

TableS3 The top 1000 features in the optimal feature set of the IncLocP

Order	Feature	F-score	kmer
1	8mer	0.10931	TCCCAAAG
2	8mer	0.107445	BCGGGGCG
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	6mer	0.088041	GGGGCG
15	8mer	0.08771	ATGGGGCG
16	8mer	0.087693	TCGGCCCC
17	8mer	0.087595	AGCGGAGC
18	8mer	0.087322	AATCGACG
19	8mer	0.085901	GCTCACGG
20	6mer	0.084846	GCGGGG
21	8mer	0.084716	GTAGGGCG
22	8mer	0.084641	CGCAGTTC
23	8mer	0.083424	CCGCAGTT
24	8mer	0.082921	BCGGAGCC
25	8mer	0.081369	CATCGGCT
26	8mer	0.079854	AGTAGGAC
27	8mer	0.079681	CCCAAAGT
28	8mer	0.079471	AAGGGCTG
29	8mer	0.078261	GGGATGCG
30	8mer	0.077741	GATCCCGA
31	6mer	0.077289	GGGCGC
32	8mer	0.077287	BCGTGAGC
33	8mer	0.077177	CGGGCGGG
34	8mer	0.077057	BCGGGGGA
35	6mer	0.076967	GGGCCT
36	8mer	0.076646	AGACGGCG
37	8mer	0.075641	CTTGTATC
38	8mer	0.07546	CTTAAACC
39	8mer	0.075416	GAGCCCGA
40	8mer	0.07508	CGTAGCTT
41	8mer	0.074414	CCCCGAA
42	6mer	0.073596	ATCGCG
43	8mer	0.073205	BGCGAGGT
44	8mer	0.073205	CCGCACCG
45	8mer	0.072737	CGCAGCGT
46	8mer	0.072567	CATCGCGG
47	8mer	0.072132	TGGGCCCC
48	6mer	0.072091	CGGGGC
49	8mer	0.072036	CCCGTTTC
50	8mer	0.072034	GTGCGTGA
51	8mer	0.071804	GGATGCGG
52	6mer	0.070845	CGTCCA
53	8mer	0.070518	AAGGGGCG
54	6mer	0.070402	AGCCGG
55	8mer	0.07038	CCGAAGGC

Order	Feature	F-score	kmer
351	8mer	0.047169	GGTCGAGA
352	8mer	0.047111	AAGTCGGG
353	8mer	0.04707	CGACCCGG
354	8mer	0.047052	BCGCCGTG
355	8mer	0.046985	TGCGTCAA
356	8mer	0.04697	TGGACAGC
357	8mer	0.046957	TCGTCAAC
358	8mer	0.0469	CACGGGAG
359	8mer	0.046795	ACGACCCC
360	8mer	0.046768	TATCCAGT
361	8mer	0.046766	GGTCATCG
362	8mer	0.046758	ACGCAGGC
363	8mer	0.046711	CGCGCCGT
364	8mer	0.046694	CTCCCAAA
365	8mer	0.046656	CGGGGCGG
366	8mer	0.046639	TTATGTGG
367	8mer	0.046637	BCGCAGCA
368	6mer	0.046583	CGGGCG
369	8mer	0.046537	GACCGCGA
370	8mer	0.046483	AGAAGCGC
371	6mer	0.04647	ACCCGC
372	6mer	0.046463	CCGAGG
373	8mer	0.046434	ATCGCCGA
374	8mer	0.04641	TGTCCGGT
375	8mer	0.046361	AGCAGTCC
376	8mer	0.046282	AGCTCGGC
377	6mer	0.04625	GTCGTG
378	8mer	0.046234	TCTGGTAT
379	8mer	0.046221	CGGGCACT
380	8mer	0.046201	GTTAATTT
381	8mer	0.046144	CAGGACGA
382	8mer	0.046108	GGTCGGGC
383	6mer	0.046078	CGCCGG
384	8mer	0.04606	CGGGGCCA
385	8mer	0.046033	CGCCCGAG
386	8mer	0.045988	TCGCCCTC
387	8mer	0.045957	GGCGGGGC
388	6mer	0.045846	CGCCCG
389	8mer	0.045842	CCCTGGGC
390	8mer	0.045821	CGCCGGAA
391	8mer	0.045807	ATCTTAAT
392	6mer	0.04578	CCGGCT
393	8mer	0.04574	CCGCGAAC
394	8mer	0.045598	TAAGGGCT
395	8mer	0.045566	CGACCTCC
396	8mer	0.045552	ACGCTCGT
397	8mer	0.045544	CGCGGGCA
398	8mer	0.045541	GCGCCAAT
399	Triplet	0.045475	Triplet
400	8mer	0.045446	GACCCGGC
401	8mer	0.045446	GGCTGATG
402	8mer	0.045399	GCACAATT
403	8mer	0.045291	GGTTGCTA
404	8mer	0.045257	TGTTTACG
405	8mer	0.045235	AGGGCGCA

56	8mer	0.069523	CCTGTGTT
57	8mer	0.069495	GTGGAAGA
58	8mer	0.068895	ATCCCGAA
59	8mer	0.068761	CTTGAACC
60	8mer	0.068296	ACGAGCCC
61	8mer	0.068213	AGACCCTC
62	8mer	0.068069	ACGTGCGG
63	8mer	0.068033	GTGCGACT
64	8mer	0.067987	GACCATCG
65	8mer	0.067781	ACGAATCA
66	8mer	0.067781	TCGTGCGC
67	8mer	0.067781	TAACCGCC
68	8mer	0.067781	GTGACCCG
69	8mer	0.067781	CGAGTCGG
70	8mer	0.067781	TATGCGGG
71	8mer	0.067781	ATGCGGGG
72	8mer	0.067692	AGCCACCG
73	8mer	0.067483	GCGTGCTC
74	8mer	0.067415	TCCGGGCG
75	6mer	0.067398	CGGGGG
76	6mer	0.067315	CCGCCG
77	8mer	0.067239	GGAATCCG
78	8mer	0.067199	TGACGAGC
79	8mer	0.067076	CGCACAGT
80	8mer	0.066653	GCGCGTCC
81	8mer	0.066595	TCCGCCGG
82	8mer	0.066489	GGCCCGTG
83	8mer	0.066467	GCGGATCT
84	8mer	0.06636	GGGCCCT
85	8mer	0.06628	CCTGGGCC
86	8mer	0.066234	GACGAATC
87	8mer	0.066039	AACGTCGG
88	8mer	0.06574	TTATGCGG
89	8mer	0.06574	CGACGAAT
90	8mer	0.065703	GCGGTAAC
91	8mer	0.065606	CCTGGGCG
92	8mer	0.065523	TTGTATCC
93	8mer	0.065484	TGGTCTCG
94	8mer	0.065447	CCACTGCA
95	6mer	0.065422	CGCCCC
96	8mer	0.065364	CGACAGGA
97	8mer	0.065247	CATGACAC
98	8mer	0.065246	ATCCGCCG
99	8mer	0.064857	TGTCGGTA
100	8mer	0.064394	CTCATTGC
101	8mer	0.06426	GTAACCGC
102	8mer	0.064055	CCGTTTGA
103	8mer	0.063797	TGTGCGCC
104	8mer	0.063633	TGCGCCGT
105	8mer	0.063586	GCACCGAA
106	8mer	0.063417	AGGAGGGC
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

406	8mer	0.045227	TAGGCATT
407	8mer	0.045213	CCGAAGTC
408	8mer	0.045208	CGTACATT
409	8mer	0.045197	CGCGTACA
410	8mer	0.045184	CCATCGGT
411	8mer	0.045154	CGAGACCA
412	8mer	0.04515	CCGGACTT
413	8mer	0.045117	ATGTCGGC
414	8mer	0.045071	CTAGCGGA
415	8mer	0.045064	TCGACGAA
416	8mer	0.045057	GGGCGCGC
417	8mer	0.044993	CCCCTTAT
418	6mer	0.044974	TCGGCC
419	8mer	0.044947	GGCAGAAA
420	6mer	0.04493	CCCGAA
421	8mer	0.044928	GTTGCTAT
422	8mer	0.044916	GGTTGGTC
423	8mer	0.044909	ACCGTGCC
424	6mer	0.044886	CCCGCC
425	8mer	0.044842	ACGCCTCG
426	8mer	0.044812	CGTGTTAC
427	8mer	0.044798	ATGCGCGA
428	8mer	0.044713	CTTTCGTC
429	5C-PseDNC	0.04466	5C-PseDNC
430	8mer	0.044572	GGGCGCGC
431	6mer	0.04457	ACGCGA
432	8mer	0.044498	ACCGCGAA
433	8mer	0.044442	GGGGATAT
434	8mer	0.044404	TTTGCTCT
435	8mer	0.044402	AGATTGCG
436	8mer	0.044323	CGAGAATC
437	8mer	0.044257	CAGTTCGG
438	8mer	0.044228	ATCGACGA
439	8mer	0.044215	CACGGCAA
440	8mer	0.044191	ACCCCCGA
441	8mer	0.044183	CGAGAGCT
442	6mer	0.04416	CTGGGC
443	6mer	0.044157	GGCGGG
444	8mer	0.044145	GCGACTCG
445	8mer	0.044099	CCGCTGGT
446	8mer	0.044081	ACTGCTCA
447	8mer	0.044067	GAGCCCCC
448	8mer	0.04403	CCAAAGTC
449	6mer	0.044018	GGGCGA
450	6mer	0.044006	GCGGTA
451	8mer	0.043915	CCGCAGCC
452	8mer	0.043904	CACCGCGC
453	8mer	0.043888	GCCGCCGC
454	8mer	0.043869	CCCCCTGA
455	8mer	0.043817	CAAAGTAC
456	8mer	0.043812	GGAGGGCC
457	8mer	0.043806	GGGGGTCC
458	8mer	0.043794	GCCGGACT
459	8mer	0.043789	GCGCCGGA
460	8mer	0.043774	GCAACGTC
461	8mer	0.043767	ACCTAGAC
462	6mer	0.043755	GCGCCG
463	8mer	0.04375	TTTTTGTA

114	8mer	0.06263	AGGTTCGAG
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	8mer	0.061952	TCACGGTG
121	8mer	0.06195	GATGCGGA
122	8mer	0.061771	GGCCCCCTG
123	8mer	0.061704	GGGGGCGA
124	8mer	0.061609	TCGTCCAA
125	8mer	0.061468	GGGGGCCT
126	8mer	0.061418	CGATGATA
127	8mer	0.06132	GGGCGAGG
128	6mer	0.061176	CCCCCG
129	8mer	0.061146	TCGGTACT
130	8mer	0.061133	CGGTGTCTG
131	8mer	0.061067	GGGGTCGA
132	8mer	0.060888	TCAAGCGA
133	6mer	0.060796	GTCGGA
134	8mer	0.060681	GGTCGTGG
135	8mer	0.06066	GGAGGGGT
136	8mer	0.060619	CCTTCCGC
137	8mer	0.060596	ACTCCCCG
138	8mer	0.060568	CAAGCGAT
139	8mer	0.060416	AGGGGATA
140	8mer	0.060304	CCTCGGCC
141	8mer	0.060137	GCATTCGG
142	6mer	0.060056	GGGCGG
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	8mer	0.059376	GACCGCAA
151	6mer	0.059325	GGGACC
152	8mer	0.059293	TGCGGATC
153	8mer	0.059292	CACGACCC
154	8mer	0.059268	CAACATGG
155	8mer	0.059228	GGGCGCGT
156	8mer	0.0589	CGTAGGGC
157	8mer	0.058785	CATCGGTA
158	8mer	0.058759	GGATCAGC
159	8mer	0.058752	AAACGTCG
160	8mer	0.058738	AGATCCCCG
161	8mer	0.058588	TCCGTAGC
162	8mer	0.05853	GGCGCAGC
163	8mer	0.058336	ACGCGGAG
164	8mer	0.058097	TAGGACAA
165	8mer	0.057951	GCCTTGGG
166	8mer	0.057646	AATCCCCG
167	8mer	0.057561	CGCCGTTA
168	8mer	0.057561	GTAAACCG
169	8mer	0.057514	GGCCTATG
170	8mer	0.057493	CGGACGTG
171	8mer	0.057406	GCCGGAAG

464	8mer	0.043725	AGATACTA
465	8mer	0.043715	AACTCGTC
466	8mer	0.043713	GGAACCTA
467	8mer	0.043673	AAGGTGCC
468	8mer	0.043639	TATTAGGG
469	8mer	0.043625	CCCGAACT
470	8mer	0.043623	TTGTACGG
471	6mer	0.043578	GATGCG
472	8mer	0.043575	CCGTGTCC
473	8mer	0.043522	CGCCTATA
474	8mer	0.043511	TATGTCGG
475	8mer	0.043503	TACTGCCA
476	8mer	0.04347	TATAGCAC
477	8mer	0.043468	GGTTGCAG
478	8mer	0.043459	TTTGCCCA
479	8mer	0.043415	CCGCGATT
480	8mer	0.043357	GGTTGTTC
481	8mer	0.043339	GAGCGGGC
482	8mer	0.04333	GAGGCCGT
483	8mer	0.04332	GTGTCTGA
484	8mer	0.043304	CTGCGAGT
485	8mer	0.043299	CGGCCCTT
486	8mer	0.043284	GCAGTTCCG
487	8mer	0.043258	TCCTGTGT
488	8mer	0.043251	GGGCGGTC
489	8mer	0.04325	CCGTTTCC
490	6mer	0.043239	GCCGGG
491	8mer	0.043201	TCGTCGAA
492	8mer	0.043186	TTGCGCCA
493	8mer	0.043135	AGCATTGC
494	8mer	0.043132	ACGCGTAC
495	8mer	0.043127	CTTCCGCG
496	8mer	0.0431	GGGGCGGT
497	8mer	0.043092	ACCCGCAC
498	8mer	0.043052	GAGCCCGC
499	8mer	0.043045	ATCCCGAC
500	8mer	0.043035	GAACGAAT
501	8mer	0.042965	CAGTACGT
502	6mer	0.042963	GCCGGC
503	8mer	0.042959	CTTAAGTA
504	8mer	0.042959	GAGATACT
505	8mer	0.04295	CGTCGTCT
506	8mer	0.042895	GACGCGTA
507	8mer	0.042872	GCACTCCA
508	8mer	0.042863	CGCGCGGC
509	8mer	0.042848	TGCTTTTT
510	6mer	0.042838	CCGGGC
511	8mer	0.04278	CCGACTA
512	8mer	0.042751	ATTGCGCC
513	8mer	0.042736	CGAGGTCT
514	8mer	0.042727	CAGGCCAC
515	6mer	0.042725	CGTCGT
516	8mer	0.042692	GATCCGGC
517	8mer	0.042663	TCCTCCGT
518	8mer	0.042634	CTAAACAT
519	8mer	0.042609	GGGCGGAC
520	6mer	0.042584	GTTTCG
521	8mer	0.042582	CACAGATC

172	8mer	0.057352	TTCCCTAA
173	8mer	0.057325	CGCGGGAA
174	8mer	0.057279	GGAATATC
175	8mer	0.057262	GGCGCGTC
176	8mer	0.057246	GGTAACCG
177	8mer	0.057244	AATCCGCC
178	8mer	0.057216	ACATCGCG
179	8mer	0.057143	GAGGGCAG
180	8mer	0.057128	TGATCCGG
181	8mer	0.057127	CTGATCCG
182	8mer	0.057105	GCCGCAGT
183	8mer	0.057074	GACGAGCC
184	8mer	0.05707	GTCGGTAC
185	8mer	0.05707	TCGTAGGG
186	8mer	0.056971	CGGGCATC
187	8mer	0.056836	CGCCGTGT
188	6mer	0.056774	GGCGCG
189	8mer	0.056612	TGCGGTAA
190	8mer	0.056504	GAGCCGGG
191	8mer	0.056504	ATAGCTAA
192	8mer	0.056455	CGGGATTC
193	8mer	0.056438	GAGACGCC
194	8mer	0.056382	CGTGCTCG
195	8mer	0.056377	CGGGCCCC
196	8mer	0.056307	CGGAGCCC
197	8mer	0.056099	AGCTACGC
198	8mer	0.056053	ATCGCGGG
199	6mer	0.056047	TCCCGA
200	8mer	0.055956	TCGCGCCG
201	8mer	0.055933	GATTCAAC
202	6mer	0.055895	CGCCGC
203	8mer	0.055575	GGAATCGT
204	8mer	0.055533	AAAAATTA
205	8mer	0.055492	CGGATCTA
206	8mer	0.05548	AGACCCGC
207	8mer	0.055231	ACCTGTAC
208	8mer	0.0552	GCCCTGGT
209	8mer	0.055116	GTCCAATG
210	8mer	0.055039	CGGCGATG
211	8mer	0.054965	TGAGTGTC
212	8mer	0.054878	TTCAAGCG
213	8mer	0.0548	GAGAATCG
214	8mer	0.054507	CAAAACTC
215	8mer	0.054459	AATATACC
216	8mer	0.05445	TATACCAG
217	8mer	0.054389	ACCCGCAG
218	8mer	0.054301	CGGTAACC
219	8mer	0.054267	TAGAGGCG
220	8mer	0.054134	CGGACTTT
221	8mer	0.053955	GCCCCGTG
222	8mer	0.053902	AAAATTAG
223	8mer	0.053867	CGAGCCCC
224	8mer	0.053538	GACCCTCC
225	8mer	0.053492	ATGCAACG
226	8mer	0.053471	TCCAATGC
227	6mer	0.053345	GCGCGG
228	8mer	0.05329	CCTCAGCC
229	8mer	0.053201	CCGCATTC

522	8mer	0.042576	STAAGGGC
523	8mer	0.042567	GGACGCAA
524	8mer	0.042554	CCTGCGCC
525	8mer	0.042547	GGCGGTGT
526	6mer	0.042512	ACCGCG
527	8mer	0.042476	TGCACTTG
528	8mer	0.042462	AGGAGGCC
529	Triplet	0.042416	Triplet
530	8mer	0.04237	CCGAAGGG
531	8mer	0.04234	TGAGCCAC
532	8mer	0.042305	TCGGGGGT
533	8mer	0.042265	CCCGGGAG
534	SC-PseDNC	0.042253	SC-PseDNC
535	8mer	0.042235	CAACGTCT
536	8mer	0.042223	CCGAAGAC
537	8mer	0.042168	TCGAGAGC
538	8mer	0.042149	ACGGCCCA
539	8mer	0.042135	CCGTGGAA
540	8mer	0.042134	AACTGCGG
541	8mer	0.042068	CCAGGCAC
542	8mer	0.042059	AGGTCTGA
543	8mer	0.042056	CCCCGGCA
544	8mer	0.042056	AAGCGAGC
545	8mer	0.04204	TGCTGCGC
546	8mer	0.042039	CGGAGGTT
547	8mer	0.041942	GTTCCCAC
548	8mer	0.041938	AGACGGTC
549	8mer	0.041901	CCATAGTG
550	8mer	0.041879	CGAAGGCC
551	8mer	0.04183	GCCGCCAG
552	8mer	0.04182	AGCTTTTG
553	8mer	0.041809	CCCGCCCC
554	8mer	0.041808	TCGACCTC
555	8mer	0.041726	CCCGGAGC
556	8mer	0.041704	CACCACTT
557	8mer	0.041699	GACCTCGG
558	6mer	0.041662	TCGCCC
559	Triplet	0.04159	Triplet
560	8mer	0.041576	TCCGTTTG
561	8mer	0.041533	CGGAAACC
562	6mer	0.041506	CTCACG
563	8mer	0.041448	ACAATGAC
564	8mer	0.041447	CCCTCGTC
565	8mer	0.041437	AGGTTCAA
566	8mer	0.041436	GCCCCCCG
567	8mer	0.041417	CCGGGCCC
568	8mer	0.041416	GGGAGATC
569	8mer	0.041376	TACTCGGG
570	8mer	0.041342	TACGTAAA
571	8mer	0.041297	TCGCGGGC
572	8mer	0.041287	CCCGTATT
573	8mer	0.041237	GACCTGGC
574	6mer	0.041214	TTCGCC
575	6mer	0.04121	AGGCAC
576	8mer	0.041175	GCCCGAGC
577	8mer	0.041136	GATTCCCG
578	8mer	0.04113	CATTTCAT
579	8mer	0.041106	GGGGCGGC

230	8mer	0.052943	ACGGAGGG
231	8mer	0.052902	GCTGCTCG
232	8mer	0.052845	GCGATGGG
233	8mer	0.052795	CTCGCGTC
234	8mer	0.052789	AATCCCAA
235	8mer	0.052597	GAATATTG
236	6mer	0.052545	CCTGGG
237	8mer	0.052512	GACCTAGC
238	6mer	0.052492	CCGAAG
239	8mer	0.052453	TGGCCGTC
240	8mer	0.052419	CCAGGTCC
241	8mer	0.052418	GGGCTGAC
242	6mer	0.052402	GCCCCG
243	8mer	0.052379	TTTGCGGA
244	8mer	0.052338	CCGCCCCC
245	8mer	0.052314	CTGACGAG
246	8mer	0.052189	GGGGCCTA
247	8mer	0.052127	GCAGAGGG
248	8mer	0.052092	CGTCGGCG
249	8mer	0.051893	CCCGAAGT
250	8mer	0.051883	GCGGTGTC
251	6mer	0.05178	CGGCCC
252	8mer	0.051652	CGTGAGCC
253	8mer	0.05164	GCCGCATT
254	8mer	0.051615	GGGTCGAG
255	8mer	0.051532	TCGAGTCG
256	8mer	0.051301	TCCCAGCT
257	8mer	0.051286	GAGCGCCA
258	8mer	0.051239	ATTGCCTG
259	8mer	0.051176	CAAGATCG
260	8mer	0.051123	GCGTTCAG
261	8mer	0.051068	GCTCCTCG
262	8mer	0.050908	ACGGGCAG
263	8mer	0.050867	AGATCGCG
264	6mer	0.050856	ACGGGA
265	8mer	0.050733	GATATTGA
266	8mer	0.050698	AGCCCGAA
267	8mer	0.050688	TTTTTAGT
268	8mer	0.050647	CTGGGCCC
269	8mer	0.050597	CCGCCAGC
270	8mer	0.050593	CTCCGTAG
271	8mer	0.050576	CGGAGGTC
272	8mer	0.050492	CGTCCAAT
273	8mer	0.050489	TGAATGCA
274	6mer	0.050433	ACCTGG
275	8mer	0.050383	GCCTGGCC
276	8mer	0.050332	GGGCGGGC
277	8mer	0.050237	TGCCGGTC
278	8mer	0.0502	TAATACCG
279	8mer	0.050185	ATTACTGC
280	8mer	0.050111	CCTGAGCA
281	8mer	0.050078	GACTGTAT
282	8mer	0.050071	GGCGTGCT
283	6mer	0.049974	CGAGAC
284	8mer	0.049934	GGAAACGG
285	8mer	0.049918	GCGGTTCC
286	8mer	0.049887	CCTCCGTA
287	8mer	0.049846	GATCGCGC

580	8mer	0.041081	AAGGGTCC
581	8mer	0.041078	ITGCGGAA
582	8mer	0.041011	CTAATGCA
583	8mer	0.040951	CCCGTCTC
584	6mer	0.040875	CGACGA
585	8mer	0.040833	CATCGCCG
586	8mer	0.040822	CCGCCGGC
587	8mer	0.040799	GGGACCGC
588	8mer	0.040795	TAACGGGA
589	8mer	0.040795	CGCGATGA
590	8mer	0.040795	ACGCGCTA
591	8mer	0.040795	CGGCTAGC
592	8mer	0.040795	GTAACGGC
593	8mer	0.040795	GTCGTAAT
594	8mer	0.040786	CTGGGCCT
595	8mer	0.040726	CCTGGCGT
596	6mer	0.040723	GAGGGG
597	8mer	0.040705	TGGGACCT
598	8mer	0.040688	CGTAATAC
599	6mer	0.040649	CTCGGC
600	8mer	0.040643	GGCGTTCA
601	8mer	0.04057	TCACGCCA
602	6mer	0.040565	TCCGCC
603	8mer	0.040562	CGAACACC
604	8mer	0.040556	TGCTTACC
605	8mer	0.040475	AACGGGTT
606	8mer	0.040464	CACGCCTC
607	8mer	0.040462	GGGATCGC
608	8mer	0.040398	CACCCCCG
609	8mer	0.040398	CGAACGGC
610	8mer	0.040395	CGTCCAGC
611	8mer	0.040386	TGGCGTGC
612	8mer	0.040378	TGGGACTG
613	8mer	0.040371	CCCCTGAC
614	8mer	0.040365	GCCGGCTG
615	8mer	0.040326	TCCTACCT
616	8mer	0.040326	CGCTGGTA
617	8mer	0.040314	GGGCCTAG
618	8mer	0.040306	CTGACGGC
619	8mer	0.040293	GGATTATA
620	8mer	0.040258	CGACCAAA
621	8mer	0.040252	GAACGTGC
622	8mer	0.040247	TTGGCCAG
623	8mer	0.04024	ACACGGGA
624	8mer	0.040219	AGCCGTTT
625	8mer	0.040205	TGCTCACG
626	8mer	0.040192	AGGAGGGC
627	8mer	0.040188	GGAGCGCT
628	8mer	0.040173	CCAATGCC
629	8mer	0.040165	ATCGCGCC
630	6mer	0.040115	GCTCAC
631	8mer	0.040044	GGGTGTTT
632	8mer	0.040027	GCCACTGC
633	8mer	0.039989	GGGAGGCT
634	8mer	0.039986	TATTGAAG
635	6mer	0.039929	GCGTTC
636	8mer	0.039906	CGAACTCG
637	8mer	0.039904	CGCTAAAT

288	8mer	0.049818	TTGGTGCT
289	8mer	0.049787	CGGAGGCG
290	8mer	0.049757	CGAGATCG
291	8mer	0.049721	TGAACCCG
292	8mer	0.049702	AATAGCCA
293	8mer	0.049633	AGCGGGGC
294	8mer	0.049554	GGTGAACC
295	8mer	0.049369	ATAAATGG
296	8mer	0.049266	CACTGCAC
297	8mer	0.049199	TGCCTGCG
298	8mer	0.049125	CAAGGAGT
299	8mer	0.049102	CATTGCCG
300	8mer	0.048983	GCCAGAGT
301	8mer	0.0489	GGCGCCCA
302	8mer	0.048871	AGCCGCAT
303	8mer	0.048848	CTCACGGG
304	8mer	0.048843	CTTCCCTA
305	6mer	0.048775	CCCGGG
306	8mer	0.048756	GATCGTCG
307	8mer	0.048723	GGCTATTG
308	6mer	0.048722	CCGGTC
309	8mer	0.04869	GGGACTGA
310	8mer	0.048651	TGAACACG
311	8mer	0.048619	GGCGGGGG
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	6mer	0.048125	CCCGCA
323	8mer	0.048104	GAACGGCC
324	8mer	0.048021	GTGGGAGG
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	6mer	0.047844	TGAGCC
333	8mer	0.047816	ATCGTAGA
334	8mer	0.047784	CTCGTCAA
335	Triplet	0.047765	Triplet
336	8mer	0.047735	AGTCGGAG
337	8mer	0.047729	CCTCTCAT
338	8mer	0.047708	TGCCCCGA
339	8mer	0.047676	ACTTCGAG
340	8mer	0.04766	TGCGGGGG
341	8mer	0.047656	GCCCGAAC
342	8mer	0.047521	TGGGCGCC
343	8mer	0.047505	ACATGGCA
344	Triplet	0.047503	Triplet
345	8mer	0.047497	TGCACGCA

638	8mer	0.039873	TAATCGAG
639	8mer	0.039868	AACCCGAA
640	8mer	0.039852	CGCGAACC
641	8mer	0.039832	CTGTGGGG
642	8mer	0.039812	CCTCGCGG
643	8mer	0.039805	CGAGATGC
644	6mer	0.039771	GCGAGA
645	8mer	0.039756	GCCGGTCT
646	8mer	0.039632	CATGTAAT
647	8mer	0.039601	TCGCGTCC
648	8mer	0.039546	TCGATGAG
649	8mer	0.039545	GGAGCACT
650	8mer	0.039509	AACGCCTG
651	8mer	0.039489	CGGCGCTG
652	8mer	0.039408	GCGCCGAC
653	8mer	0.039396	CTCGAGAT
654	8mer	0.039384	CGAAATTA
655	8mer	0.039362	CAATGCCC
656	8mer	0.039358	TCGACCCG
657	8mer	0.039265	TATACTCG
658	8mer	0.039222	GGCGCGTG
659	8mer	0.039208	GGTAGCGT
660	8mer	0.039192	GGTCGAGG
661	8mer	0.039183	TTTTTTAG
662	8mer	0.039178	CCGCCAAT
663	8mer	0.039146	TGGGCGCG
664	8mer	0.039133	TACCGGCA
665	8mer	0.039092	CATGCTAC
666	8mer	0.03906	CCCGCAAA
667	8mer	0.039038	AACTGTAC
668	8mer	0.039002	GCAGAAAA
669	8mer	0.039	CACGCAGC
670	6mer	0.039	GCCCGC
671	8mer	0.038999	GCATCGTC
672	8mer	0.038984	GAGGGCGA
673	8mer	0.038973	GGTGAAAC
674	8mer	0.038944	CGGATGTC
675	8mer	0.038939	GCGGAGCC
676	8mer	0.03893	AAGCCGCA
677	6mer	0.038902	CGCGGC
678	8mer	0.03885	TGACGCGA
679	8mer	0.038789	CCACCGAT
680	6mer	0.038761	GCGCCC
681	8mer	0.038745	TGCGTGAG
682	8mer	0.038735	TCATTGCC
683	8mer	0.03869	AGCGAGAC
684	8mer	0.038687	TGTGTTAC
685	8mer	0.038675	CCGTAGCT
686	8mer	0.038664	ACATCCGT
687	8mer	0.03865	CCTCGTCG
688	8mer	0.038649	CTGCGCCC
689	8mer	0.038618	CTATGGGA
690	8mer	0.038616	TGCAACCT
691	8mer	0.038607	GATTAGAG
692	8mer	0.0386	TCGCCGCC
693	8mer	0.038583	AGGATGTT
694	8mer	0.038573	CCGCGGAT
695	8mer	0.038571	ATGAGCTA

346	8mer	0.047407	CGCATTCC
347	8mer	0.047393	CACCGGAC
348	6mer	0.047356	GAGGGC
349	8mer	0.047279	CGCTTGAA
350	8mer	0.047215	GAGATCCC

696	8mer	0.038566	ACTCGTCA
697	8mer	0.038562	GCGGAGGT
698	6mer	0.038552	AGGGCG
699	8mer	0.038544	ATAACGAC
700	8mer	0.038543	GAATATCT

TableS4 The top 1000 features in the optimal feature set of the IncLocPr

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	TACGGGA
8	8mer	0.09567	TGCGGTTT
9	8mer	0.09254	GGGGGACC
10	8mer	0.091784	CCACGACC
11	8mer	0.091036	GCCCCGG
12	5mer	0.090816	GCCGG
13	8mer	0.09051	TCAACGTC
14	8mer	0.088855	CGAAGTCG
15	8mer	0.088667	GGGGGGAC
16	6mer	0.088041	GGGGCG
17	8mer	0.08771	ATGGGGCG
18	8mer	0.087693	TCGGCCCC
19	8mer	0.087595	AGCGGAGC
20	8mer	0.087322	AATCGACG
21	8mer	0.085901	GCTCACGG
22	6mer	0.084846	GCGGGG
23	8mer	0.084716	GTAGGGCG
24	8mer	0.084641	CGCAGTTC
25	8mer	0.083424	CCGCAGTT
26	8mer	0.082921	GCGGAGCC
27	8mer	0.081369	CATCGGCT
28	8mer	0.079854	AGTAGGAC
29	8mer	0.079681	CCCAAAGT
30	8mer	0.079471	AAGGGCTG
31	5mer	0.078959	CGCCG
32	8mer	0.078261	GGGATGCG
33	8mer	0.077741	GATCCCGA
34	6mer	0.077289	GGGCGC
35	8mer	0.077287	GCGTGAGC
36	8mer	0.077177	GGGGCGGG
37	8mer	0.077057	GCGGGGGA
38	6mer	0.076967	GGGCCT
39	8mer	0.076646	AGACGGCG
40	8mer	0.075641	CTTGTATC
41	8mer	0.07546	CTTAAACC
42	8mer	0.075416	GAGCCCGA
43	8mer	0.07508	CGTAGCTT
44	5mer	0.074635	CCCGC
45	8mer	0.074414	CCCCCGAA
46	5mer	0.073675	CGCCC
47	6mer	0.073596	ATCGCG
48	8mer	0.073205	GCGGAGGT

Order	Feature	F-score	kmer
351	8mer	0.048104	GAACGGCC
352	8mer	0.048021	GTGGGAGC
353	5mer	0.048021	CCGAG
354	8mer	0.048009	AGGTCCCT
355	8mer	0.047939	GCGCCATC
356	8mer	0.047913	CGAGACTC
357	8mer	0.047904	CCCGAAGC
358	8mer	0.047877	GTTCTGCA
359	8mer	0.047858	CCTGTCGG
360	8mer	0.047854	GCGCGGGC
361	6mer	0.047844	TGAGCC
362	8mer	0.047816	ATCGTAGA
363	8mer	0.047784	CTCGTCAA
364	Triplet	0.047765	Triplet
365	8mer	0.047735	AGTCGGAC
366	8mer	0.047729	CCTCTCAT
367	8mer	0.047708	TGCCCCGA
368	8mer	0.047676	ACTTCGAG
369	8mer	0.04766	GCGGGGCG
370	8mer	0.047656	GCCCCAAC
371	8mer	0.047521	TGGGCGCC
372	8mer	0.047505	ACATGGCA
373	Triplet	0.047503	Triplet
374	8mer	0.047497	TGCACGCA
375	8mer	0.047407	CGCATTCC
376	8mer	0.047393	CACCGGAC
377	6mer	0.047356	GAGGGC
378	8mer	0.047279	CGCTTGAA
379	8mer	0.047215	GAGATCCC
380	8mer	0.047169	GGTGAGAA
381	8mer	0.047111	AAGTCGGC
382	8mer	0.04707	GACCCCGC
383	8mer	0.047052	GCGCCGTC
384	8mer	0.046985	TGCGTCAA
385	8mer	0.04697	TGGACAGC
386	8mer	0.046957	TCGTCAAC
387	8mer	0.0469	CACGGGAG
388	8mer	0.046795	ACGACCCG
389	8mer	0.046768	TATCCAGT
390	8mer	0.046766	GGTCATCG
391	5mer	0.046764	CCGCA
392	8mer	0.046758	ACGCAGGC
393	8mer	0.046711	GCGCCGCT
394	8mer	0.046694	CTCCCAAA
395	8mer	0.046656	GGGGGCGC
396	8mer	0.046639	TTATGTGCG
397	8mer	0.046637	GCGCAGCA
398	6mer	0.046583	GCGGCG

49	8mer	0.073205	CCGCACCG
50	8mer	0.072737	CGCAGCGT
51	8mer	0.072567	CATCGCGG
52	8mer	0.072132	TGGGCCCC
53	6mer	0.072091	CGGGGC
54	8mer	0.072036	CCCGTTTC
55	8mer	0.072034	GTGCGTGA
56	8mer	0.071804	GGATGCGG
57	5mer	0.071377	CGAGA
58	6mer	0.070845	CGTCCA
59	8mer	0.070518	AAGGGGCG
60	6mer	0.070402	AGCCGG
61	8mer	0.07038	CCGAAGGC
62	8mer	0.069523	CCTGTGTT
63	8mer	0.069495	GTGGAAGA
64	8mer	0.068895	ATCCCGAA
65	8mer	0.068761	CTTGAACC
66	8mer	0.068296	ACGAGCCC
67	8mer	0.068213	AGACCCTC
68	8mer	0.068069	ACGTGCGG
69	8mer	0.068033	GTGCGACT
70	8mer	0.067987	GACCATCG
71	8mer	0.067781	ACGAATCA
72	8mer	0.067781	TCGTGCAC
73	8mer	0.067781	TAACCGCC
74	8mer	0.067781	GTGACCCG
75	8mer	0.067781	CGAGTCGG
76	8mer	0.067781	TATGCGGG
77	8mer	0.067781	ATGCGGGG
78	8mer	0.067692	AGCCACCG
79	8mer	0.067483	GCGTGCTC
80	8mer	0.067415	TCCGGGCG
81	6mer	0.067398	CGGGGG
82	5mer	0.067361	GGAGG
83	6mer	0.067315	CCGCCG
84	8mer	0.067239	GGAATCCG
85	8mer	0.067199	TGACGAGC
86	8mer	0.067076	CGCACAGT
87	8mer	0.066653	GCGCGTCC
88	8mer	0.066595	TCCGCCGG
89	8mer	0.066489	GGCCCGTG
90	8mer	0.066467	GCGGATCT
91	8mer	0.06636	GGGCCCCCT
92	8mer	0.06628	CCTGGGCC
93	8mer	0.066234	GACGAATC
94	8mer	0.066039	AACGTCGG
95	5mer	0.065994	CCGCC
96	5mer	0.065986	CGGGG
97	8mer	0.06574	TTATGCGG
98	8mer	0.06574	CGACGAAT
99	8mer	0.065703	GCGGTAAC
100	8mer	0.065606	CCTGGGCG
101	8mer	0.065523	TTGTATCC
102	8mer	0.065484	TGGTCTCG
103	8mer	0.065447	CCACTGCA
104	6mer	0.065422	CGCCCC
105	8mer	0.065364	CGACAGGA
106	8mer	0.065247	CATGACAC

399	8mer	0.046537	GACCGCGA
400	8mer	0.046483	AGAAGCGC
401	6mer	0.04647	ACCCGC
402	5mer	0.046467	AGGGC
403	6mer	0.046463	CCGAGG
404	8mer	0.046434	ATCGCCGA
405	8mer	0.04641	TGTCCGGT
406	8mer	0.046361	AGCAGTCC
407	8mer	0.046282	AGCTCGGC
408	6mer	0.04625	GTCGTG
409	8mer	0.046234	TCTGGTAT
410	8mer	0.046221	CGGGCACT
411	8mer	0.046201	GTTAATTT
412	8mer	0.046144	CAGGACGA
413	8mer	0.046108	GGTCGGGC
414	6mer	0.046078	CGCCGG
415	8mer	0.04606	CGGGGCCA
416	8mer	0.046033	CGCCCGAC
417	8mer	0.045988	TCGCCCTC
418	8mer	0.045957	GCGGGGGC
419	6mer	0.045846	CGCCCG
420	8mer	0.045842	CCCTGGGC
421	8mer	0.045821	CGCCGGAA
422	8mer	0.045807	ATCTTAAT
423	6mer	0.04578	CCGGCT
424	8mer	0.04574	CCGCGAAC
425	8mer	0.045598	TAAGGGCT
426	8mer	0.045566	CGACCTCC
427	8mer	0.045552	ACGCTCGT
428	8mer	0.045544	GCGGGGCA
429	8mer	0.045541	GCGCCAAT
430	Triplet	0.045475	Triplet
431	8mer	0.045446	GACCCGGC
432	8mer	0.045446	GGCTGATC
433	8mer	0.045399	GCACAATT
434	5mer	0.045344	CGGCC
435	8mer	0.045291	GGTTGCTA
436	8mer	0.045257	TGTTTACG
437	8mer	0.045235	AGGGCGCA
438	8mer	0.045227	TAGGCATT
439	8mer	0.045213	CCGAAGTC
440	8mer	0.045208	CGTACATT
441	8mer	0.045197	CGCGTACA
442	8mer	0.045184	CCATCGGT
443	8mer	0.045154	CGAGACCA
444	8mer	0.04515	CCGGACTT
445	8mer	0.045117	ATGTCGGC
446	8mer	0.045071	CTAGCGGA
447	8mer	0.045064	TCGACGAA
448	8mer	0.045057	GCGCGCGC
449	8mer	0.044993	CCCCTTAT
450	6mer	0.044974	TCGGCC
451	8mer	0.044947	GCGAGAAA
452	6mer	0.04493	CCCGAA
453	8mer	0.044928	GTTGCTAT
454	8mer	0.044916	GGTTGGTC
455	8mer	0.044909	ACCGTGCC
456	6mer	0.044886	CCCGCC

107	8mer	0.065246	ATCCGCCG
108	5mer	0.065211	GCGGG
109	8mer	0.064857	TGTCGGTA
110	8mer	0.064394	CTCATTGC
111	8mer	0.06426	GTAACCGC
112	8mer	0.064055	CCGTTGA
113	5mer	0.063847	CCCCG
114	8mer	0.063797	TGTGCGCC
115	8mer	0.063633	TGCGCCGT
116	8mer	0.063586	GCACCGAA
117	8mer	0.063417	AGGAGGGC
118	8mer	0.063183	AACCGCGG
119	8mer	0.063161	CGGTACTT
120	8mer	0.063124	CCCCAGCG
121	8mer	0.062931	ATACCGCC
122	8mer	0.062913	TCCAGGCG
123	8mer	0.062893	TCCCGCAC
124	8mer	0.062656	ACCGGACT
125	8mer	0.06263	AGGTTCGAG
126	8mer	0.062472	GAAACCCG
127	8mer	0.062448	TTGCCGGT
128	8mer	0.06233	TCAAGCAC
129	8mer	0.062312	CCAAGATC
130	8mer	0.062211	ATTGCCGG
131	5mer	0.062101	CGGGC
132	8mer	0.061952	TCACGGTG
133	8mer	0.06195	GATGCGGA
134	5mer	0.061919	GCGCC
135	8mer	0.061771	GGCCCTG
136	8mer	0.061704	GGGGCGCA
137	8mer	0.061609	TCGTCCAA
138	8mer	0.061468	CGGGGCCT
139	8mer	0.061418	CGATGATA
140	8mer	0.06132	GGGCGAGG
141	6mer	0.061176	CCCCCG
142	8mer	0.061146	TCGGTACT
143	8mer	0.061133	CGGTGTCTG
144	8mer	0.061067	GGGGTCTGA
145	5mer	0.061032	GCCCG
146	8mer	0.060888	TCAAGCGA
147	6mer	0.060796	GTCGGA
148	8mer	0.060681	GGTCGTGG
149	8mer	0.06066	GGAGGGGT
150	8mer	0.060619	CCTTCCGC
151	8mer	0.060596	ACTCCCCG
152	8mer	0.060568	CAAGCGAT
153	8mer	0.060416	AGGGGATA
154	8mer	0.060304	CCTCGGCC
155	8mer	0.060137	GCATTCGG
156	6mer	0.060056	GGGCGG
157	8mer	0.060047	GGGCGCCC
158	8mer	0.059996	TTGGGCGC
159	8mer	0.059859	ATGCGGAT
160	8mer	0.059776	GGGGCCTT
161	8mer	0.059758	GCTTCGCC
162	8mer	0.059737	GTGTTACT
163	8mer	0.059509	CCAACATG
164	5mer	0.059434	GGCCG

457	8mer	0.044842	ACGCCTCG
458	8mer	0.044812	CGTGTTAC
459	8mer	0.044798	ATGCGCGA
460	8mer	0.044713	CTTTCGTC
461	5C-PseDNC	0.04466	5C-PseDNC
462	8mer	0.044572	GGGGCGCC
463	6mer	0.04457	ACGCGA
464	8mer	0.044498	ACCGCGAA
465	8mer	0.044442	GGGGATAT
466	8mer	0.044404	TTTGTCTC
467	8mer	0.044402	AGATTGCG
468	8mer	0.044323	CGAGAATC
469	5mer	0.044304	GGCGT
470	8mer	0.044257	CAGTTCGG
471	8mer	0.044228	ATCGACGA
472	8mer	0.044215	CACGGCAA
473	5mer	0.044197	TCGCG
474	8mer	0.044191	ACCCCCGA
475	8mer	0.044183	CGAGAGCT
476	6mer	0.04416	CTGGGC
477	6mer	0.044157	GGCGGG
478	8mer	0.044145	GCGACTCG
479	8mer	0.044099	CCGCTGGT
480	8mer	0.044081	ACTGCTCA
481	8mer	0.044067	GAGCCCCG
482	8mer	0.04403	CCAAAGTC
483	6mer	0.044018	GGGCGA
484	6mer	0.044006	GCGGTA
485	8mer	0.043915	CCGCAGCC
486	8mer	0.043904	CACCGCGC
487	8mer	0.043888	GCCGCCGC
488	8mer	0.043869	GCCCCTGA
489	8mer	0.043817	CAAAGTAC
490	8mer	0.043812	GGAGGGCC
491	8mer	0.043806	GGGGGTCC
492	8mer	0.043794	GCCGGACT
493	8mer	0.043789	GCGCCGGA
494	8mer	0.043774	GCAACGTC
495	8mer	0.043767	ACCTAGAC
496	6mer	0.043755	GCGCCG
497	8mer	0.04375	TTTTTGTA
498	8mer	0.043725	AGATACTA
499	8mer	0.043715	AACTCGTC
500	8mer	0.043713	GGAACCTA
501	8mer	0.043673	AAGGTGCC
502	8mer	0.043639	TATTAGGG
503	8mer	0.043625	CCCGAACT
504	8mer	0.043623	TTGTACGG
505	6mer	0.043578	GATGCG
506	8mer	0.043575	CCGTGTCC
507	8mer	0.043522	CGCCTATA
508	8mer	0.043511	TATGTCGG
509	8mer	0.043503	TACTGCCA
510	8mer	0.04347	TATAGCAC
511	8mer	0.043468	GGTTGCAG
512	8mer	0.043459	TTTGCCCA
513	8mer	0.043415	CCGCGATT
514	8mer	0.043357	GGTTGTTC

165	8mer	0.059376	GACCGCAA
166	6mer	0.059325	GGGACC
167	8mer	0.059293	TGCGGATC
168	8mer	0.059292	CACGACCC
169	5mer	0.059282	GTCGG
170	8mer	0.059268	CAACATGG
171	8mer	0.059228	GGGCGCGT
172	8mer	0.0589	CGTAGGGC
173	8mer	0.058785	CATCGGTA
174	8mer	0.058759	GGATCAGC
175	8mer	0.058752	AAACGTCG
176	8mer	0.058738	AGATCCCG
177	8mer	0.058588	TCCGTAGC
178	8mer	0.05853	GGCGCAGC
179	8mer	0.058336	ACGCGGAG
180	8mer	0.058097	TAGGACAA
181	8mer	0.057951	GCCTTGGG
182	5mer	0.057679	GGCGC
183	8mer	0.057646	AATCCCCG
184	8mer	0.057561	CGCCGTTA
185	8mer	0.057561	GTAAACCG
186	8mer	0.057514	GGCCTATG
187	8mer	0.057493	CGGACGTG
188	8mer	0.057406	GCCGGAAG
189	8mer	0.057352	TTCCCTAA
190	8mer	0.057325	CGCGGGAA
191	8mer	0.057279	GGAATATC
192	8mer	0.057262	GGCGCGTC
193	8mer	0.057246	GGTAACCG
194	8mer	0.057244	AATCCGCC
195	8mer	0.057216	ACATCGCG
196	8mer	0.057143	GAGGGCAG
197	8mer	0.057128	TGATCCGG
198	8mer	0.057127	CTGATCCG
199	8mer	0.057105	GCCGCAGT
200	8mer	0.057074	GACGAGCC
201	8mer	0.05707	GTCGGTAC
202	8mer	0.05707	TCGTAGGG
203	5mer	0.057046	GGGCC
204	8mer	0.056971	CGGGCATC
205	5mer	0.056908	CCCGA
206	8mer	0.056836	CGCCGTGT
207	6mer	0.056774	GGCGCG
208	8mer	0.056612	TGCGGTAA
209	8mer	0.056504	GGAGCCGG
210	8mer	0.056504	ATAGCTAA
211	8mer	0.056455	CGGGATTG
212	8mer	0.056438	GAGACGCC
213	8mer	0.056382	CGTGCTCG
214	8mer	0.056377	CGGGCCCC
215	8mer	0.056307	CGGAGCCC
216	8mer	0.056099	AGCTACGC
217	8mer	0.056053	ATCGCGGG
218	6mer	0.056047	TCCCGA
219	8mer	0.055956	TCGCGCCG
220	8mer	0.055933	GATTCAAC
221	6mer	0.055895	CGCCGC
222	5mer	0.055828	CGGAG

515	8mer	0.043339	GAGCGGGC
516	8mer	0.04333	GAGGCCGT
517	8mer	0.04332	GTGTCGTA
518	8mer	0.043304	CTGCGAGT
519	8mer	0.043299	CGGCCCCT
520	8mer	0.043284	GCAGTTCC
521	8mer	0.043258	TCCTGTGT
522	8mer	0.043251	GGGCGGTC
523	8mer	0.04325	CCGTTTCC
524	6mer	0.043239	GCCGGG
525	8mer	0.043201	TCGTCGAA
526	8mer	0.043186	TTGCGCCA
527	8mer	0.043135	AGCATTGC
528	8mer	0.043132	ACGCGTAC
529	8mer	0.043127	CTTCCGCG
530	5mer	0.043117	CCCGG
531	8mer	0.0431	GGGGCGGT
532	8mer	0.043092	ACCCGCAC
533	8mer	0.043052	GAGCCCGC
534	8mer	0.043045	ATCCCGAC
535	8mer	0.043035	GAACGAAT
536	8mer	0.042965	CAGTACGT
537	6mer	0.042963	GCCGGC
538	8mer	0.042959	CTTAAGTA
539	8mer	0.042959	GAGATACT
540	8mer	0.04295	CGTCGTCT
541	8mer	0.042895	GACGCGTA
542	8mer	0.042872	GCACTCCA
543	8mer	0.042863	CGCGCGGC
544	8mer	0.042848	TGCTTTTT
545	6mer	0.042838	CCGGGC
546	8mer	0.04278	CCGGACTA
547	8mer	0.042751	ATTGCGCC
548	8mer	0.042736	CGAGGTCT
549	8mer	0.042727	CAGGCCAC
550	6mer	0.042725	CGTCGT
551	8mer	0.042692	GATCCGGC
552	8mer	0.042663	TCCTCCGT
553	8mer	0.042634	CTAAACAT
554	5mer	0.042622	ACGGG
555	8mer	0.042609	CGGGCGAC
556	6mer	0.042584	GTTTCG
557	8mer	0.042582	CACAGATC
558	8mer	0.042576	GTAAGGGC
559	8mer	0.042567	GGACGCAA
560	8mer	0.042554	CCTGCGCC
561	8mer	0.042547	GGCGGTGT
562	6mer	0.042512	ACCGCG
563	8mer	0.042476	TGCACTTG
564	8mer	0.042462	AGGAGGCC
565	Triplet	0.042416	Triplet
566	8mer	0.04237	CCGAAGGC
567	8mer	0.04234	TGAGCCAC
568	8mer	0.042305	TCGGGGGT
569	8mer	0.042265	CCCGGGAC
570	5C-PseDNC	0.042253	5C-PseDNC
571	8mer	0.042235	CAACGTCT
572	5mer	0.042224	CGAGG

223	8mer	0.055575	GGAATCGT
224	8mer	0.055533	AAAAATTA
225	8mer	0.055492	CGGATCTA
226	8mer	0.05548	AGACCCGC
227	8mer	0.055231	ACCTGTAC
228	8mer	0.0552	GCCCTGGT
229	8mer	0.055116	GTCCAATG
230	8mer	0.055039	CGGCGATG
231	8mer	0.054965	TGAGTGTC
232	8mer	0.054878	TTCAAGCG
233	8mer	0.0548	GAGAATCG
234	5mer	0.054796	CCTGG
235	8mer	0.054507	CAAAACTC
236	8mer	0.054459	AATATACC
237	8mer	0.05445	TATACCAG
238	8mer	0.054389	ACCCGCAG
239	5mer	0.054319	CCGGC
240	8mer	0.054301	CGGTAACC
241	8mer	0.054267	TAGAGGCG
242	8mer	0.054134	CGGACTTT
243	5mer	0.05409	CCGAA
244	8mer	0.053955	GCCCCGTG
245	8mer	0.053902	AAAATTAG
246	8mer	0.053867	CGAGCCCG
247	8mer	0.053538	GACCCTCC
248	8mer	0.053492	ATGCAACG
249	8mer	0.053471	TCCAATGC
250	6mer	0.053345	GCGCGG
251	8mer	0.05329	CCTCAGCC
252	8mer	0.053201	CCGCATTC
253	5mer	0.053022	GGGGC
254	8mer	0.052943	ACGGAGGG
255	8mer	0.052902	GCTGCTCG
256	8mer	0.052845	GCGATGGG
257	8mer	0.052795	CTCGCGTC
258	8mer	0.052789	AATCCCAA
259	8mer	0.052597	GAATATTG
260	6mer	0.052545	CCTGGG
261	8mer	0.052512	GACCTAGC
262	6mer	0.052492	CCGAAG
263	8mer	0.052453	TGGCCGTC
264	8mer	0.052419	CCAGGTCC
265	8mer	0.052418	GGGCTGAC
266	6mer	0.052402	GCCCCG
267	8mer	0.052379	TTGCGGA
268	8mer	0.052338	CCGCCCCC
269	8mer	0.052314	CTGACGAG
270	5mer	0.052297	CGCGG
271	8mer	0.052189	GGGGCCTA
272	8mer	0.052127	GAGAGGG
273	8mer	0.052092	CGTCGGCG
274	8mer	0.051893	CCCGAAGT
275	8mer	0.051883	GCGGTGTC
276	6mer	0.05178	CGGCCC
277	8mer	0.051652	CGTGAGCC
278	8mer	0.05164	GCCGCATT
279	8mer	0.051615	GGGTGAG
280	8mer	0.051532	TCGAGTCG

573	8mer	0.042223	CCGAAGAC
574	5mer	0.042205	CTCGG
575	8mer	0.042168	TCGAGAGC
576	8mer	0.042149	ACGGCCCA
577	8mer	0.042135	CCGTGGAA
578	8mer	0.042134	AACTGCGC
579	8mer	0.042068	CCAGGCAC
580	8mer	0.042059	AGGTCTGA
581	8mer	0.042056	GCCCGGCA
582	8mer	0.042056	AAGCGAGC
583	8mer	0.04204	TGCTGCGC
584	8mer	0.042039	CGGAGGTT
585	8mer	0.041942	GTTCCAC
586	8mer	0.041938	AGACGGTC
587	8mer	0.041901	CCATAGTG
588	8mer	0.041879	CGAAGGCC
589	8mer	0.04183	GCCGCCAC
590	8mer	0.04182	AGCTTTTG
591	8mer	0.041809	CCCGCCCC
592	8mer	0.041808	TCGACCTC
593	8mer	0.041726	CCCGGAGC
594	8mer	0.041704	CACCACTT
595	8mer	0.041699	GACCTCGC
596	6mer	0.041662	TCGCCC
597	Triplet	0.04159	Triplet
598	8mer	0.041576	TCCGGTTG
599	8mer	0.041533	CGGAAACC
600	6mer	0.041506	CTCACG
601	8mer	0.041448	ACAATGAG
602	8mer	0.041447	CCCTCGTC
603	8mer	0.041437	AGGTTCAA
604	8mer	0.041436	GCCCCCCC
605	8mer	0.041417	CCGGGCCC
606	8mer	0.041416	GGGAGATC
607	8mer	0.041376	TACTCGGG
608	8mer	0.041342	TACGTAAA
609	8mer	0.041297	TCGCGGGC
610	8mer	0.041287	CCCGTATT
611	8mer	0.041237	GACCTGGC
612	6mer	0.041214	TTCGCC
613	6mer	0.04121	AGGCAC
614	8mer	0.041175	GCCCGAGC
615	8mer	0.041136	GCAATCCG
616	8mer	0.04113	CATTTCAT
617	8mer	0.041106	GGGGCGGC
618	8mer	0.041081	AAGGGTCC
619	8mer	0.041078	TTGCGGAA
620	5mer	0.041038	GCGCG
621	8mer	0.041011	CTAATGCA
622	8mer	0.040951	CCCGTCTC
623	5mer	0.040906	GCCCC
624	6mer	0.040875	CGACGA
625	8mer	0.040833	CATCGCCC
626	8mer	0.040822	CCGCCGGC
627	8mer	0.040799	GGGACCGC
628	8mer	0.040795	TACGGGA
629	8mer	0.040795	CGCGATGA
630	8mer	0.040795	ACGCGCTA

281	8mer	0.051301	TCCCAGCT
282	8mer	0.051286	GAGCGCCA
283	8mer	0.051239	ATTGCCTG
284	8mer	0.051176	CAAGATCG
285	8mer	0.051123	GCGTTCAG
286	8mer	0.051068	GCTCCTCG
287	5mer	0.050999	GCACC
288	8mer	0.050908	ACGGGCAG
289	8mer	0.050867	AGATCGCG
290	6mer	0.050856	ACGGGA
291	8mer	0.050733	GATATTGA
292	8mer	0.050698	AGCCCGAA
293	8mer	0.050688	TTTTTAGT
294	8mer	0.050647	CTGGGCCC
295	8mer	0.050597	CCGCCAGC
296	8mer	0.050593	CTCCGTAG
297	8mer	0.050576	CGGAGGTC
298	8mer	0.050492	CGTCCAAT
299	8mer	0.050489	TGAATGCA
300	6mer	0.050433	ACCTGG
301	8mer	0.050383	GCCTGGCC
302	8mer	0.050332	GCGCGGGC
303	8mer	0.050237	TGCCGGTC
304	8mer	0.0502	TAATACCG
305	8mer	0.050185	ATTACTGC
306	8mer	0.050111	CCTGAGCA
307	8mer	0.050078	GA CTGTAT
308	8mer	0.050071	GGCGTGCT
309	6mer	0.049974	CGAGAC
310	8mer	0.049934	GGAACGG
311	8mer	0.049918	GCGGTTCC
312	5mer	0.049887	TCGGC
313	8mer	0.049887	CCTCCGTA
314	8mer	0.049846	GATCGCGC
315	8mer	0.049818	TTGGTGCT
316	8mer	0.049787	CGGAGGCG
317	8mer	0.049757	CGAGATCG
318	8mer	0.049721	TGAACCCG
319	8mer	0.049702	AATAGCCA
320	8mer	0.049633	AGCGGGGC
321	8mer	0.049554	GGTGAACC
322	8mer	0.049369	ATAAATGG
323	8mer	0.049266	CACTGCAC
324	8mer	0.049199	TGCCTGCG
325	8mer	0.049125	CAAGGAGT
326	8mer	0.049102	CATTGCCG
327	5mer	0.049087	CCGGG
328	8mer	0.048983	GCCAGAGT
329	8mer	0.0489	GCGCCCA
330	8mer	0.048871	AGCCGCAT
331	8mer	0.048848	CTCACGGG
332	8mer	0.048843	CTTCCCTA
333	6mer	0.048775	CCCGGG
334	8mer	0.048756	GATCGTCG
335	8mer	0.048723	GGCTATTG
336	6mer	0.048722	CCGGTC
337	8mer	0.04869	GGGACTGA
338	8mer	0.048651	TGAACACG

631	8mer	0.040795	GCGCTAGC
632	8mer	0.040795	GTAACGGC
633	8mer	0.040795	GTCGTAAT
634	8mer	0.040786	CTGGGCCCT
635	8mer	0.040726	CCTGGCGT
636	6mer	0.040723	GAGGGG
637	8mer	0.040705	TGGGACCT
638	8mer	0.040688	CGTAATAC
639	6mer	0.040649	CTCGGC
640	8mer	0.040643	GGCGTTCA
641	8mer	0.04057	TCACGCCA
642	6mer	0.040565	TCCGCC
643	8mer	0.040562	GGAACACC
644	8mer	0.040556	TGCTTACC
645	8mer	0.040475	AACGGGTT
646	8mer	0.040464	CACGCCTC
647	8mer	0.040462	GGGATCGC
648	8mer	0.040398	CACCCCCG
649	8mer	0.040398	GGAACGGC
650	8mer	0.040395	CGTCCAGC
651	8mer	0.040386	TGGCGTGC
652	8mer	0.040378	TGGGACTG
653	8mer	0.040371	CCCCTGAC
654	8mer	0.040365	GCCGGCTG
655	8mer	0.040326	TCCTACCT
656	8mer	0.040326	CGCTGGTA
657	8mer	0.040314	GGGCCTAG
658	8mer	0.040306	CTGACGGC
659	8mer	0.040293	GGATTATA
660	8mer	0.040258	CGACCAAA
661	8mer	0.040252	GAACGTGC
662	8mer	0.040247	TGGCCAG
663	8mer	0.04024	ACACGGGA
664	8mer	0.040219	AGCCGTTT
665	8mer	0.040205	TGCTCACG
666	8mer	0.040192	AGGAGGGC
667	8mer	0.040188	GGAGCGCT
668	8mer	0.040173	CCAATGCC
669	8mer	0.040165	ATCGCGCC
670	6mer	0.040115	GCTCAC
671	8mer	0.040044	GGGTGTTT
672	8mer	0.040027	GCCACTGC
673	8mer	0.039989	GGGAGGCT
674	8mer	0.039986	TATTGAAG
675	6mer	0.039929	GCGTTC
676	8mer	0.039906	GGAACCTC
677	8mer	0.039904	CGCTAAAT
678	8mer	0.039873	TAATCGAG
679	8mer	0.039868	AACCCGAA
680	8mer	0.039852	GCGAACC
681	8mer	0.039832	CTGTGGGC
682	8mer	0.039812	CCTCGCGC
683	8mer	0.039805	CGAGATGC
684	6mer	0.039771	GCGAGA
685	8mer	0.039756	GCCGGTCT
686	8mer	0.039632	CATGTAAT
687	8mer	0.039601	TCGCGTCC
688	8mer	0.039546	TCGATGAG

339	8mer	0.048619	GGCGGGG
340	8mer	0.048591	ACTGTGGG
341	8mer	0.048586	GACCCGCA
342	8mer	0.048551	AGACCTGG
343	8mer	0.048515	GACTCGCG
344	8mer	0.048481	CGCCGCCG
345	8mer	0.048465	CGGGAGCG
346	8mer	0.048459	CGCCCGGC
347	8mer	0.048409	GGCGCGGC
348	8mer	0.048382	ATCGTCGT
349	8mer	0.048319	GTTGTACG
350	6mer	0.048125	CCCGCA

689	8mer	0.039545	GGAGCACT
690	8mer	0.039509	AACGCCTG
691	8mer	0.039489	CGGCGCTG
692	8mer	0.039408	GCGCCGAG
693	8mer	0.039396	CTCGAGAT
694	8mer	0.039384	CGAAATTA
695	8mer	0.039362	CAATGCCC
696	8mer	0.039358	TCGACCCG
697	5mer	0.039357	GGTCG
698	8mer	0.039265	TATACTCG
699	8mer	0.039222	GGCGCGTG
700	8mer	0.039208	GGTAGCGT

redtp68 according to F-score order

Order	Feature	F-score	kmer
701	8mer	0.038531	GACACGGG
702	6mer	0.038521	CCCCGC
703	8mer	0.038508	AACATGGT
704	8mer	0.038481	ATAGTGCG
705	6mer	0.038469	GCGGAG
706	8mer	0.038451	GACGGGCG
707	8mer	0.038446	TGCAAGTT
708	8mer	0.038361	CGCGGATT
709	8mer	0.03836	GCCCGGAG
710	8mer	0.038343	ACGTCCGG
711	8mer	0.038332	TAGATCGT
712	8mer	0.038297	GCGGAGTT
713	8mer	0.038275	ACGCCACA
714	8mer	0.03827	ATGGGACC
715	8mer	0.038252	CGTCATGA
716	8mer	0.038251	CATCCGTC
717	8mer	0.038249	GAGGGCGC
718	8mer	0.038244	TGATTCCG
719	8mer	0.038228	TGCGCTCC
720	8mer	0.038226	GGGCTCAC
721	8mer	0.038193	GCACGTGG
722	8mer	0.038168	GGTGTCTG
723	8mer	0.038135	CGGCTGGG
724	8mer	0.038124	CTGACGCG
725	8mer	0.03805	TTAAATGA
726	8mer	0.038048	TCACCGCG
727	8mer	0.038015	CTCACATC
728	8mer	0.037947	TGCGCCGA
729	8mer	0.037943	GCAAACGT
730	8mer	0.037942	CCGGGAAC
731	8mer	0.037932	TGGCCAGG
732	8mer	0.037889	CGAATGAT
733	8mer	0.037866	ACGCGACA
734	8mer	0.037866	GACGCGAC
735	8mer	0.037866	TGTCGACC
736	8mer	0.037866	TAAACCGC
737	8mer	0.037866	GCTCGCGT
738	8mer	0.037861	CGGCCTCC
739	8mer	0.037833	GACGCCTC
740	8mer	0.037833	TTCAATGA
741	8mer	0.037822	CTCGTCGA
742	6mer	0.037789	TCGCGC
743	8mer	0.037764	GGCCGTCG
744	8mer	0.037759	TGGGGTAT
745	8mer	0.037755	TGTCCCGC
746	8mer	0.03775	CCTTCACT
747	8mer	0.037746	ATATGTGA
748	8mer	0.037736	CCGAACTC
749	8mer	0.037682	TGGTTGCT
750	8mer	0.037677	CTGCTCAC
751	8mer	0.037644	CAGGCGGG
752	8mer	0.03761	GCCCGTCG
753	8mer	0.037582	GCGGGCAT
754	8mer	0.037572	CACGAGAC
755	8mer	0.037572	GCAATGCC

756	8mer	0.037561	CTGCTGCG
757	8mer	0.037556	CCGAGGGC
758	8mer	0.037538	AGGTCGTG
759	8mer	0.037517	GGCTGGTC
760	8mer	0.037512	GAAAGCGA
761	8mer	0.037508	GGGGCTGA
762	8mer	0.037503	GGCTGCTA
763	8mer	0.037495	CTTCATCG
764	8mer	0.03748	CGTGGGTG
765	8mer	0.03746	ATTAGCCG
766	8mer	0.037429	GCTGCCGG
767	8mer	0.037404	AAGCCGGC
768	8mer	0.037394	GGGGGCGG
769	8mer	0.037359	CCCGTCTG
770	8mer	0.037355	CACGGGAT
771	8mer	0.037317	ACCCGAAT
772	8mer	0.037314	TCGCCCCG
773	8mer	0.037312	GCCTATGG
774	8mer	0.03727	AATACCGG
775	8mer	0.037266	CCCGCCC
776	8mer	0.037257	CGGCCTAT
777	8mer	0.037252	GCTGCGCA
778	8mer	0.037222	CCCGGCAC
779	8mer	0.037219	ACCCGGGG
780	8mer	0.037189	AAGCGAAA
781	6mer	0.037183	CCCGTC
782	8mer	0.037161	GAATCGCT
783	8mer	0.037139	GCGCCCGG
784	8mer	0.037131	ACCCGGCC
785	8mer	0.037086	ATTAGGGA
786	8mer	0.037082	CCGCTTAG
787	6mer	0.037077	GGCCCG
788	8mer	0.037036	GGCATTCC
789	8mer	0.037023	GGCATTCC
790	8mer	0.036989	TAAAATGG
791	8mer	0.036988	TCCTTTCC
792	6mer	0.036974	GACCTC
793	8mer	0.036971	CCTCCCAA
794	8mer	0.036962	CGGAATAT
795	8mer	0.036961	TTGGGCCC
796	8mer	0.036937	CGCCAATC
797	8mer	0.036924	TGACCTTA
798	8mer	0.036905	GCGAAATT
799	8mer	0.036883	CTTCAAGC
800	6mer	0.036873	ACACGG
801	8mer	0.036856	AGGCTGGT
802	8mer	0.036827	GCGCGCGG
803	8mer	0.036785	GTGCCGAA
804	8mer	0.036765	TTTCCCGA
805	8mer	0.036755	GAATCCGG
806	8mer	0.036745	CCGACTTT
807	6mer	0.036732	GAGCCC
808	8mer	0.036719	GTAGTGGG
809	6mer	0.036697	GGCCCC
810	8mer	0.036678	CAAAAAAA
811	8mer	0.036676	GCGACATC
812	8mer	0.036655	ACGGGTTC
813	8mer	0.036648	TTCAAGGT

814	8mer	0.036629	TGTACGTT
815	8mer	0.036619	GTCGACCC
816	8mer	0.036608	GCGCCACT
817	8mer	0.036604	CTGGTATA
818	8mer	0.036599	CGTCAAGG
819	8mer	0.036577	GACTGCGA
820	8mer	0.036576	CAGTAGGA
821	8mer	0.036575	CGGCCAC
822	8mer	0.036574	GAAGTCGG
823	8mer	0.036528	ATGCACGC
824	8mer	0.036518	TAACGAGG
825	8mer	0.036479	TCGGGGG
826	8mer	0.036471	CGGGTGGC
827	8mer	0.036463	CAATATAC
828	8mer	0.036425	TCGGCTGT
829	8mer	0.03639	CTATGTTG
830	6mer	0.036385	TCGCGT
831	8mer	0.036376	ACGCTCAT
832	8mer	0.036373	GCGTTCTC
833	6mer	0.036294	TGTCGG
834	8mer	0.03628	GACGGCCT
835	8mer	0.036274	GAGGTCAT
836	8mer	0.036257	ACGGCCTG
837	8mer	0.036192	GTTATCTG
838	8mer	0.03619	CTACAGTA
839	8mer	0.036189	TATAAGGT
840	8mer	0.036163	CGACTTTG
841	8mer	0.036153	TGAGCCGA
842	8mer	0.036139	CGTAGATA
843	8mer	0.036137	CTAAACCG
844	6mer	0.036132	TCGACG
845	8mer	0.036098	CTTGCGTC
846	8mer	0.036073	AATTATCC
847	8mer	0.036068	GAGCGACG
848	8mer	0.036049	GTCCAGAC
849	8mer	0.036048	CGGGGGCC
850	8mer	0.036041	ACGCCCCGA
851	8mer	0.036032	GGCGCCCG
852	8mer	0.036022	GTTTCGCC
853	8mer	0.036005	TCCTGGGC
854	8mer	0.036004	TTGTTAAT
855	8mer	0.03599	CAGCCTCC
856	8mer	0.035982	GGTGGCGC
857	8mer	0.035958	CGTTCAGG
858	6mer	0.035953	CCAACG
859	8mer	0.035936	CGAGTTCT
860	8mer	0.03593	CGCGCGGG
861	8mer	0.035852	ACCCGCGC
862	6mer	0.035848	CAAGCG
863	8mer	0.035833	ACCTCGAC
864	8mer	0.035761	CGAACAGA
865	8mer	0.035756	CTGTCCCG
866	8mer	0.035747	CTCCTGTG
867	8mer	0.035746	GGCTCACG
868	8mer	0.035744	GACCTGGG
869	8mer	0.035735	GAACCTCGT
870	8mer	0.035724	ACTGCACG
871	8mer	0.035716	GAGGGCCG

872	8mer	0.03568	ACCATTAA
873	6mer	0.03567	CGGAGC
874	8mer	0.035668	CAAGCGAA
875	8mer	0.035666	GCGGCCAA
876	8mer	0.035652	GCCATCTA
877	8mer	0.035597	GTCTAGGC
878	8mer	0.035581	ATCATTAT
879	8mer	0.035577	TGGGCCCC
880	8mer	0.035562	TAACCGCG
881	8mer	0.035558	GGCTGGGA
882	8mer	0.035557	GTAAACAT
883	6mer	0.035553	CCGGGG
884	8mer	0.035551	CGTCCAGG
885	6mer	0.035534	GCACGA
886	8mer	0.035506	CGGACTAG
887	6mer	0.035485	GCCCTA
888	8mer	0.035482	GGACCTAG
889	8mer	0.035439	TAGTGTCC
890	8mer	0.035432	CACTACAG
891	8mer	0.035424	CGCTTCTT
892	8mer	0.035384	TCCCAACG
893	8mer	0.035383	CCGGGGTG
894	8mer	0.035364	CGGCGCGC
895	8mer	0.03536	TGGGATTA
896	8mer	0.035351	AGCCGGGG
897	8mer	0.035334	TTCACTAG
898	8mer	0.03528	CCCTTCGA
899	6mer	0.035237	GGCCGG
900	8mer	0.03522	CTGCGCCT
901	6mer	0.035216	CCCAGG
902	8mer	0.03518	CTCGGCTG
903	8mer	0.035126	CACGCCAC
904	8mer	0.035113	CTCGGGTC
905	8mer	0.035093	CGGCGTGC
906	6mer	0.035071	CCGGCA
907	8mer	0.035069	CTGTGAC
908	8mer	0.035065	TAGTGGGA
909	8mer	0.035051	CTCAGCCT
910	8mer	0.03505	GGCCGCTC
911	8mer	0.035019	GGCCAGGT
912	8mer	0.035018	TTACGCC
913	8mer	0.035013	CTGGCCAC
914	8mer	0.034942	CTGGGCTG
915	8mer	0.034939	GTGGA
916	8mer	0.034897	GCACTTGT
917	8mer	0.034893	TCGCTTGA
918	8mer	0.034842	ACTCAGTG
919	8mer	0.034833	GAACTTAA
920	8mer	0.034822	CCCAGGC
921	8mer	0.034799	TCCCGCAA
922	8mer	0.034779	TCTTAAGT
923	8mer	0.034736	TGTGGCCG
924	8mer	0.034735	CCCTGATT
925	8mer	0.03473	CGCAAAG
926	8mer	0.034724	ACAAGCGA
927	8mer	0.034711	CGGGCGTT
928	8mer	0.034709	GAGGGCCT
929	8mer	0.034708	GGCGTCGG

930	8mer	0.034707	ACCAGGGG
931	8mer	0.0347	AGGCACCT
932	8mer	0.034691	CTAATGGT
933	8mer	0.034684	ACCGCTGG
934	6mer	0.034677	GGGCAG
935	6mer	0.034677	AGGCCG
936	8mer	0.034665	CATGGCAT
937	8mer	0.034628	GCGTTGTT
938	6mer	0.034587	TGTTGG
939	8mer	0.034567	CACCGAAG
940	6mer	0.034562	CCCGGC
941	8mer	0.034557	CTGAAAAA
942	8mer	0.034523	CGCGGGAG
943	6mer	0.034513	CCAGGC
944	8mer	0.034485	CGGTTAAC
945	8mer	0.034454	CGGGAGTT
946	8mer	0.03441	CGCCGCCT
947	8mer	0.034398	CCCTCACG
948	8mer	0.034398	AGGGCCGG
949	8mer	0.034374	CCCGGGCT
950	8mer	0.034353	AGGGGCGT
951	8mer	0.034352	GGGGGACG
952	8mer	0.034348	CCGTATTA
953	8mer	0.034348	GGATTACA
954	6mer	0.034343	CGAGAT
955	8mer	0.034336	CCCTACCA
956	8mer	0.03431	GAAATTGA
957	8mer	0.034267	GTCGGACC
958	6mer	0.034264	CTAGCG
959	8mer	0.034215	TTTTGTAT
960	8mer	0.034207	CCGGGCAT
961	8mer	0.034187	TTGGAAAG
962	8mer	0.034168	GATGCCCG
963	8mer	0.034158	GTGCCGGG
964	8mer	0.034099	CCGGGGCG
965	8mer	0.034094	GAAGGCGG
966	8mer	0.034065	GCAATGCG
967	8mer	0.034064	TCCCCTCG
968	8mer	0.034041	GCCCACAA
969	8mer	0.034036	GCTGCTAG
970	8mer	0.034008	GGCGGCAC
971	8mer	0.034007	ACCTGGTC
972	8mer	0.033999	AACAAGCG
973	8mer	0.03399	TGCCTAAG
974	8mer	0.033985	TACAAGCG
975	8mer	0.033952	CTGGTCTC
976	6mer	0.033927	TGCGCC
977	8mer	0.033917	GGTCTCGA
978	8mer	0.033911	TATGATAC
979	8mer	0.0339	CCGCGCTT
980	8mer	0.033872	GATTTTTA
981	8mer	0.033869	GCACCCCT
982	6mer	0.033863	CCTCCC
983	8mer	0.033854	CCGTAGAC
984	8mer	0.033778	TCGTGACA
985	8mer	0.033767	GTCCCTGG
986	8mer	0.033762	CGAATTAC
987	8mer	0.03376	GTGCCTAA

988	8mer	0.033714	ACTGCGAT
989	8mer	0.033696	GCACCACG
990	8mer	0.033673	CCGCCGTT
991	6mer	0.033647	CCTCGG
992	8mer	0.033602	CGCAGCGA
993	8mer	0.033572	TGTTACAT
994	8mer	0.033562	TGACGGGC
995	8mer	0.033561	AATACGGA
996	8mer	0.033552	ATAGGTAC
997	8mer	0.033546	TTTGCAGC
998	8mer	0.033524	AACCCGCG
999	8mer	0.03352	CGTATGTG
1000	8mer	0.033512	CTGCTGTT

edtp568 according to F-score order

Order	Feature	F-score	kmer
701	8mer	0.039192	GGTCGAGG
702	8mer	0.039183	TTTTTTAG
703	8mer	0.039178	CCGCCAAT
704	8mer	0.039146	TGGGGCGG
705	8mer	0.039133	TACCGGCA
706	8mer	0.039092	CATGCTAC
707	8mer	0.03906	CCCGCAAA
708	5mer	0.039041	AGCCG
709	8mer	0.039038	AACTGTAC
710	8mer	0.039002	GCAGAAAA
711	8mer	0.039	CACGCAGG
712	6mer	0.039	GCCCCG
713	8mer	0.038999	GCATCGTC
714	8mer	0.038984	GAGGGCGA
715	8mer	0.038973	GGTGA AAC
716	8mer	0.038944	CGGATGTC
717	8mer	0.038939	GCGGAGCG
718	8mer	0.03893	AAGCCGCA
719	6mer	0.038902	CGCGGC
720	8mer	0.03885	TGACGCGA
721	8mer	0.038789	CCACCGAT
722	6mer	0.038761	GCGCCC
723	8mer	0.038745	TGCGTGAG
724	8mer	0.038735	TCATTGCC
725	8mer	0.03869	AGCGAGAC
726	8mer	0.038687	TGTGTTAC
727	8mer	0.038675	CCGTAGCT
728	8mer	0.038664	ACATCCGT
729	8mer	0.03865	CCTCGTCG
730	8mer	0.038649	CTGCGCCG
731	8mer	0.038618	CTATGGGA
732	8mer	0.038616	TGCAACCT
733	8mer	0.038607	GATTAGAG
734	8mer	0.0386	TCGCCGCC
735	8mer	0.038583	AGGATGTT
736	8mer	0.038573	CCGCGGAT
737	8mer	0.038571	ATGAGCTA
738	8mer	0.038566	ACTCGTCA
739	8mer	0.038562	GCGGAGGT
740	6mer	0.038552	AGGGCG
741	8mer	0.038544	ATAACGAG
742	8mer	0.038543	GAATATCT
743	8mer	0.038531	GACACGGG
744	6mer	0.038521	CCCCGC
745	8mer	0.038508	AACATGGT
746	8mer	0.038481	ATAGTGCG
747	6mer	0.038469	GCGGAG
748	8mer	0.038451	GACGGGCG

749	8mer	0.038446	TGCAAGTT
750	8mer	0.038361	CGCGGATT
751	8mer	0.03836	GCCCGGAG
752	8mer	0.038343	ACGTCCGG
753	8mer	0.038332	TAGATCGT
754	5mer	0.038321	TGGTA
755	8mer	0.038297	GCGGAGTT
756	8mer	0.038275	ACGCCACA
757	8mer	0.03827	ATGGGACC
758	8mer	0.038252	CGTCATGA
759	8mer	0.038251	CATCCGTC
760	8mer	0.038249	GAGGGCGC
761	8mer	0.038244	TGATTTCG
762	8mer	0.038228	TGCGCTCC
763	8mer	0.038226	GGGCTCAC
764	5mer	0.038202	GGCGG
765	8mer	0.038193	GCACGTGG
766	8mer	0.038168	GGTGTCTG
767	8mer	0.038135	CGGCTGGG
768	8mer	0.038124	CTGACGCG
769	8mer	0.03805	TTAAATGA
770	8mer	0.038048	TCACCGCG
771	8mer	0.038015	CTCACATC
772	8mer	0.037947	TGCGCCGA
773	8mer	0.037943	GCAAACGT
774	8mer	0.037942	CCGGAAC
775	8mer	0.037932	TGGCCAGG
776	8mer	0.037889	CGAATGAT
777	8mer	0.037866	ACGCGACA
778	8mer	0.037866	GACGCGAC
779	8mer	0.037866	TGTCGACC
780	8mer	0.037866	TAAACCGC
781	8mer	0.037866	GCTCGCGT
782	8mer	0.037861	CGGCCTCC
783	8mer	0.037833	GACGCCTC
784	8mer	0.037833	TTCAATGA
785	8mer	0.037822	CTCGTCGA
786	6mer	0.037789	TCGCGC
787	8mer	0.037764	GGCCGTCT
788	8mer	0.037759	TGGGGTAT
789	8mer	0.037755	TGTCCCGC
790	8mer	0.03775	CCTTCACT
791	8mer	0.037746	ATATGTGA
792	8mer	0.037736	CCGAAGTC
793	5mer	0.037698	TGCCC
794	8mer	0.037682	TGGTTGCT
795	8mer	0.037677	CTGCTCAC
796	8mer	0.037644	CAGGCGGG
797	8mer	0.03761	GCCCGTCG
798	8mer	0.037582	GCGGGCAT
799	8mer	0.037572	CACGAGAC
800	8mer	0.037572	GCAATGCC
801	8mer	0.037561	CTGCTGCG
802	8mer	0.037556	CCGAGGGC
803	8mer	0.037538	AGGTCGTG
804	8mer	0.037517	GGCTGGTC
805	8mer	0.037512	GAAAGCGA
806	8mer	0.037508	GGGGCTGA

807	5mer	0.037508	GGCGA
808	8mer	0.037503	GGCTGCTA
809	8mer	0.037495	CTTCATCG
810	8mer	0.03748	CGTGGGTG
811	8mer	0.03746	ATTAGCCG
812	8mer	0.037429	GCTGCCGG
813	8mer	0.037404	AAGCCGGC
814	8mer	0.037394	GGGGCGG
815	5mer	0.037369	CAGGC
816	8mer	0.037359	CCCGTCTG
817	8mer	0.037355	CACGGGAT
818	8mer	0.037317	ACCCGAAT
819	8mer	0.037314	TCGCCCCG
820	8mer	0.037312	GCCTATGG
821	8mer	0.03727	AATACCGG
822	8mer	0.037266	CCCCGCCC
823	8mer	0.037257	CGGCCTAT
824	8mer	0.037252	GCTGCGCA
825	8mer	0.037222	CCCGGCAC
826	8mer	0.037219	ACCCGGGG
827	8mer	0.037189	AAGCGAAA
828	6mer	0.037183	CCCGTC
829	8mer	0.037161	GAATCGCT
830	8mer	0.037139	GCGCCCGG
831	8mer	0.037131	ACCCGGCC
832	8mer	0.037086	ATTAGGGA
833	8mer	0.037082	CCGCTTAG
834	6mer	0.037077	GGCCCG
835	8mer	0.037036	GGCATTGG
836	8mer	0.037023	GGCATTCC
837	8mer	0.036989	TAAAATGG
838	8mer	0.036988	TCCTTTCC
839	6mer	0.036974	GACCTC
840	8mer	0.036971	CCTCCCAA
841	8mer	0.036962	CGGAATAT
842	8mer	0.036961	TTGGGCCC
843	8mer	0.036937	CGCCAATC
844	8mer	0.036924	TGACCTTA
845	8mer	0.036905	GCGAAATT
846	8mer	0.036883	CTTCAAGC
847	6mer	0.036873	ACACGG
848	8mer	0.036856	AGGCTGGT
849	8mer	0.036827	GCGCGCGG
850	8mer	0.036785	GTGCCGAA
851	8mer	0.036765	TTTCCCGA
852	8mer	0.036755	GAATCCGG
853	8mer	0.036745	CCGACTTT
854	6mer	0.036732	GAGCCC
855	8mer	0.036719	GTAGTGGG
856	6mer	0.036697	GGCCCC
857	5mer	0.036696	GAGGG
858	8mer	0.036678	CAAAAAAA
859	8mer	0.036676	GCGACATC
860	8mer	0.036655	ACGGGTTC
861	8mer	0.036648	TTCAAGGT
862	8mer	0.036629	TGTACGTT
863	8mer	0.036619	GTCGACCC
864	8mer	0.036608	GCGCCACT

865	8mer	0.036604	CTGGTATA
866	8mer	0.036599	CGTCAAGG
867	8mer	0.036577	GACTGCGA
868	8mer	0.036576	CAGTAGGA
869	5mer	0.036576	CGGGA
870	8mer	0.036575	CGGCCAC
871	8mer	0.036574	GAAGTCGG
872	8mer	0.036528	ATGCACGC
873	8mer	0.036518	TAACGAGG
874	8mer	0.036479	TCGGGGG
875	8mer	0.036471	CGGGTGGC
876	8mer	0.036463	CAATATAC
877	8mer	0.036425	TCGGCTGT
878	8mer	0.03639	CTATGTTG
879	6mer	0.036385	TCGCGT
880	8mer	0.036376	ACGCTCAT
881	8mer	0.036373	GCGTTCTC
882	6mer	0.036294	TGTCGG
883	8mer	0.03628	GACGGCCT
884	8mer	0.036274	GAGGTCAT
885	8mer	0.036257	ACGGCCTG
886	8mer	0.036192	GTTATCTG
887	8mer	0.03619	CTACAGTA
888	8mer	0.036189	TATAAGGT
889	8mer	0.036163	CGACTTTG
890	8mer	0.036153	TGAGCCGA
891	8mer	0.036139	CGTAGATA
892	8mer	0.036137	CTAAACCG
893	6mer	0.036132	TCGACG
894	8mer	0.036098	CTTGCGTC
895	8mer	0.036073	AATTATCC
896	8mer	0.036068	GAGCGACG
897	8mer	0.036049	GTCCAGAC
898	8mer	0.036048	CGGGGGCC
899	8mer	0.036041	ACGCCCGA
900	5mer	0.036039	GGCCC
901	8mer	0.036032	GGCGCCCG
902	8mer	0.036022	GTTTCGCC
903	8mer	0.036005	TCCTGGGC
904	8mer	0.036004	TTGTTAAT
905	8mer	0.03599	CAGCCTCC
906	8mer	0.035982	GGTGGCGC
907	8mer	0.035958	CGTTCAGG
908	6mer	0.035953	CCAACG
909	8mer	0.035936	CGAGTTCT
910	8mer	0.03593	CGCGCGGG
911	5mer	0.035879	GCGGA
912	8mer	0.035852	ACCCGCGC
913	6mer	0.035848	CAAGCG
914	8mer	0.035833	ACCTCGAC
915	8mer	0.035761	CGAACAGA
916	8mer	0.035756	CTGTCCCG
917	8mer	0.035747	CTCCTGTG
918	8mer	0.035746	GGCTCACG
919	8mer	0.035744	GACCTGGG
920	8mer	0.035735	GAACCTCGT
921	8mer	0.035724	ACTGCACG
922	8mer	0.035716	GAGGGCCG

923	8mer	0.03568	ACCATTAA
924	6mer	0.03567	CGGAGC
925	8mer	0.035668	CAAGCGAA
926	8mer	0.035666	GCGGCCAA
927	8mer	0.035652	GCCATCTA
928	5mer	0.035632	ACCTC
929	8mer	0.035597	GTCTAGGC
930	8mer	0.035581	ATCATTAT
931	8mer	0.035577	TGGGCCCC
932	8mer	0.035562	TAACCGCG
933	8mer	0.035558	GGCTGGGA
934	8mer	0.035557	GTAAACAT
935	6mer	0.035553	CCGGGG
936	8mer	0.035551	CGTCCAGG
937	6mer	0.035534	GCACGA
938	8mer	0.035506	CGGACTAG
939	6mer	0.035485	GCCCTA
940	8mer	0.035482	GGACCTAG
941	8mer	0.035439	TAGTGTCC
942	8mer	0.035432	CACTACAG
943	8mer	0.035424	CGCTTCTT
944	8mer	0.035384	TCCCAACG
945	8mer	0.035383	CCGGGGTG
946	8mer	0.035364	CGGCGCGC
947	8mer	0.03536	TGGGATTA
948	8mer	0.035351	AGCCGGGG
949	8mer	0.035334	TTCCTAG
950	8mer	0.03528	CCCTTCGA
951	6mer	0.035237	GGCCGG
952	8mer	0.03522	CTGCGCCT
953	6mer	0.035216	CCCAGG
954	8mer	0.03518	CTCGGCTG
955	8mer	0.035126	CACGCCAC
956	8mer	0.035113	CTCGGGTC
957	8mer	0.035093	CGGCGTGC
958	6mer	0.035071	CCGGCA
959	8mer	0.035069	CTGTGAC
960	8mer	0.035065	TAGTGGGA
961	8mer	0.035051	CTCAGCCT
962	8mer	0.03505	GGCCGCTC
963	8mer	0.035019	GGCCAGGT
964	8mer	0.035018	TTCACGCC
965	8mer	0.035013	CTGGCCAC
966	8mer	0.034942	CTGGGCTG
967	8mer	0.034939	GTGGA
968	8mer	0.034897	GCACTTGT
969	8mer	0.034893	TCGCTTGA
970	5mer	0.034853	GCCGA
971	8mer	0.034842	ACTCAGTG
972	8mer	0.034833	GAACCTAA
973	8mer	0.034822	CCCCAGGC
974	8mer	0.034799	TCCCGCAA
975	8mer	0.034779	TCTTAAGT
976	5mer	0.034747	CGCAC
977	8mer	0.034736	TGTGGCCG
978	8mer	0.034735	CCCTGATT
979	8mer	0.03473	CGCAAAAG
980	8mer	0.034724	ACAAGCGA

981	8mer	0.034711	CGGGCGTT
982	8mer	0.034709	GAGGGCCT
983	8mer	0.034708	GGCGTCGG
984	8mer	0.034707	ACCAGGGG
985	8mer	0.0347	AGGCACCT
986	8mer	0.034691	CTAATGGT
987	8mer	0.034684	ACCGCTGG
988	6mer	0.034677	GGGCAG
989	6mer	0.034677	AGGCCG
990	8mer	0.034665	CATGGCAT
991	8mer	0.034628	GCGTTGTT
992	6mer	0.034587	TGTTGG
993	8mer	0.034567	CACCGAAG
994	6mer	0.034562	CCCGGC
995	8mer	0.034557	CTGAAAAA
996	8mer	0.034523	CGCGGGAG
997	6mer	0.034513	CCAGGC
998	8mer	0.034485	CGGTTAAC
999	8mer	0.034454	CGGGAGTT
1000	8mer	0.03441	CGCCGCCT

-

;
;
;
;
;

;
;
;
;
;

;
;
-