0.078959	CGCCG
30	8mer	0.078261	GGGATGCG
31	8mer	0.077741	GATCCCGA
32	8mer	0.077287	GCGTGAGC
33	8mer	0.077177	CGGGCGGG
34	8mer	0.077057	GCGGGGGA
35	8mer	0.076646	AGACGGCG
36	8mer	0.075641	CTTGTATC
37	8mer	0.07546	CTTAAACC
38	8mer	0.075416	GAGCCCGA
39	8mer	0.07508	CGTAGCTT
40	5mer	0.074635	CCCGC
41	8mer	0.074414	CCCCCGAA
42	5mer	0.073675	CGCCC
43	8mer	0.073205	GGCGAGGT
44	8mer	0.073205	CCGCACCG
45	8mer	0.072737	CGCAGCGT
46	8mer	0.072567	CATCGCGG
47	8mer	0.072132	TGGGCCCG
48	8mer	0.072036	CCCGTTTC

Order	Feature	F-score
351	8mer	0.04707
352	8mer	0.047052
353	8mer	0.046985
354	8mer	0.04697
355	8mer	0.046957
356	8mer	0.0469
357	8mer	0.046795
358	8mer	0.046768
359	8mer	0.046766
360	5mer	0.046764
361	8mer	0.046758
362	8mer	0.046711
363	8mer	0.046694
364	8mer	0.046656
365	8mer	0.046639
366	8mer	0.046637
367	8mer	0.046537
368	8mer	0.046483
369	5mer	0.046467
370	8mer	0.046434
371	8mer	0.04641
372	8mer	0.046361
373	8mer	0.046282
374	8mer	0.046234
375	8mer	0.046221
376	8mer	0.046201
377	8mer	0.046144
378	8mer	0.046108
379	8mer	0.04606
380	8mer	0.046033
381	8mer	0.045988
382	8mer	0.045957
383	8mer	0.045842
384	8mer	0.045821
385	8mer	0.045807
386	8mer	0.04574
387	8mer	0.045598
388	8mer	0.045566
389	8mer	0.045552
390	8mer	0.045544
391	8mer	0.045541
392	Triplet	0.045475
393	8mer	0.045446
394	8mer	0.045446
395	8mer	0.045399
396	5mer	0.045344
397	8mer	0.045291
398	8mer	0.045257

49	8mer	0.072034	GTGCGTGA
50	8mer	0.071804	GGATGCGG
51	5mer	0.071377	CGAGA
52	8mer	0.070518	AAGGGGCG
53	8mer	0.07038	CCGAAGGC
54	8mer	0.069523	CCTGTGTT
55	8mer	0.069495	GTGGAAGA
56	8mer	0.068895	ATCCCGAA
57	8mer	0.068761	CTTGAACC
58	8mer	0.068296	ACGAGCCC
59	8mer	0.068213	AGACCCTC
60	8mer	0.068069	ACGTGCGG
61	8mer	0.068033	GTGCGACT
62	8mer	0.067987	GACCATCG
63	8mer	0.067781	ACGAATCA
64	8mer	0.067781	TCGTCGAC
65	8mer	0.067781	TAACCGCC
66	8mer	0.067781	GTCGACCG
67	8mer	0.067781	CGAGTCGG
68	8mer	0.067781	TATGCGGG
69	8mer	0.067781	ATGCGGGG
70	8mer	0.067692	AGCCACCG
71	8mer	0.067483	GCGTGCTC
72	8mer	0.067415	TCCGGGCG
73	5mer	0.067361	GGAGG
74	8mer	0.067239	GGAATCCG
75	8mer	0.067199	TGACGAGC
76	8mer	0.067076	CGCACAGT
77	8mer	0.066653	GCGCGTCC
78	8mer	0.066595	TCCGCCGG
79	8mer	0.066489	GGCCCGTG
80	8mer	0.066467	GCGGATCT
81	8mer	0.06636	GGGCCCCT
82	8mer	0.06628	CCTGGGCC
83	8mer	0.066234	GACGAATC
84	8mer	0.066039	AACGTCGG
85	5mer	0.065994	CCGCC
86	5mer	0.065986	CGGGG
87	8mer	0.06574	TTATGCGG
88	8mer	0.06574	CGACGAAT
89	8mer	0.065703	GCGGTAAC
90	8mer	0.065606	CCTGGGCG
91	8mer	0.065523	TTGTATCC
92	8mer	0.065484	TGGTCTCG
93	8mer	0.065447	CCACTGCA
94	8mer	0.065364	CGACAGGA
95	8mer	0.065247	CATGACAC
96	8mer	0.065246	ATCCGCCG
97	5mer	0.065211	GCGGG
98	8mer	0.064857	TGTCGGTA
99	8mer	0.064394	CTCATTGC
100	8mer	0.06426	GTAACCGC
101	8mer	0.064055	CCGGTTGA
102	5mer	0.063847	CCCCG
103	8mer	0.063797	TGTGCGCC
104	8mer	0.063633	TGCGCCGT
105	8mer	0.063586	GCACCGAA
106	8mer	0.063417	AGGAGGGC

399	8mer	0.045235
400	8mer	0.045227
401	8mer	0.045213
402	8mer	0.045208
403	8mer	0.045197
404	8mer	0.045184
405	8mer	0.045154
406	8mer	0.04515
407	8mer	0.045117
408	8mer	0.045071
409	8mer	0.045064
410	8mer	0.045057
411	8mer	0.044993
412	8mer	0.044947
413	8mer	0.044928
414	8mer	0.044916
415	8mer	0.044909
416	8mer	0.044842
417	8mer	0.044812
418	8mer	0.044798
419	8mer	0.044713
420	C-PseDNC	0.04466
421	8mer	0.044572
422	8mer	0.044498
423	8mer	0.044442
424	8mer	0.044404
425	8mer	0.044402
426	8mer	0.044323
427	5mer	0.044304
428	8mer	0.044257
429	8mer	0.044228
430	8mer	0.044215
431	5mer	0.044197
432	8mer	0.044191
433	8mer	0.044183
434	8mer	0.044145
435	8mer	0.044099
436	8mer	0.044081
437	8mer	0.044067
438	8mer	0.04403
439	8mer	0.043915
440	8mer	0.043904
441	8mer	0.043888
442	8mer	0.043869
443	8mer	0.043817
444	8mer	0.043812
445	8mer	0.043806
446	8mer	0.043794
447	8mer	0.043789
448	8mer	0.043774
449	8mer	0.043767
450	8mer	0.04375
451	8mer	0.043725
452	8mer	0.043715
453	8mer	0.043713
454	8mer	0.043673
455	8mer	0.043639
456	8mer	0.043625

107	8mer	0.063183	AACCGCGG
108	8mer	0.063161	CGGTACTT
109	8mer	0.063124	CCCCAGCG
110	8mer	0.062931	ATACCGCC
111	8mer	0.062913	TCCAGGCG
112	8mer	0.062893	TCCCGCAC
113	8mer	0.062656	ACCGGACT
114	8mer	0.06263	AGGTCGAG
115	8mer	0.062472	GAAACCCG
116	8mer	0.062448	TTGCCGGT
117	8mer	0.06233	TCAAGCAC
118	8mer	0.062312	CCAAGATC
119	8mer	0.062211	ATTGCCGG
120	5mer	0.062101	CGGGC
121	8mer	0.061952	TCACGGTG
122	8mer	0.06195	GATGCGGA
123	5mer	0.061919	GCGCC
124	8mer	0.061771	GGCCCCTG
125	8mer	0.061704	CGGGGCGA
126	8mer	0.061609	TCGTCCAA
127	8mer	0.061468	CGGGGCCT
128	8mer	0.061418	CGATGATA
129	8mer	0.06132	GGGCGAGG
130	8mer	0.061146	TCGGTACT
131	8mer	0.061133	CGGTGTCG
132	8mer	0.061067	GGGGTCGA
133	5mer	0.061032	GCCCG
134	8mer	0.060888	TCAAGCGA
135	8mer	0.060681	GGTCGTGG
136	8mer	0.06066	GGAGGGGT
137	8mer	0.060619	CCTTCCGC
138	8mer	0.060596	ACTCCCCG
139	8mer	0.060568	CAAGCGAT
140	8mer	0.060416	AGGGGATA
141	8mer	0.060304	CCTCGGCC
142	8mer	0.060137	GCATTCGG
143	8mer	0.060047	GGGCGCCC
144	8mer	0.059996	TTGGGCGC
145	8mer	0.059859	ATGCGGAT
146	8mer	0.059776	GGGGCCTT
147	8mer	0.059758	GCTTCGCC
148	8mer	0.059737	GTGTTACT
149	8mer	0.059509	CCAACATG
150	5mer	0.059434	GGCCG
151	8mer	0.059376	GACCGCAA
152	8mer	0.059293	TGCGGATC
153	8mer	0.059292	CACGACCC
154	5mer	0.059282	GTCGG
155	8mer	0.059268	CAACATGG
156	8mer	0.059228	GGGCGCGT
157	8mer	0.0589	CGTAGGGC
158	8mer	0.058785	CATCGGTA
159	8mer	0.058759	GGATCAGC
160	8mer	0.058752	AAACGTCG
161	8mer	0.058738	AGATCCCG
162	8mer	0.058588	TCCGTAGC
163	8mer	0.05853	GGCGCAGC
164	8mer	0.058336	ACGCGGAG

457	8mer	0.043623
458	8mer	0.043575
459	8mer	0.043522
460	8mer	0.043511
461	8mer	0.043503
462	8mer	0.04347
463	8mer	0.043468
464	8mer	0.043459
465	8mer	0.043415
466	8mer	0.043357
467	8mer	0.043339
468	8mer	0.04333
469	8mer	0.04332
470	8mer	0.043304
471	8mer	0.043299
472	8mer	0.043284
473	8mer	0.043258
474	8mer	0.043251
475	8mer	0.04325
476	8mer	0.043201
477	8mer	0.043186
478	8mer	0.043135
479	8mer	0.043132
480	8mer	0.043127
481	5mer	0.043117
482	8mer	0.0431
483	8mer	0.043092
484	8mer	0.043052
485	8mer	0.043045
486	8mer	0.043035
487	8mer	0.042965
488	8mer	0.042959
489	8mer	0.042959
490	8mer	0.04295
491	8mer	0.042895
492	8mer	0.042872
493	8mer	0.042863
494	8mer	0.042848
495	8mer	0.04278
496	8mer	0.042751
497	8mer	0.042736
498	8mer	0.042727
499	8mer	0.042692
500	8mer	0.042663
501	8mer	0.042634
502	5mer	0.042622
503	8mer	0.042609
504	8mer	0.042582
505	8mer	0.042576
506	8mer	0.042567
507	8mer	0.042554
508	8mer	0.042547
509	8mer	0.042476
510	8mer	0.042462
511	Triplet	0.042416
512	8mer	0.04237
513	8mer	0.04234
514	8mer	0.042305

165	8mer	0.058097	TAGGACAA
166	8mer	0.057951	GCCTTGGG
167	5mer	0.057679	GGCGC
168	8mer	0.057646	AATCCCCG
169	8mer	0.057561	CGCCGTTA
170	8mer	0.057561	GTAAACCG
171	8mer	0.057514	GGCCTATG
172	8mer	0.057493	CGGACGTG
173	8mer	0.057406	GCCGGAAG
174	8mer	0.057352	TTCCCTAA
175	8mer	0.057325	CGCGGGAA
176	8mer	0.057279	GGAATATC
177	8mer	0.057262	GGCGCGTC
178	8mer	0.057246	GGTAACCG
179	8mer	0.057244	AATCCGCC
180	8mer	0.057216	ACATCGCG
181	8mer	0.057143	GAGGGCAG
182	8mer	0.057128	TGATCCGG
183	8mer	0.057127	CTGATCCG
184	8mer	0.057105	GCCGCAGT
185	8mer	0.057074	GACGAGCC
186	8mer	0.05707	GTCGGTAC
187	8mer	0.05707	TCGTAGGG
188	5mer	0.057046	GGGCC
189	8mer	0.056971	CGGGCATC
190	5mer	0.056908	CCCGA
191	8mer	0.056836	CGCCGTGT
192	8mer	0.056612	TGCGGTAA
193	8mer	0.056504	GGAGCCGG
194	8mer	0.056504	ATAGCTAA
195	8mer	0.056455	CGGGATTC
196	8mer	0.056438	GAGACGCC
197	8mer	0.056382	CGTGCTCG
198	8mer	0.056377	CGGGCCCC
199	8mer	0.056307	CGGAGCCC
200	8mer	0.056099	AGCTACGC
201	8mer	0.056053	ATCGCGGG
202	8mer	0.055956	TCGCGCCG
203	8mer	0.055933	GATTCAAC
204	5mer	0.055828	CGGAG
205	8mer	0.055575	GGAATCGT
206	8mer	0.055533	AAAAATTA
207	8mer	0.055492	CGGATCTA
208	8mer	0.05548	AGACCCGC
209	8mer	0.055231	ACCTGTAC
210	8mer	0.0552	GCCCTGGT
211	8mer	0.055116	GTCCAATG
212	8mer	0.055039	CGGCGATG
213	8mer	0.054965	TGAGTGTC
214	8mer	0.054878	TTCAAGCG
215	8mer	0.0548	GAGAATCG
216	5mer	0.054796	CCTGG
217	8mer	0.054507	CAAAACTC
218	8mer	0.054459	AATATACC
219	8mer	0.05445	TATACCAG
220	8mer	0.054389	ACCCGCAG
221	5mer	0.054319	CCGGC
222	8mer	0.054301	CGGTAACC

515	8mer	0.042265
516	SC-PseDNC	0.042253
517	8mer	0.042235
518	5mer	0.042224
519	8mer	0.042223
520	5mer	0.042205
521	8mer	0.042168
522	8mer	0.042149
523	8mer	0.042135
524	8mer	0.042134
525	8mer	0.042068
526	8mer	0.042059
527	8mer	0.042056
528	8mer	0.042056
529	8mer	0.04204
530	8mer	0.042039
531	8mer	0.041942
532	8mer	0.041938
533	8mer	0.041901
534	8mer	0.041879
535	8mer	0.04183
536	8mer	0.04182
537	8mer	0.041809
538	8mer	0.041808
539	8mer	0.041726
540	8mer	0.041704
541	8mer	0.041699
542	Triplet	0.04159
543	8mer	0.041576
544	8mer	0.041533
545	8mer	0.041448
546	8mer	0.041447
547	8mer	0.041437
548	8mer	0.041436
549	8mer	0.041417
550	8mer	0.041416
551	8mer	0.041376
552	8mer	0.041342
553	8mer	0.041297
554	8mer	0.041287
555	8mer	0.041237
556	8mer	0.041175
557	8mer	0.041136
558	8mer	0.04113
559	8mer	0.041106
560	8mer	0.041081
561	8mer	0.041078
562	5mer	0.041038
563	8mer	0.041011
564	8mer	0.040951
565	5mer	0.040906
566	8mer	0.040833
567	8mer	0.040822
568	8mer	0.040799
569	8mer	0.040795
570	8mer	0.040795
571	8mer	0.040795
572	8mer	0.040795

223	8mer	0.054267	TAGAGGCG
224	8mer	0.054134	CGGACTTT
225	5mer	0.05409	CCGAA
226	8mer	0.053955	GCCCCGTG
227	8mer	0.053902	AAAATTAG
228	8mer	0.053867	CGAGCCCG
229	8mer	0.053538	GACCCTCC
230	8mer	0.053492	ATGCAACG
231	8mer	0.053471	TCCAATGC
232	8mer	0.05329	CCTCAGCC
233	8mer	0.053201	CCGCATTC
234	5mer	0.053022	GGGGC
235	8mer	0.052943	ACGGAGGG
236	8mer	0.052902	GCTGCTCG
237	8mer	0.052845	GCGATGGG
238	8mer	0.052795	CTCGCGTC
239	8mer	0.052789	AATCCCAA
240	8mer	0.052597	GAATATTG
241	8mer	0.052512	GACCTAGC
242	8mer	0.052453	TGGCCGTC
243	8mer	0.052419	CCAGGTCC
244	8mer	0.052418	GGGCTGAC
245	8mer	0.052379	TTTGCGGA
246	8mer	0.052338	CCGCCCCC
247	8mer	0.052314	CTGACGAG
248	5mer	0.052297	CGCGG
249	8mer	0.052189	GGGGCCTA
250	8mer	0.052127	GCAGAGGG
251	8mer	0.052092	CGTCGGCG
252	8mer	0.051893	CCCGAAGT
253	8mer	0.051883	GCGGTGTC
254	8mer	0.051652	CGTGAGCC
255	8mer	0.05164	GCCGCATT
256	8mer	0.051615	GGGTCGAG
257	8mer	0.051532	TCGAGTCG
258	8mer	0.051301	TCCCAGCT
259	8mer	0.051286	GAGCGCCA
260	8mer	0.051239	ATTGCCTG
261	8mer	0.051176	CAAGATCG
262	8mer	0.051123	GCGTTCAG
263	8mer	0.051068	GCTCCTCG
264	5mer	0.050999	GCACC
265	8mer	0.050908	ACGGGCAG
266	8mer	0.050867	AGATCGCG
267	8mer	0.050733	GATATTGA
268	8mer	0.050698	AGCCCGAA
269	8mer	0.050688	TTTTTAGT
270	8mer	0.050647	CTGGGCCC
271	8mer	0.050597	CCGCCAGC
272	8mer	0.050593	CTCCGTAG
273	8mer	0.050576	CGGAGGTC
274	8mer	0.050492	CGTCCAAT
275	8mer	0.050489	TGAATGCA
276	8mer	0.050383	GCCTGGCC
277	8mer	0.050332	GGGCGGGC
278	8mer	0.050237	TGCCGGTC
279	8mer	0.0502	TAATACCG
280	8mer	0.050185	ATTACTGC

573	8mer	0.040795
574	8mer	0.040795
575	8mer	0.040786
576	8mer	0.040726
577	8mer	0.040705
578	8mer	0.040688
579	8mer	0.040643
580	8mer	0.04057
581	8mer	0.040562
582	8mer	0.040556
583	8mer	0.040475
584	8mer	0.040464
585	8mer	0.040462
586	8mer	0.040398
587	8mer	0.040398
588	8mer	0.040395
589	8mer	0.040386
590	8mer	0.040378
591	8mer	0.040371
592	8mer	0.040365
593	8mer	0.040326
594	8mer	0.040326
595	8mer	0.040314
596	8mer	0.040306
597	8mer	0.040293
598	8mer	0.040258
599	8mer	0.040252
600	8mer	0.040247
601	8mer	0.04024
602	8mer	0.040219
603	8mer	0.040205
604	8mer	0.040192
605	8mer	0.040188
606	8mer	0.040173
607	8mer	0.040165
608	8mer	0.040044
609	8mer	0.040027
610	8mer	0.039989
611	8mer	0.039986
612	8mer	0.039906
613	8mer	0.039904
614	8mer	0.039873
615	8mer	0.039868
616	8mer	0.039852
617	8mer	0.039832
618	8mer	0.039812
619	8mer	0.039805
620	8mer	0.039756
621	8mer	0.039632
622	8mer	0.039601
623	8mer	0.039546
624	8mer	0.039545
625	8mer	0.039509
626	8mer	0.039489
627	8mer	0.039408
628	8mer	0.039396
629	8mer	0.039384
630	8mer	0.039362

281	8mer	0.050111	CCTGAGCA
282	8mer	0.050078	GACTGTAT
283	8mer	0.050071	GGCGTGCT
284	8mer	0.049934	GGAAACGG
285	8mer	0.049918	GCGGTTCC
286	5mer	0.049887	TCGGC
287	8mer	0.049887	CCTCCGTA
288	8mer	0.049846	GATCGCGC
289	8mer	0.049818	TTGGTGCT
290	8mer	0.049787	CGGAGGCG
291	8mer	0.049757	CGAGATCG
292	8mer	0.049721	TGAACCCG
293	8mer	0.049702	AATAGCCA
294	8mer	0.049633	AGCGGGGC
295	8mer	0.049554	GGTGAACC
296	8mer	0.049369	ATAAATGG
297	8mer	0.049266	CACTGCAC
298	8mer	0.049199	TGCCTGCG
299	8mer	0.049125	CAAGGAGT
300	8mer	0.049102	CATTGCCG
301	5mer	0.049087	CCGGG
302	8mer	0.048983	GCCAGAGT
303	8mer	0.0489	GGCGCCCA
304	8mer	0.048871	AGCCGCAT
305	8mer	0.048848	CTCACGGG
306	8mer	0.048843	CTTCCCTA
307	8mer	0.048756	GATCGTCG
308	8mer	0.048723	GGCTATTG
309	8mer	0.04869	GGGACTGA
310	8mer	0.048651	TGAACACG
311	8mer	0.048619	GGGCGGGG
312	8mer	0.048591	ACTGTGGG
313	8mer	0.048586	GACCCGCA
314	8mer	0.048551	AGACCTGG
315	8mer	0.048515	GACTCGCG
316	8mer	0.048481	CGCCGCCG
317	8mer	0.048465	CGGGAGCG
318	8mer	0.048459	CGCCCGGC
319	8mer	0.048409	GGCGCGGC
320	8mer	0.048382	ATCGTCGT
321	8mer	0.048319	GTTGTACG
322	8mer	0.048104	GAACGGCC
323	8mer	0.048021	GTGGGAGG
324	5mer	0.048021	CCGAG
325	8mer	0.048009	AGGTCCCT
326	8mer	0.047939	CGGCCATC
327	8mer	0.047913	CGAGACTG
328	8mer	0.047904	CCCGAAGG
329	8mer	0.047877	GTTCTGCA
330	8mer	0.047858	CCTGTCGG
331	8mer	0.047854	CGGCGGGG
332	8mer	0.047816	ATCGTAGA
333	8mer	0.047784	CTCGTCAA
334	Triplet	0.047765	Triplet
335	8mer	0.047735	AGTCGGAG
336	8mer	0.047729	CCTCTCAT
337	8mer	0.047708	TGCCCCGA
338	8mer	0.047676	ACTTCGAG

631	8mer	0.039358
632	5mer	0.039357
633	8mer	0.039265
634	8mer	0.039222
635	8mer	0.039208
636	8mer	0.039192
637	8mer	0.039183
638	8mer	0.039178
639	8mer	0.039146
640	8mer	0.039133
641	8mer	0.039092
642	8mer	0.03906
643	5mer	0.039041
644	8mer	0.039038
645	8mer	0.039002
646	8mer	0.039
647	8mer	0.038999
648	8mer	0.038984
649	8mer	0.038973
650	8mer	0.038944
651	8mer	0.038939
652	8mer	0.03893
653	8mer	0.03885
654	8mer	0.038789
655	8mer	0.038745
656	8mer	0.038735
657	8mer	0.03869
658	8mer	0.038687
659	8mer	0.038675
660	8mer	0.038664
661	8mer	0.03865
662	8mer	0.038649
663	8mer	0.038618
664	8mer	0.038616
665	8mer	0.038607
666	8mer	0.0386
667	8mer	0.038583
668	8mer	0.038573
669	8mer	0.038571
670	8mer	0.038566
671	8mer	0.038562
672	8mer	0.038544
673	8mer	0.038543
674	8mer	0.038531
675	8mer	0.038508
676	8mer	0.038481
677	8mer	0.038451
678	8mer	0.038446
679	8mer	0.038361
680	8mer	0.03836
681	8mer	0.038343
682	8mer	0.038332
683	5mer	0.038321
684	8mer	0.038297
685	8mer	0.038275
686	8mer	0.03827
687	8mer	0.038252
688	8mer	0.038251

339	8mer	0.04766	TGCGGGGG
340	8mer	0.047656	GCCCGAAC
341	8mer	0.047521	TGGGCGCC
342	8mer	0.047505	ACATGGCA
343	Triplet	0.047503	Triplet
344	8mer	0.047497	TGCACGCA
345	8mer	0.047407	CGCATTCC
346	8mer	0.047393	CACCGGAC
347	8mer	0.047279	CGCTTGAA
348	8mer	0.047215	GAGATCCC
349	8mer	0.047169	GGTCGAGA
350	8mer	0.047111	AAGTCGGG

689	8mer	0.038249
690	8mer	0.038244
691	8mer	0.038228
692	8mer	0.038226
693	5mer	0.038202
694	8mer	0.038193
695	8mer	0.038168
696	8mer	0.038135
697	8mer	0.038124
698	8mer	0.03805
699	8mer	0.038048
700	8mer	0.038015

f the lncLocPredtp8 according to F-score order

kmer
GGCGGGGC
CCCTGGGC
CGCCGGAA
ATCTTAAT
CCGCGAAC
TAAGGGCT
CGACCTCC
ACGCTCGT
CGCGGGCA
GCGCCAAT
Triplet
GACCCGGC
GGCTGATG
GCACAATT
GGTTGCTA
TGTTTACG
AGGGCGCA
TAGGCATT
CCGAAGTC
CGTACATT
CGCGTACA
CCATCGGT
CGAGACCA
CCGGACTT
ATGTCGGC
CTAGCGGA
TCGACGAA
GGGCGCGG
CCCCTTAT
GGCAGAAA
GTTGCTAT
GGTTGGTC
ACCGTGCC
ACGCCTCG
CGTGTTAC
ATGCGCGA
CTTTCGTC
SC-PseDNC
GGGGCGCG
ACCGCGAA
GGGGATAT
TTTGTCTC
AGATTGCG
CGAGAATC
CAGTTCGG
ATCGACGA
CACGGCAA
ACCCCCGA
CGAGAGCT
GCGACTCG
CCGCTGGT
ACTGCTCA
GAGCCCCG
CCAAAGTG
CCGCAGCG

Order	Feature	F-score	kmer
701	8mer	0.037312	GCCTATGG
702	8mer	0.03727	AATACCGG
703	8mer	0.037266	CCCCGCCC
704	8mer	0.037257	CGGCCTAT
705	8mer	0.037252	GCTGCGCA
706	8mer	0.037222	CCCGGCAC
707	8mer	0.037219	ACCCGGGG
708	8mer	0.037189	AAGCGAAA
709	8mer	0.037161	GAATCGCT
710	8mer	0.037139	GCGCCCGG
711	8mer	0.037131	ACCCGGCC
712	8mer	0.037086	ATTAGGGA
713	8mer	0.037082	CCGCTTAG
714	8mer	0.037036	GGCATTCG
715	8mer	0.037023	GGCATTCC
716	8mer	0.036989	TAAAATGG
717	8mer	0.036988	TCCTTTCC
718	8mer	0.036971	CCTCCCAA
719	8mer	0.036962	CGGAATAT
720	8mer	0.036961	TTGGGCCC
721	8mer	0.036937	CGCCAATC
722	8mer	0.036924	TGACCTTA
723	8mer	0.036905	GCGAAATT
724	8mer	0.036883	CTTCAAGC
725	8mer	0.036856	AGGCTGGT
726	8mer	0.036827	GCGCGCGG
727	8mer	0.036785	GTGCCGAA
728	8mer	0.036765	TTTCCCGA
729	8mer	0.036755	GAATCCGG
730	8mer	0.036745	CCGACTTT
731	8mer	0.036719	GTAGTGGG
732	8mer	0.036678	CAAAAAAA
733	8mer	0.036676	GCGACATC
734	8mer	0.036655	ACGGGTTC
735	8mer	0.036648	TTCAAGGT
736	8mer	0.036629	TGTACGTT
737	8mer	0.036619	GTCGACCC
738	8mer	0.036608	GCGCCACT
739	8mer	0.036604	CTGGTATA
740	8mer	0.036599	CGTCAAGG
741	8mer	0.036577	GACTGCGA
742	8mer	0.036576	CAGTAGGA
743	8mer	0.036575	CGGCCCAC
744	8mer	0.036574	GAAGTCGG
745	8mer	0.036528	ATGCACGC
746	8mer	0.036518	TAACGAGG
747	8mer	0.036479	TTCGGGGG
748	8mer	0.036471	CGGGTGGC
749	8mer	0.036463	CAATATAC
750	8mer	0.036425	TCGGCTGT
751	8mer	0.03639	CTATGTTG
752	8mer	0.036376	ACGCTCAT
753	8mer	0.036373	GCGTTCTC
754	8mer	0.03628	GACGGCCT
755	8mer	0.036274	GAGGTCAT

CACCGCGC	756	8mer	0.036257	ACGGCCTG
GCCGCCGC	757	8mer	0.036192	GTTATCTG
GCCCCTGA	758	8mer	0.03619	CTACAGTA
CAAAGTAC	759	8mer	0.036189	TATAAGGT
GGAGGGCG	760	8mer	0.036163	CGACTTTG
GGGGGTCG	761	8mer	0.036153	TGAGCCGA
GCCGGACT	762	8mer	0.036139	CGTAGATA
GCGCCGGA	763	8mer	0.036137	CTAAACCG
GCAACGTC	764	8mer	0.036098	CTTGCGTC
ACCTAGAC	765	8mer	0.036073	AATTATCC
TTTTTGTA	766	8mer	0.036068	GAGCGACG
AGATACTA	767	8mer	0.036049	GTCCAGAC
AACTCGTC	768	8mer	0.036048	CGGGGGCC
GGAACCTA	769	8mer	0.036041	ACGCCCGA
AAGGTGCC	770	8mer	0.036032	GGCGCCCG
TATTAGGG	771	8mer	0.036022	GTTTCGCC
CCCGAACT	772	8mer	0.036005	TCCTGGGC
TTGTACGG	773	8mer	0.036004	TTGTTAAT
CCGTGTCC	774	8mer	0.03599	CAGCCTCC
CGCCTATA	775	8mer	0.035982	GGTGGCGC
TATGTCGG	776	8mer	0.035958	CGTTCAGG
TACTGCCA	777	8mer	0.035936	CGAGTTCT
TATAGCAC	778	8mer	0.03593	CGCGCGGG
GGTTGCAG	779	8mer	0.035852	ACCCGCGC
TTTGGCCA	780	8mer	0.035833	ACCTCGAC
CCGCGATT	781	8mer	0.035761	CGAACAGA
GGTTGTTC	782	8mer	0.035756	CTGTCCCG
GAGCGGGG	783	8mer	0.035747	CTCCTGTG
GAGGCCGT	784	8mer	0.035746	GGCTCACG
GTGTCGTA	785	8mer	0.035744	GACCTGGG
CTGCGAGT	786	8mer	0.035735	GAACTCGT
CGGCCCCT	787	8mer	0.035724	ACTGCACG
GCAGTTCG	788	8mer	0.035716	GAGGGCCG
TCCTGTGT	789	8mer	0.03568	ACCATTAA
GGGCGGTG	790	8mer	0.035668	CAAGCGAA
CCGTTTCC	791	8mer	0.035666	GCGGCCAA
TCGTCGAA	792	8mer	0.035652	GCCATCTA
TTGCGCCA	793	8mer	0.035597	GTCTAGGC
AGCATTGC	794	8mer	0.035581	ATCATTAT
ACGCGTAC	795	8mer	0.035577	TGGGCCCC
CTTCCGCG	796	8mer	0.035562	TAACCGCG
GGGGCGGT	797	8mer	0.035558	GGCTGGGA
ACCCGCAC	798	8mer	0.035557	GTAAACAT
GAGCCCGG	799	8mer	0.035551	CGTCCAGG
ATCCCGAC	800	8mer	0.035506	CGGACTAG
GAACGAAT	801	8mer	0.035482	GGACCTAG
CAGTACGT	802	8mer	0.035439	TAGTGTCC
CTTAAGTA	803	8mer	0.035432	CACTACAG
GAGATACT	804	8mer	0.035424	CGCTTCTT
CGTCGTCT	805	8mer	0.035384	TCCCAACG
GACGCGTA	806	8mer	0.035383	CCGGGGTG
GCACTCCA	807	8mer	0.035364	CGGCGCGC
CGCGCGGC	808	8mer	0.03536	TGGGATTA
TGCTTTTT	809	8mer	0.035351	AGCCGGGG
CCGGACTA	810	8mer	0.035334	TTCACTAG
ATTGCGCC	811	8mer	0.03528	CCCTTCGA
CGAGGTCT	812	8mer	0.03522	CTGCGCCT
CAGGCCAC	813	8mer	0.03518	CTCGGCTG

GATCCGGG	814	8mer	0.035126	CACGCCAC
TCCTCCGT	815	8mer	0.035113	CTCGGGTC
CTAAACAT	816	8mer	0.035093	CGGCGTGC
CGGGCGAG	817	8mer	0.035069	CTGTCGAC
CACAGATG	818	8mer	0.035065	TAGTGGGA
GTAAGGGC	819	8mer	0.035051	CTCAGCCT
GGACGCAA	820	8mer	0.03505	GGCCGCTC
CCTGCGCC	821	8mer	0.035019	GGCCAGGT
GGCGGTGT	822	8mer	0.035018	TTCACGCC
TGCACTTG	823	8mer	0.035013	CTGGCCAC
AGGAGGCC	824	8mer	0.034942	CTGGGCTG
Triplet	825	8mer	0.034939	GTCGGAAA
CCGAAGGG	826	8mer	0.034897	GCACTTGT
TGAGCCAC	827	8mer	0.034893	TCGCTTGA
TCGGGGGT	828	8mer	0.034842	ACTCAGTG
CCCGGGAG	829	8mer	0.034833	GAACTTAA
SC-PseDNC	830	8mer	0.034822	CCCCAGGC
CAACGTCT	831	8mer	0.034799	TCCCGCAA
CCGAAGAC	832	8mer	0.034779	TCTTAAGT
TCGAGAGC	833	8mer	0.034736	TGTGGCCG
ACGGCCCA	834	8mer	0.034735	CCCTGATT
CCGTGGAA	835	8mer	0.03473	CGCAAAAG
AACTGCGG	836	8mer	0.034724	ACAAGCGA
CCAGGCAC	837	8mer	0.034711	CGGGCGTT
AGGTCGTA	838	8mer	0.034709	GAGGGCCT
GCCCGGCA	839	8mer	0.034708	GGCGTCGG
AAGCGAGG	840	8mer	0.034707	ACCAGGGG
TGCTGCGC	841	8mer	0.0347	AGGCACCT
CGGAGGTT	842	8mer	0.034691	CTAATGGT
GTTCCCAC	843	8mer	0.034684	ACCGCTGG
AGACGGTC	844	8mer	0.034665	CATGGCAT
CCATAGTG	845	8mer	0.034628	GCGTTGTT
CGAAGGCG	846	8mer	0.034567	CACCGAAG
GCCGCCAG	847	8mer	0.034557	CTGAAAAA
AGCTTTTG	848	8mer	0.034523	CGCGGGAG
CCCGCCCC	849	8mer	0.034485	CGGTTAAC
TCGACCTC	850	8mer	0.034454	CGGGAGTT
CCCGGAGG	851	8mer	0.03441	CGCCGCCT
CACCACTT	852	8mer	0.034398	CCCTCACG
GACCTCGG	853	8mer	0.034398	AGGGCCGG
Triplet	854	8mer	0.034374	CCCGGGCT
TCCGGTTG	855	8mer	0.034353	AGGGGCGT
CGGAAACG	856	8mer	0.034352	GGGGGACG
ACAATGAG	857	8mer	0.034348	CCGTATTA
CCCTCGTC	858	8mer	0.034348	GGATTACA
AGGTTCAA	859	8mer	0.034336	CCCTACCA
GCCCCCCG	860	8mer	0.03431	GAAATTGA
CCGGGCCC	861	8mer	0.034267	GTCGGACC
GGGAGATC	862	8mer	0.034215	TTTTGTAT
TACTCGGG	863	8mer	0.034207	CCGGGCAT
TACGTAAA	864	8mer	0.034187	TTGGAAAG
TCGCGGGC	865	8mer	0.034168	GATGCCCG
CCCGTATT	866	8mer	0.034158	GTGCCGGG
GACCTGGC	867	8mer	0.034099	CCGGGGCG
GCCCGAGG	868	8mer	0.034094	GAAGGCGG
GCATTCCG	869	8mer	0.034065	GCAATGCG
CATTCGAT	870	8mer	0.034064	TCCCCTCG
GGGGCGGG	871	8mer	0.034041	GCCCACAA

AAGGGTCC	872	8mer	0.034036	GCTGCTAG
TTGCGGAA	873	8mer	0.034008	GGCGGCAC
CTAATGCA	874	8mer	0.034007	ACCTGGTC
CCCGTCTC	875	8mer	0.033999	AACAAGCG
CATCGCCG	876	8mer	0.03399	TGCCTAAG
CCGCCGGC	877	8mer	0.033985	TACAAGCG
GGGACCGC	878	8mer	0.033952	CTGGTCTC
TAACGGGA	879	8mer	0.033917	GGTCTCGA
CGCGATGA	880	8mer	0.033911	TATGATAC
ACGCGCTA	881	8mer	0.0339	CCGCGCTT
GCGCTAGG	882	8mer	0.033872	GATTTTTA
GTAACGGG	883	8mer	0.033869	GCACCCCT
GTCGTAAT	884	8mer	0.033854	CCGTAGAC
CTGGGCCT	885	8mer	0.033778	TCGTGACA
CCTGGCGT	886	8mer	0.033767	GTCCCTGG
TGGGACCT	887	8mer	0.033762	CGAATTAC
CGTAATAC	888	8mer	0.03376	GTGCCTAA
GGCGTTCA	889	8mer	0.033714	ACTGCGAT
TCACGCCA	890	8mer	0.033696	GCACCACG
CGAACACC	891	8mer	0.033673	CCGCCGTT
TGCTTACC	892	8mer	0.033602	CGCAGCGA
AACGGGTT	893	8mer	0.033572	TGTTACAT
CACGCCTC	894	8mer	0.033562	TGACGGGC
GGGATCGC	895	8mer	0.033561	AATACGGA
CACCCCCG	896	8mer	0.033552	ATAGGTAC
CGAACGGG	897	8mer	0.033546	TTTGCAGC
CGTCCAGC	898	8mer	0.033524	AACCCGCG
TGGCGTGC	899	8mer	0.03352	CGTATGTG
TGGGACTG	900	8mer	0.033512	CTGCTGTT
CCCCTGAC	901	8mer	0.033509	AATCACAA
GCCGGCTG	902	8mer	0.033465	TCGGGACG
TCCTACCT	903	8mer	0.033421	AGCGAAAT
CGCTGGTA	904	8mer	0.03342	GCCCTGGC
GGGCCTAG	905	8mer	0.033374	CTCGGTTC
CTGACGGC	906	8mer	0.033372	AACGGACG
GGATTATA	907	8mer	0.033367	GGTCGCCA
CGACCAAA	908	8mer	0.033338	CGAATCCC
GAACGTGC	909	8mer	0.033335	GGGCGGCA
TTGGCCAG	910	8mer	0.033317	GTGCGGTA
ACACGGGA	911	8mer	0.033292	GAATGCAC
AGCCGTTT	912	8mer	0.033279	CTGTCGGT
TGCTCACG	913	8mer	0.033276	CGGCCGGG
AGGAGGGG	914	8mer	0.033224	CGTCACCT
GGAGCGCT	915	8mer	0.033193	GCCGTGTC
CCAATGCC	916	8mer	0.033192	CATCATTA
ATCGCGCC	917	8mer	0.033165	TCGGCCTC
GGGTGTTT	918	8mer	0.033164	GGGATATC
GCCACTGC	919	8mer	0.033131	CCTATATC
GGGAGGCT	920	8mer	0.033087	CAGCCAAG
TATTGAAG	921	8mer	0.033084	AGGGGCGG
CGAACTCG	922	8mer	0.033055	GCCGGCCG
CGCTAAAT	923	8mer	0.033036	GGCCGAAT
TAATCGAG	924	8mer	0.033024	TCGTTGAA
AACCCGAA	925	8mer	0.033004	GCTCACAC
CGCGAACG	926	8mer	0.032975	CCCAGTGC
CTGTGGGG	927	8mer	0.032933	CCGCCGCC
CCTCGCGG	928	8mer	0.032904	GACCTCCC
CGAGATGG	929	8mer	0.032903	GCCGATCG

GCCGGTCT	930	8mer	0.032903	CGACGAAG
CATGTAAT	931	8mer	0.032903	CCGTATAC
TCGCGTCC	932	8mer	0.032903	ACTCGCGC
TCGATGAG	933	8mer	0.032903	CCGCTATA
GGAGCACT	934	8mer	0.032903	CGGTGCTA
AACGCCTG	935	8mer	0.032903	TTTAGCGC
CGGCGCTG	936	8mer	0.032903	CACGAACG
GCGCCGAG	937	8mer	0.032903	CCGTATCG
CTCGAGAT	938	8mer	0.032903	CTTACGCG
CGAAATTA	939	8mer	0.032903	CGGTAATG
CAATGCCC	940	8mer	0.032903	CGGCTAAC
TCGACCCG	941	8mer	0.032903	CGGGGTAA
TATACTCG	942	8mer	0.032903	CGCGTTGA
GGCGCGTG	943	8mer	0.032903	CGACGATA
GGTAGCGT	944	8mer	0.032903	ATCGGCTA
GGTCGAGG	945	8mer	0.032903	ACGTCGTC
TTTTTTAG	946	8mer	0.032903	CGCTTATG
CCGCCAAT	947	8mer	0.032903	TTGCGTTG
TGGGGCGG	948	8mer	0.032903	ACGACGAT
TACCGGCA	949	8mer	0.032903	ATCGCGCA
CATGCTAC	950	8mer	0.032903	TCGTACAC
CCCGCAAA	951	8mer	0.032903	CCGACGAC
AACTGTAC	952	8mer	0.032903	GGTGCGAC
GCAGAAAA	953	8mer	0.032903	CGTTCGGC
CACGCAGG	954	8mer	0.032903	TCGAGACG
GCATCGTC	955	8mer	0.032903	ACGCGTCG
GAGGGCGA	956	8mer	0.032903	CACCCGAT
GGTGAAAC	957	8mer	0.032903	CCGGATTA
CGGATGTC	958	8mer	0.032903	GTCGGAGT
GCGGAGCG	959	8mer	0.032903	ATAGCGCC
AAGCCGCA	960	8mer	0.032903	TATAGCGC
TGACGCGA	961	8mer	0.032903	ACGAGCCG
CCACCGAT	962	8mer	0.032903	CGACTCGA
TGCGTGAG	963	8mer	0.032903	TACGGCGC
TCATTGCC	964	8mer	0.032903	ACGTAATC
AGCGAGAC	965	8mer	0.032903	ACCGACGG
TGTGTTAC	966	8mer	0.032903	GTGCGACG
CCGTAGCT	967	8mer	0.032903	ACGAGTCA
ACATCCGT	968	8mer	0.032903	CTATCACG
CCTCGTCG	969	8mer	0.032903	ATCTCGAA
CTGCGCCG	970	8mer	0.032903	CGATCCGA
CTATGGGA	971	8mer	0.032903	CCCGGTTA
TGCAACCT	972	8mer	0.032903	TCGGAACG
GATTAGAG	973	8mer	0.032903	TGCCGCTA
TCGCCGCC	974	8mer	0.032903	ATCGGTCC
AGGATGTT	975	8mer	0.032903	CGTACCAA
CCGCGGAT	976	8mer	0.032903	ATCCCGTA
ATGAGCTA	977	8mer	0.032903	CCCGTACC
ACTCGTCA	978	8mer	0.032903	CGGAACCC
GCGGAGGT	979	8mer	0.032903	ATCGTCCG
ATAACGAG	980	8mer	0.032903	TTTGTCCG
GAATATCT	981	8mer	0.032903	TCGCGCTA
GACACGGG	982	8mer	0.032903	TACCCGCA
AACATGGT	983	8mer	0.032903	CGACGGAC
ATAGTGCG	984	8mer	0.032903	ACTCGCCG
GACGGGCG	985	8mer	0.032903	CGGCACGA
TGCAAGTT	986	8mer	0.032903	TGCGACGA
CGCGGATT	987	8mer	0.032903	TGTGCGAC

GCCCGGAG
ACGTCCGG
TAGATCGT
GCGGAGTT
ACGCCACA
ATGGGACC
CGTCATGA
CATCCGTC
GAGGGCGC
TGATTCGC
TGCGCTCC
GGGCTCAC
GCACGTGG
GGTGTCGT
CGGCTGGG
CTGACGCG
TTAAATGA
TCACCGCG
CTCACATC
TGCGCCGA
GCAAACGT
CCGGGAAC
TGGCCAGG
CGAATGAT
ACGCGACA
GACGCGAC
TGTCGACC
TAAACCGC
GCTCGCGT
CGGCCTCC
GACGCCTC
TTCAATGA
CTCGTCGA
GGCCGTCG
TGGGGTAT
TGTCCCGC
CCTTCACT
ATATGTGA
CCGAACTC
TGGTTGCT
CTGCTCAC
CAGGCGGG
GCCCGTCG
GCGGGCAT
CACGAGAC
GCAATGCC
CTGCTGCG
CCGAGGGC
AGGTCGTG
GGCTGGTC
GAAAGCGA
GGGGCTGA
GGCTGCTA
CTTCATCG
CGTGGGTG
ATTAGCCG
GCTGCCGG
AAGCCGGC

988	8mer	0.032903	CGCGGATC
989	8mer	0.032903	GTCGGCGG
990	8mer	0.032903	CGAGGCGT
991	8mer	0.032869	CGGGGTGG
992	8mer	0.032866	CCCCGAAG
993	8mer	0.032861	CTCACTGC
994	8mer	0.032859	GGACGACT
995	8mer	0.032837	AAACCGCT
996	8mer	0.03282	GACCATTA
997	8mer	0.03279	GGTTTCGC
998	8mer	0.032776	AGTCGGGA
999	8mer	0.032775	TGCCCGTC
1000	Triplet	0.032742	Triplet

GGGGGCGG
CCCGTCTG
CACGGGAT
ACCCGAAT
TCGCCCCG

f the lncLocPredtp58 according to F-score order

kmer
CGACCCGG
GCGCCGTG
TGCGTCAA
TGGACAGC
TCGTCAAC
CACGGGAG
ACGACCCG
TATCCAGT
GGTCATCG
CCGCA
ACGCAGGG
CGCGCCGT
CTCCCAAA
CGGGGCGG
TTATGTCG
GCGCAGCA
GACCGCGA
AGAAGCGG
AGGGC
ATCGCCGA
TGTCCGGT
AGCAGTCC
AGCTCGGC
TCTGGTAT
CGGGCACT
GTTAATTT
CAGGACGA
GGTCGGGC
CGGGGCCA
CGCCCGAG
TCGCCCTC
GGCGGGGC
CCCTGGGC
CGCCGGAA
ATCTTAAT
CCGCGAAC
TAAGGGCT
CGACCTCC
ACGCTCGT
CGCGGGCA
GCGCCAAT
Triplet
GACCCGGC
GGCTGATG
GCACAATT
CGGCC
GGTTGCTA
TGTTTACG

Order	Feature	F-score	kmer
701	8mer	0.037947	TGCGCCGA
702	8mer	0.037943	GCAAACGT
703	8mer	0.037942	CCGGGAAC
704	8mer	0.037932	TGGCCAGG
705	8mer	0.037889	CGAATGAT
706	8mer	0.037866	ACGCGACA
707	8mer	0.037866	GACGCGAC
708	8mer	0.037866	TGTCGACC
709	8mer	0.037866	TAAACCGC
710	8mer	0.037866	GCTCGCGT
711	8mer	0.037861	CGGCCTCC
712	8mer	0.037833	GACGCCTC
713	8mer	0.037833	TTCAATGA
714	8mer	0.037822	CTCGTCGA
715	8mer	0.037764	GGCCGTCG
716	8mer	0.037759	TGGGGTAT
717	8mer	0.037755	TGTCCCGC
718	8mer	0.03775	CCTTCACT
719	8mer	0.037746	ATATGTGA
720	8mer	0.037736	CCGAACTC
721	5mer	0.037698	TGCCC
722	8mer	0.037682	TGGTTGCT
723	8mer	0.037677	CTGCTCAC
724	8mer	0.037644	CAGGCGGG
725	8mer	0.03761	GCCCGTCG
726	8mer	0.037582	GCGGGCAT
727	8mer	0.037572	CACGAGAC
728	8mer	0.037572	GCAATGCC
729	8mer	0.037561	CTGCTGCG
730	8mer	0.037556	CCGAGGGC
731	8mer	0.037538	AGGTCGTG
732	8mer	0.037517	GGCTGGTC
733	8mer	0.037512	GAAAGCGA
734	8mer	0.037508	GGGGCTGA
735	5mer	0.037508	GGCGA
736	8mer	0.037503	GGCTGCTA
737	8mer	0.037495	CTTCATCG
738	8mer	0.03748	CGTGGGTG
739	8mer	0.03746	ATTAGCCG
740	8mer	0.037429	GCTGCCGG
741	8mer	0.037404	AAGCCGGC
742	8mer	0.037394	GGGGGCGG
743	5mer	0.037369	CAGGC
744	8mer	0.037359	CCCGTCTG
745	8mer	0.037355	CACGGGAT
746	8mer	0.037317	ACCCGAAT
747	8mer	0.037314	TCGCCCCG
748	8mer	0.037312	GCCTATGG

AGGGCGCA	749	8mer	0.03727	AATACCGG
TAGGCATT	750	8mer	0.037266	CCCCGCCC
CCGAAGTC	751	8mer	0.037257	CGGCCTAT
CGTACATT	752	8mer	0.037252	GCTGCGCA
CGCGTACA	753	8mer	0.037222	CCCGGCAC
CCATCGGT	754	8mer	0.037219	ACCCGGGG
CGAGACCA	755	8mer	0.037189	AAGCGAAA
CCGGACTT	756	8mer	0.037161	GAATCGCT
ATGTCGGC	757	8mer	0.037139	GCGCCCGG
CTAGCGGA	758	8mer	0.037131	ACCCGGCC
TCGACGAA	759	8mer	0.037086	ATTAGGGA
GGGCGCGG	760	8mer	0.037082	CCGCTTAG
CCCCTTAT	761	8mer	0.037036	GGCATTCG
GGCAGAAA	762	8mer	0.037023	GGCATTCC
GTTGCTAT	763	8mer	0.036989	TAAAATGG
GGTTGGTC	764	8mer	0.036988	TCCTTTCC
ACCGTGCC	765	8mer	0.036971	CCTCCCAA
ACGCCTCG	766	8mer	0.036962	CGGAATAT
CGTGTTAC	767	8mer	0.036961	TTGGGCCC
ATGCGCGA	768	8mer	0.036937	CGCCAATC
CTTTCGTC	769	8mer	0.036924	TGACCTTA
SC-PseDNC	770	8mer	0.036905	GCGAAATT
GGGGCGCG	771	8mer	0.036883	CTTCAAGC
ACCGCGAA	772	8mer	0.036856	AGGCTGGT
GGGGATAT	773	8mer	0.036827	GCGCGCGG
TTTGTCTC	774	8mer	0.036785	GTGCCGAA
AGATTGCG	775	8mer	0.036765	TTTCCCGA
CGAGAATC	776	8mer	0.036755	GAATCCGG
GGCGT	777	8mer	0.036745	CCGACTTT
CAGTTCGG	778	8mer	0.036719	GTAGTGGG
ATCGACGA	779	5mer	0.036696	GAGGG
CACGGCAA	780	8mer	0.036678	CAAAAAAA
TCGCG	781	8mer	0.036676	GCGACATC
ACCCCCGA	782	8mer	0.036655	ACGGGTTC
CGAGAGCT	783	8mer	0.036648	TTCAAGGT
GCGACTCG	784	8mer	0.036629	TGTACGTT
CCGCTGGT	785	8mer	0.036619	GTCGACCC
ACTGCTCA	786	8mer	0.036608	GCGCCACT
GAGCCCCG	787	8mer	0.036604	CTGGTATA
CCAAAGTG	788	8mer	0.036599	CGTCAAGG
CCGCAGCG	789	8mer	0.036577	GACTGCGA
CACCGCGC	790	8mer	0.036576	CAGTAGGA
GCCGCCGC	791	5mer	0.036576	CGGGA
GCCCCTGA	792	8mer	0.036575	CGGCCCAC
CAAAGTAC	793	8mer	0.036574	GAAGTCGG
GGAGGGCG	794	8mer	0.036528	ATGCACGC
GGGGGTCG	795	8mer	0.036518	TAACGAGG
GCCGGACT	796	8mer	0.036479	TTCGGGGG
GCGCCGGA	797	8mer	0.036471	CGGGTGGC
GCAACGTC	798	8mer	0.036463	CAATATAC
ACCTAGAC	799	8mer	0.036425	TCGGCTGT
TTTTTGTA	800	8mer	0.03639	CTATGTTG
AGATACTA	801	8mer	0.036376	ACGCTCAT
AACTCGTC	802	8mer	0.036373	GCGTTCTC
GGAACCTA	803	8mer	0.03628	GACGGCCT
AAGGTGCC	804	8mer	0.036274	GAGGTCAT
TATTAGGG	805	8mer	0.036257	ACGGCCTG
CCCGAACT	806	8mer	0.036192	GTTATCTG

TTGTACGG	807	8mer	0.03619	CTACAGTA
CCGTGTCC	808	8mer	0.036189	TATAAGGT
CGCCTATA	809	8mer	0.036163	CGACTTTG
TATGTCGG	810	8mer	0.036153	TGAGCCGA
TACTGCCA	811	8mer	0.036139	CGTAGATA
TATAGCAC	812	8mer	0.036137	CTAAACCG
GGTTGCAG	813	8mer	0.036098	CTTGCGTC
TTTGGCCA	814	8mer	0.036073	AATTATCC
CCGCGATT	815	8mer	0.036068	GAGCGACG
GGTTGTTC	816	8mer	0.036049	GTCCAGAC
GAGCGGGG	817	8mer	0.036048	CGGGGGCC
GAGGCCGT	818	8mer	0.036041	ACGCCCGA
GTGTCGTA	819	5mer	0.036039	GGCCC
CTGCGAGT	820	8mer	0.036032	GGCGCCCG
CGGCCCCT	821	8mer	0.036022	GTTTCGCC
GCAGTTCG	822	8mer	0.036005	TCCTGGGC
TCCTGTGT	823	8mer	0.036004	TTGTTAAT
GGGCGGTG	824	8mer	0.03599	CAGCCTCC
CCGTTTCC	825	8mer	0.035982	GGTGGCGC
TCGTCGAA	826	8mer	0.035958	CGTTCAGG
TTGCGCCA	827	8mer	0.035936	CGAGTTCT
AGCATTGC	828	8mer	0.03593	CGCGCGGG
ACGCGTAC	829	5mer	0.035879	GCGGA
CTTCCGCG	830	8mer	0.035852	ACCCGCGC
CCCGG	831	8mer	0.035833	ACCTCGAC
GGGGCGGT	832	8mer	0.035761	CGAACAGA
ACCCGCAC	833	8mer	0.035756	CTGTCCCG
GAGCCCGG	834	8mer	0.035747	CTCCTGTG
ATCCCGAC	835	8mer	0.035746	GGCTCACG
GAACGAAT	836	8mer	0.035744	GACCTGGG
CAGTACGT	837	8mer	0.035735	GAACTCGT
CTTAAGTA	838	8mer	0.035724	ACTGCACG
GAGATACT	839	8mer	0.035716	GAGGGCCG
CGTCGTCT	840	8mer	0.03568	ACCATTAA
GACGCGTA	841	8mer	0.035668	CAAGCGAA
GCACTCCA	842	8mer	0.035666	GCGGCCAA
CGCGCGGC	843	8mer	0.035652	GCCATCTA
TGCTTTTT	844	5mer	0.035632	ACCTC
CCGGACTA	845	8mer	0.035597	GTCTAGGC
ATTGCGCC	846	8mer	0.035581	ATCATTAT
CGAGGTCT	847	8mer	0.035577	TGGGCCCC
CAGGCCAC	848	8mer	0.035562	TAACCGCG
GATCCGGG	849	8mer	0.035558	GGCTGGGA
TCCTCCGT	850	8mer	0.035557	GTAAACAT
CTAAACAT	851	8mer	0.035551	CGTCCAGG
ACGGG	852	8mer	0.035506	CGGACTAG
CGGGCGAG	853	8mer	0.035482	GGACCTAG
CACAGATG	854	8mer	0.035439	TAGTGTCC
GTAAGGGC	855	8mer	0.035432	CACTACAG
GGACGCAA	856	8mer	0.035424	CGCTTCTT
CCTGCGCC	857	8mer	0.035384	TCCCAACG
GGCGGTGT	858	8mer	0.035383	CCGGGGTG
TGCACTTG	859	8mer	0.035364	CGGCGCGC
AGGAGGCC	860	8mer	0.03536	TGGGATTA
Triplet	861	8mer	0.035351	AGCCGGGG
CCGAAGGG	862	8mer	0.035334	TTCACTAG
TGAGCCAC	863	8mer	0.03528	CCCTTCGA
TCGGGGGT	864	8mer	0.03522	CTGCGCCT

CCCGGGAG		865	8mer	0.03518	CTCGGCTG
SC-PseDNC		866	8mer	0.035126	CACGCCAC
CAACGTCT		867	8mer	0.035113	CTCGGGTC
CGAGG		868	8mer	0.035093	CGGCGTGC
CCGAAGAC		869	8mer	0.035069	CTGTCGAC
CTCGG		870	8mer	0.035065	TAGTGGGA
TCGAGAGC		871	8mer	0.035051	CTCAGCCT
ACGGCCCA		872	8mer	0.03505	GGCCGCTC
CCGTGGAA		873	8mer	0.035019	GGCCAGGT
AACTGCGG		874	8mer	0.035018	TTCACGCC
CCAGGCAC		875	8mer	0.035013	CTGGCCAC
AGGTCGTA		876	8mer	0.034942	CTGGGCTG
GCCCGGCA		877	8mer	0.034939	GTCGGAAA
AAGCGAGG		878	8mer	0.034897	GCACTTGT
TGCTGCGC		879	8mer	0.034893	TCGCTTGA
CGGAGGTT		880	5mer	0.034853	GCCGA
GTTCCCAC		881	8mer	0.034842	ACTCAGTG
AGACGGTC		882	8mer	0.034833	GAACTTAA
CCATAGTG		883	8mer	0.034822	CCCCAGGC
CGAAGGCG		884	8mer	0.034799	TCCCGCAA
GCCGCCAG		885	8mer	0.034779	TCTTAAGT
AGCTTTTG		886	5mer	0.034747	CGCAC
CCCGCCCC		887	8mer	0.034736	TGTGGCCG
TCGACCTC		888	8mer	0.034735	CCCTGATT
CCCGGAGG		889	8mer	0.03473	CGCAAAAG
CACCACTT		890	8mer	0.034724	ACAAGCGA
GACCTCGG		891	8mer	0.034711	CGGGCGTT
Triplet		892	8mer	0.034709	GAGGGCCT
TCCGGTTG		893	8mer	0.034708	GGCGTCGG
CGGAAACG		894	8mer	0.034707	ACCAGGGG
ACAATGAG		895	8mer	0.0347	AGGCACCT
CCCTCGTC		896	8mer	0.034691	CTAATGGT
AGGTTCAA		897	8mer	0.034684	ACCGCTGG
GCCCCCCG		898	8mer	0.034665	CATGGCAT
CCGGGCCC		899	8mer	0.034628	GCGTTGTT
GGGAGATC		900	8mer	0.034567	CACCGAAG
TACTCGGG		901	8mer	0.034557	CTGAAAAA
TACGTAAA		902	8mer	0.034523	CGCGGGAG
TCGCGGGC		903	8mer	0.034485	CGGTTAAC
CCCGTATT		904	8mer	0.034454	CGGGAGTT
GACCTGGC		905	8mer	0.03441	CGCCGCCT
GCCCGAGG		906	8mer	0.034398	CCCTCACG
GCATTCCG		907	5mer	0.034398	GGCCT
CATTCGAT		908	8mer	0.034398	AGGGCCGG
GGGGCGGG		909	5mer	0.034375	CGGCA
AAGGGTCC		910	8mer	0.034374	CCCGGGCT
TTGCGGAA		911	8mer	0.034353	AGGGGCGT
GCGCG		912	8mer	0.034352	GGGGGACG
CTAATGCA		913	8mer	0.034348	CCGTATTA
CCCGTCTC		914	8mer	0.034348	GGATTACA
GCCCC		915	8mer	0.034336	CCCTACCA
CATCGCCG		916	8mer	0.03431	GAAATTGA
CCGCCGGC		917	8mer	0.034267	GTCGGACC
GGGACCGC		918	8mer	0.034215	TTTTGTAT
TAACGGGA		919	8mer	0.034207	CCGGGCAT
CGCGATGA		920	8mer	0.034187	TTGGAAAG
ACGCGCTA		921	8mer	0.034168	GATGCCCG
GCGCTAGG		922	8mer	0.034158	GTGCCGGG

GTAACGGG	923	8mer	0.034099	CCGGGGCG
GTCGTAAT	924	8mer	0.034094	GAAGGCGG
CTGGGCCT	925	8mer	0.034065	GCAATGCG
CCTGGCGT	926	8mer	0.034064	TCCCCTCG
TGGGACCT	927	8mer	0.034041	GCCCACAA
CGTAATAC	928	8mer	0.034036	GCTGCTAG
GGCGTTCA	929	8mer	0.034008	GGCGGCAC
TCACGCCA	930	8mer	0.034007	ACCTGGTC
CGAACACC	931	8mer	0.033999	AACAAGCG
TGCTTACC	932	8mer	0.03399	TGCCTAAG
AACGGGTT	933	8mer	0.033985	TACAAGCG
CACGCCTC	934	8mer	0.033952	CTGGTCTC
GGGATCGC	935	8mer	0.033917	GGTCTCGA
CACCCCCG	936	8mer	0.033911	TATGATAC
CGAACGGG	937	8mer	0.0339	CCGCGCTT
CGTCCAGC	938	8mer	0.033872	GATTTTTA
TGGCGTGC	939	8mer	0.033869	GCACCCCT
TGGGACTG	940	8mer	0.033854	CCGTAGAC
CCCCTGAC	941	8mer	0.033778	TCGTGACA
GCCGGCTG	942	8mer	0.033767	GTCCCTGG
TCCTACCT	943	8mer	0.033762	CGAATTAC
CGCTGGTA	944	8mer	0.03376	GTGCCTAA
GGGCCTAG	945	5mer	0.033727	GCGTG
CTGACGGC	946	8mer	0.033714	ACTGCGAT
GGATTATA	947	8mer	0.033696	GCACCACG
CGACCAAA	948	8mer	0.033673	CCGCCGTT
GAACGTGC	949	8mer	0.033602	CGCAGCGA
TTGGCCAG	950	8mer	0.033572	TGTTACAT
ACACGGGA	951	8mer	0.033562	TGACGGGC
AGCCGTTT	952	8mer	0.033561	AATACGGA
TGCTCACG	953	8mer	0.033552	ATAGGTAC
AGGAGGGG	954	8mer	0.033546	TTTGCAGC
GGAGCGCT	955	8mer	0.033524	AACCCGCG
CCAATGCC	956	8mer	0.03352	CGTATGTG
ATCGCGCC	957	8mer	0.033512	CTGCTGTT
GGGTGTTT	958	8mer	0.033509	AATCACAA
GCCACTGC	959	5mer	0.033471	CTGGG
GGGAGGCT	960	8mer	0.033465	TCGGGACG
TATTGAAG	961	8mer	0.033421	AGCGAAAT
CGAACTCG	962	8mer	0.03342	GCCCTGGC
CGCTAAAT	963	8mer	0.033374	CTCGGTTC
TAATCGAG	964	8mer	0.033372	AACGGACG
AACCCGAA	965	8mer	0.033367	GGTCGCCA
CGCGAACG	966	8mer	0.033338	CGAATCCC
CTGTGGGG	967	8mer	0.033335	GGGCGGCA
CCTCGCGG	968	8mer	0.033317	GTGCGGTA
CGAGATGG	969	8mer	0.033292	GAATGCAC
GCCGGTCT	970	8mer	0.033279	CTGTCGGT
CATGTAAT	971	8mer	0.033276	CGGCCGGG
TCGCGTCC	972	8mer	0.033224	CGTCACCT
TCGATGAG	973	8mer	0.033193	GCCGTGTC
GGAGCACT	974	8mer	0.033192	CATCATTA
AACGCCTG	975	8mer	0.033165	TCGGCCTC
CGGCGCTG	976	8mer	0.033164	GGGATATC
GCGCCGAG	977	8mer	0.033131	CCTATATC
CTCGAGAT	978	8mer	0.033087	CAGCCAAG
CGAAATTA	979	8mer	0.033084	AGGGGCGG
CAATGCCC	980	8mer	0.033055	GCCGGCCG

TCGACCCG
GGTCG
TATACTCG
GGCGCGTG
GGTAGCGT
GGTCGAGG
TTTTTTAG
CCGCCAAT
TGGGGCGG
TACCGGCA
CATGCTAC
CCCGCAAA
AGCCG
AACTGTAC
GCAGAAAA
CACGCAGG
GCATCGTC
GAGGGCGA
GGTGAAAC
CGGATGTC
GCGGAGCG
AAGCCGCA
TGACGCGA
CCACCGAT
TGCGTGAG
TCATTGCC
AGCGAGAC
TGTGTTAC
CCGTAGCT
ACATCCGT
CCTCGTCG
CTGCGCCG
CTATGGGA
TGCAACCT
GATTAGAG
TCGCCGCC
AGGATGTT
CCGCGGAT
ATGAGCTA
ACTCGTCA
GCGGAGGT
ATAACGAG
GAATATCT
GACACGGG
AACATGGT
ATAGTGCG
GACGGGCG
TGCAAGTT
CGCGGATT
GCCCGGAG
ACGTCCGG
TAGATCGT
TGGTA
GCGGAGTT
ACGCCACA
ATGGGACC
CGTCATGA
CATCCGTC

981	5mer	0.033046	ACCCG
982	8mer	0.033036	GGCCGAAT
983	8mer	0.033024	TCGTTGAA
984	8mer	0.033004	GCTCACAC
985	8mer	0.032975	CCCAGTGC
986	5mer	0.032954	GCCGC
987	8mer	0.032933	CCGCCGCC
988	8mer	0.032904	GACCTCCC
989	8mer	0.032903	GCCGATCG
990	8mer	0.032903	CGACGAAG
991	8mer	0.032903	CCGTATAC
992	8mer	0.032903	ACTCGCGC
993	8mer	0.032903	CCGCTATA
994	8mer	0.032903	CGGTGCTA
995	8mer	0.032903	TTTAGCGC
996	8mer	0.032903	CACGAACG
997	8mer	0.032903	CCGTATCG
998	8mer	0.032903	CTTACGCG
999	8mer	0.032903	CGGTAATG
1000	8mer	0.032903	CGGCTAAC

GAGGGCGC
TGATTCGC
TGCGCTCC
GGGCTCAC
GGCGG
GCACGTGG
GGTGTCGT
CGGCTGGG
CTGACGCG
TTAAATGA
TCACCGCG
CTCACATC