

TableS1 The top 1000 features in the optimal feature set of the IncLocPredtp8 according to F-score order

Order	Feature	F-score	kmer	Order	Feature	F-score	kmer	Order	Feature	F-score	kmer
1	8mer	0.10931	TCCCAAAG	351	8mer	0.04596	GGCGGGGC	701	8mer	0.037312	GCCTATGG
2	8mer	0.107445	GCGGGGCG	352	8mer	0.04584	CCCTGGGC	702	8mer	0.03727	AATACCGG
3	8mer	0.106471	AACGGCCC	353	8mer	0.04582	CGCCGGAA	703	8mer	0.037266	CCCCGCC
4	8mer	0.105616	CGGTGACG	354	8mer	0.04581	ATCTTAAT	704	8mer	0.037257	CGGCCTAT
5	8mer	0.098194	CGGGTCAC	355	8mer	0.04574	CCGCGAAC	705	8mer	0.037252	GCTGCGCA
6	8mer	0.095685	TCACGGGA	356	8mer	0.0456	TAAGGGCT	706	8mer	0.037222	CCCCGCAC
7	8mer	0.09567	TGCGGTTT	357	8mer	0.04557	CGACCTCC	707	8mer	0.037219	ACCCGGGG
8	8mer	0.09254	GGGGGACC	358	8mer	0.04555	ACGCTCGT	708	8mer	0.037189	AAGCGAAA
9	8mer	0.091784	CCACGACC	359	8mer	0.04554	CGCGGGCA	709	8mer	0.037161	GAATCGCT
10	8mer	0.091036	CGCCCCGG	360	8mer	0.04554	GCGCCAAT	710	8mer	0.037139	GCGCCCGG
11	8mer	0.09051	TCAACGTC	361	Triplet	0.04548	Triplet	711	8mer	0.037131	ACCCGGCC
12	8mer	0.088855	CGAAGTCG	362	8mer	0.04545	GACCCGGC	712	8mer	0.037086	ATTAGGGA
13	8mer	0.088667	CGGGGGAC	363	8mer	0.04545	GGCTGATG	713	8mer	0.037082	CCGCTTAG
14	8mer	0.08771	ATGGGGCG	364	8mer	0.0454	GCACAATT	714	8mer	0.037036	GGCATTCC
15	8mer	0.087693	TCGGCCCC	365	8mer	0.04529	GGTGTCTA	715	8mer	0.037023	GGCATTCC
16	8mer	0.087595	AGCGGAGC	366	8mer	0.04526	TGTTTACG	716	8mer	0.036989	TAAAGATG
17	8mer	0.087322	AATCGACG	367	8mer	0.04524	AGGGCGCA	717	8mer	0.036988	TCCTTTCC
18	8mer	0.085901	GCTCACGG	368	8mer	0.04523	TAGGCATT	718	8mer	0.036971	CCTCCCAA
19	8mer	0.084716	GTAGGGCG	369	8mer	0.04521	CCGAAGTC	719	8mer	0.036962	CGGAATAT
20	8mer	0.084641	CGCAGTTC	370	8mer	0.04521	CGTACATT	720	8mer	0.036961	TTGGGCCC
21	8mer	0.083424	CCGCAGTT	371	8mer	0.0452	CGCGTACA	721	8mer	0.036937	CGCCAATC
22	8mer	0.082921	GCGGAGCC	372	8mer	0.04518	CCATCGGT	722	8mer	0.036924	TGACCTTA
23	8mer	0.081369	CATCGGCT	373	8mer	0.04515	CGAGACCA	723	8mer	0.036905	GCGAAATT
24	8mer	0.079854	AGTAGGAC	374	8mer	0.04515	CCGGACTT	724	8mer	0.036883	CTTCAAGC
25	8mer	0.079681	CCCAAAGT	375	8mer	0.04512	ATGTGCGG	725	8mer	0.036856	AGGCTGGT
26	8mer	0.079471	AAGGGCTG	376	8mer	0.04507	CTAGCGGA	726	8mer	0.036827	GCGCGCGG
27	8mer	0.078261	GGGATGCG	377	8mer	0.04506	TCGACGAA	727	8mer	0.036785	GTGCCGAA
28	8mer	0.077741	GATCCCGA	378	8mer	0.04506	GGGCGCGG	728	8mer	0.036765	TTTCCCGA
29	8mer	0.077287	GCGTGAGC	379	8mer	0.04499	CCCCTTAT	729	8mer	0.036755	GAATCCGG
30	8mer	0.077177	CGGGCGGG	380	8mer	0.04495	GGCAGAAA	730	8mer	0.036745	CCGACTTT
31	8mer	0.077057	GCGGGGGA	381	8mer	0.04493	GTTGCTAT	731	8mer	0.036719	GTAGTGGG
32	8mer	0.076646	AGACGGCG	382	8mer	0.04492	GGTTGGTC	732	8mer	0.036678	CAAAAAAA
33	8mer	0.075641	CTTGATAT	383	8mer	0.04491	ACCGTGCC	733	8mer	0.036676	GCGACATC
34	8mer	0.07546	CTTAAACC	384	8mer	0.04484	ACGCCTCG	734	8mer	0.036655	ACGGGTTT
35	8mer	0.075416	GAGCCCGA	385	8mer	0.04481	CGTGTTC	735	8mer	0.036648	TTCAGAGT
36	8mer	0.07508	CGTAGCTT	386	8mer	0.0448	ATGCGCGA	736	8mer	0.036629	TGTACGTT
37	8mer	0.074414	CCCCGAA	387	8mer	0.04471	CTTTCGTC	737	8mer	0.036619	GTCGACCC
38	8mer	0.073205	GGCGAGGT	388	SC-PseDNC	0.04466	SC-PseDNC	738	8mer	0.036608	GCGGCACT
39	8mer	0.073205	CGGCACCG	389	8mer	0.04457	GGGCGCGG	739	8mer	0.036604	CTGGTATA
40	8mer	0.072737	CGCAGCGT	390	8mer	0.0445	ACCGCGAA	740	8mer	0.036599	CGTCAAGG
41	8mer	0.072567	CATCGCGG	391	8mer	0.04444	GGGGATAT	741	8mer	0.036577	GACTGCGA
42	8mer	0.072132	TGGGCCCG	392	8mer	0.0444	TTTGTCTC	742	8mer	0.036576	CAGTAGGA
43	8mer	0.072036	CCCGTTTC	393	8mer	0.0444	AGATTGCG	743	8mer	0.036575	CGGCCCAC
44	8mer	0.072034	GTGCGTGA	394	8mer	0.04432	CGAGAATC	744	8mer	0.036574	GAAGTCGG
45	8mer	0.071804	GGATGCGG	395	8mer	0.04426	CAGTTCGG	745	8mer	0.036528	ATGCACGC
46	8mer	0.070518	AAGGGGCG	396	8mer	0.04423	ATCGACGA	746	8mer	0.036518	TAACGAGG
47	8mer	0.07038	CCGAAGGC	397	8mer	0.04422	CACGGCAA	747	8mer	0.036479	TTGCGGGG
48	8mer	0.069523	CCTGTGTT	398	8mer	0.04419	ACCCCCGA	748	8mer	0.036471	CGGTTGGC
49	8mer	0.069495	GTGAAGA	399	8mer	0.04418	CGAGAGCT	749	8mer	0.036463	CAATATAC
50	8mer	0.068895	ATCCCGAA	400	8mer	0.04415	GCGACTCG	750	8mer	0.036425	TCGGCTGT
51	8mer	0.068761	CTTGAACC	401	8mer	0.0441	CCGCTGTT	751	8mer	0.03639	CTATGTGT
52	8mer	0.068296	ACGAGCCC	402	8mer	0.04408	ACTGCTCA	752	8mer	0.036376	ACGCTCAT
53	8mer	0.068213	AGACCTCT	403	8mer	0.04407	GAGCCCCG	753	8mer	0.036373	GCGTTTCT
54	8mer	0.068069	ACGTGCGG	404	8mer	0.04403	CCAAAGTG	754	8mer	0.03628	GACGGCCT
55	8mer	0.068033	GTGCGACT	405	8mer	0.04392	CCGCAGCG	755	8mer	0.036274	GAGGTCAT
56	8mer	0.067987	GACCATCG	406	8mer	0.0439	CACGCGCG	756	8mer	0.036257	ACGGCCGT
57	8mer	0.067781	ACGAATCA	407	8mer	0.04389	CGCCGCCG	757	8mer	0.036192	GTTATCTG
58	8mer	0.067781	TGTCGAC	408	8mer	0.04387	GCCCTCTGA	758	8mer	0.03619	CTACAGTA
59	8mer	0.067781	TAACCGCC	409	8mer	0.04382	CAAAGTAC	759	8mer	0.036189	TATAAGGT
60	8mer	0.067781	GTCACCGC	410	8mer	0.04381	GAGGGGCG	760	8mer	0.036163	CGACTTTG
61	8mer	0.067781	CGAGTCGG	411	8mer	0.04381	GGGGGTCG	761	8mer	0.036153	TGAGCCGA
62	8mer	0.067781	TATGCGGG	412	8mer	0.04379	GCCGGACT	762	8mer	0.036139	CGTAGATA
63	8mer	0.067781	ATGCGGGG	413	8mer	0.04379	CGCGCCGA	763	8mer	0.036137	CTAAACCG
64	8mer	0.067692	AGCCACCG	414	8mer	0.04377	GCAACGTC	764	8mer	0.036098	CTGCGTCT
65	8mer	0.067483	GCGTGCTC	415	8mer	0.04377	ACCTAGAC	765	8mer	0.036073	AATTATCC
66	8mer	0.067415	TCCGGGCG	416	8mer	0.04375	TTTTTGTA	766	8mer	0.036068	GAGCGACG
67	8mer	0.067239	GGAATCCG	417	8mer	0.04373	AGATACTA	767	8mer	0.036049	GTCCAGAC
68	8mer	0.067199	TGACGAGC	418	8mer	0.04372	AACTCGTC	768	8mer	0.036048	CGGGGGCC
69	8mer	0.067076	CGCACAGT	419	8mer	0.04371	GGAACCTA	769	8mer	0.036041	ACGCCCGA
70	8mer	0.066653	GCGCGTCC	420	8mer	0.04367	AAGGTGCC	770	8mer	0.036032	GCGCGCCG
71	8mer	0.066595	TCCGCCGG	421	8mer	0.04364	TATTAGGG	771	8mer	0.036022	GTTTCGCC
72	8mer	0.066489	GGCCCGTG	422	8mer	0.04363	CCGCAACT	772	8mer	0.036005	TCCTGGGC
73	8mer	0.066467	GCGGATCT	423	8mer	0.04362	TTGTACGG	773	8mer	0.036004	TTGTTAAT
74	8mer	0.06636	GGGCCCTT	424	8mer	0.04358	CCGTGTCC	774	8mer	0.03599	CAGCCTCC
75	8mer	0.06628	CCTGGGCC	425	8mer	0.04352	CGCCTATA	775	8mer	0.035982	GGTGGCGC
76	8mer	0.066234	GACGAATC	426	8mer	0.04351	TATGTCCG	776	8mer	0.035958	CGTTACGG
77	8mer	0.066039	AACGTCCG	427	8mer	0.0435	TACTGCCA	777	8mer	0.035936	CGAGTTCT
78	8mer	0.06574	TTATGCGG	428	8mer	0.04347	TATAGCAC	778	8mer	0.03593	CGCGCGGG
79	8mer	0.06574	CGACGAAT	429	8mer	0.04347	GGTTGCAG	779	8mer	0.035852	ACCCGCGC
80	8mer	0.065703	GCGGTAAC	430	8mer	0.04346	TTTTGGCA	780	8mer	0.035833	ACCTCGAC
81	8mer	0.065606	CCTGGGCG	431	8mer	0.04342	CCGCGATT	781	8mer	0.035761	CGAACAGA
82	8mer	0.065523	TTGTATCC	432	8mer	0.04336	GGTTGTTC	782	8mer	0.035756	CTGTCCCG
83	8mer	0.065484	TGGTCTCG	433	8mer	0.04334	GAGCGGGG	783	8mer	0.035747	CTCCTGTG
84	8mer	0.065447	CCACTGCA	434	8mer	0.04333	GAGGCCGT	784	8mer	0.035746	GGCTCAGG
85	8mer	0.065364	CGACAGGA	435	8mer	0.04332	GTGTGCTA	785	8mer	0.035744	GACCTGGG
86	8mer	0.065247	CATGACAC	436	8mer	0.0433	CTGCGAGT	786	8mer	0.035735	GAACCTGT
87	8mer	0.065246	ATCCGCCG	437	8mer	0.0433	CGGCCCTT	787	8mer	0.035724	ACTGCACG
88	8mer	0.064857	TGTCGGTA	438	8mer	0.04328	GCAGTTCG	788	8mer	0.035716	GAGGGCCG
89	8mer	0.064394	CTCATTGC	439	8mer	0.04326	TCCTGTGT	789	8mer	0.03568	ACCATTAA

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	AACCGCCG
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	AGGTGCGA
104	8mer	0.062472	GAAACCCG
105	8mer	0.062448	TTGCCGGT
106	8mer	0.06233	TCAAGCAC
107	8mer	0.062312	CCAAGATC
108	8mer	0.062211	ATTGCGCG
109	8mer	0.061952	TCACGGTG
110	8mer	0.06195	GATGCGGA
111	8mer	0.061771	GGCCCCCT
112	8mer	0.061704	CGGGGCGA
113	8mer	0.061609	TCGTCCAA
114	8mer	0.061468	CGGGCGCT
115	8mer	0.061418	CGATGATA
116	8mer	0.06132	GGGCGAGG
117	8mer	0.061146	TCGGTACT
118	8mer	0.061133	CGGTGTCT
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	ACTCCCGC
125	8mer	0.060568	CAAGCGAT
126	8mer	0.060416	AGGGGATA
127	8mer	0.060304	CCTCGGCC
128	8mer	0.060137	GCATTCCG
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	GTGTACT
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	AAACGTGC
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	GTCCGTAT
170	8mer	0.05707	TCGTAGGG
171	8mer	0.056971	CGGGCATC
172	8mer	0.056836	GCCTGTGT
173	8mer	0.056612	TGCGGTAA
174	8mer	0.056504	GGAGCCCG
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

440	8mer	0.04325	GGGCGGTG
441	8mer	0.04325	CCGTTTCC
442	8mer	0.0432	TCGTGCGA
443	8mer	0.04319	TTGCGCCA
444	8mer	0.04314	AGCATTGC
445	8mer	0.04313	ACGCGTAC
446	8mer	0.04313	CTTCCCGC
447	8mer	0.0431	GGGCGGGT
448	8mer	0.04309	ACCCGCAC
449	8mer	0.04305	GAGCCCGG
450	8mer	0.04305	ATCCCGAC
451	8mer	0.04304	GAACGAAT
452	8mer	0.04297	CAGTACGT
453	8mer	0.04296	CTTAAGTA
454	8mer	0.04296	GAGATACT
455	8mer	0.04295	CGTCGTCT
456	8mer	0.0429	GACGCGTA
457	8mer	0.04287	GCACTCCA
458	8mer	0.04286	CGCGCGCG
459	8mer	0.04285	TGCTTTTT
460	8mer	0.04278	CCGGAATA
461	8mer	0.04275	ATTGCGCC
462	8mer	0.04274	CGAGGTCT
463	8mer	0.04273	CAGGCCAC
464	8mer	0.04269	GATCCGGG
465	8mer	0.04266	TCCTCCGT
466	8mer	0.04263	CTAAACAT
467	8mer	0.04261	CGGGCGAG
468	8mer	0.04258	CACAGATG
469	8mer	0.04258	GTAAGGGC
470	8mer	0.04257	GGACGCAA
471	8mer	0.04255	CCTGCGCG
472	8mer	0.04255	GGCGGTGT
473	8mer	0.04248	TGCATTGT
474	8mer	0.04246	AGGAGGCC
475	Triplet	0.04242	Triplet
476	8mer	0.04237	CCGAAGGG
477	8mer	0.04234	TGAGCCAC
478	8mer	0.04231	TCGGGGGT
479	8mer	0.04227	CCCGGGAG
480	SC-PseDNC	0.04225	SC-PseDNC
481	8mer	0.04224	CAACGTCT
482	8mer	0.04222	CCGAAGAC
483	8mer	0.04217	TCGAGAGC
484	8mer	0.04215	ACGCGCCA
485	8mer	0.04214	CCGTGGAA
486	8mer	0.04213	AACTGCGG
487	8mer	0.04207	CGAGGCAC
488	8mer	0.04206	AGGTCGTA
489	8mer	0.04206	GCCCGGCA
490	8mer	0.04206	AAGCGAGG
491	8mer	0.04204	TGCTGCGC
492	8mer	0.04204	CGGAGGTT
493	8mer	0.04194	GTCCACAC
494	8mer	0.04194	AGACGGTC
495	8mer	0.0419	CCATAGTG
496	8mer	0.04188	CGAAGGGC
497	8mer	0.04183	GCCGCCAG
498	8mer	0.04182	AGCTTTTT
499	8mer	0.04181	CCCGCCCC
500	8mer	0.04181	TCGACCTC
501	8mer	0.04173	CCCGGAGG
502	8mer	0.0417	CACCACTT
503	8mer	0.0417	GACCTCGG
504	Triplet	0.04159	Triplet
505	8mer	0.04158	TCGGTTTG
506	8mer	0.04153	CGGAAACG
507	8mer	0.04145	ACAATGAG
508	8mer	0.04145	CCCTCGTC
509	8mer	0.04144	AGGTTCAA
510	8mer	0.04144	GCCCCCGG
511	8mer	0.04142	CGGGGCCC
512	8mer	0.04142	GGGAGATC
513	8mer	0.04138	TACTCGGG
514	8mer	0.04134	TACGTAAA
515	8mer	0.0413	TCGCGGGC
516	8mer	0.04129	CCGCTATT
517	8mer	0.04124	GACCTGGC
518	8mer	0.04118	GCCCGAGG
519	8mer	0.04114	GCATTCCG
520	8mer	0.04113	CATTGCGT
521	8mer	0.04111	GGGGCGGG
522	8mer	0.04108	AAGGTTCC
523	8mer	0.04108	TTGCGGAA
524	8mer	0.04101	CTAATGCA
525	8mer	0.04095	CCGTCTCT
526	8mer	0.04083	CATCGCCG
527	8mer	0.04082	CCGCGCGC
528	8mer	0.0408	GGGACCGC
529	8mer	0.0408	TAACGGGA
530	8mer	0.0408	CGCGATGA
531	8mer	0.0408	ACGCGCTA

790	8mer	0.035668	CAAGCGAA
791	8mer	0.035666	GCGGCCAA
792	8mer	0.035652	GCCATCTA
793	8mer	0.035597	GTCTAGGC
794	8mer	0.035581	ATCATTAT
795	8mer	0.035577	TGGGCCCC
796	8mer	0.035562	TAACCGCG
797	8mer	0.035558	GGCTGGGA
798	8mer	0.035557	GTAACAT
799	8mer	0.035551	CGTCCAGG
800	8mer	0.035506	CGGACTAG
801	8mer	0.035482	GGACCTAG
802	8mer	0.035439	TAGTGTCC
803	8mer	0.035432	CACTACAG
804	8mer	0.035424	CGCTTCTT
805	8mer	0.035384	TCCCAACG
806	8mer	0.035383	CCGGGGTG
807	8mer	0.035364	CGGCGCGC
808	8mer	0.03536	TGGGATTA
809	8mer	0.035351	AGCGGGGG
810	8mer	0.035334	TTCACTAG
811	8mer	0.03528	CCCTTCGA
812	8mer	0.03522	CTGCGCCT
813	8mer	0.03518	CTCGGCTG
814	8mer	0.035126	CACGCCAC
815	8mer	0.035113	CTCGGGTC
816	8mer	0.035093	CGGCGTGC
817	8mer	0.035069	CTGTGCAC
818	8mer	0.035065	TAGTGGGA
819	8mer	0.035051	CTCAGCCT
820	8mer	0.03505	GGCCGCTC
821	8mer	0.035019	GGCCAGGT
822	8mer	0.035018	TTACAGCC
823	8mer	0.035013	CTGGCCAC
824	8mer	0.034942	CTGGGCTG
825	8mer	0.034939	GTCCGAAA
826	8mer	0.034897	GCACTTGT
827	8mer	0.034893	TCGCTTGA
828	8mer	0.034842	ACTCAGTG
829	8mer	0.034833	GAACCTAA
830	8mer	0.034822	CCCCAGGC
831	8mer	0.034799	TCCCGCAA
832	8mer	0.034779	TCTTAAGT
833	8mer	0.034736	TGTGGCCG
834	8mer	0.034735	CCCTGACT
835	8mer	0.03473	CGCAAAAG
836	8mer	0.034724	ACAAGCGT
837	8mer	0.034711	CGGGCGTT
838	8mer	0.034709	GAGGGCCT
839	8mer	0.034708	GGCGTCGG
840	8mer	0.034707	ACCAGGGG
841	8mer	0.0347	AGGCACCT
842	8mer	0.034691	CTAATGTT
843	8mer	0.034684	ACCGCTGG
844	8mer	0.034665	CATGGCAT
845	8mer	0.034628	GCGTTGTT
846	8mer	0.034567	CACCGAAG
847	8mer	0.034557	CTGAAAAA
848	8mer	0.034523	CGCGGGAG
849	8mer	0.034485	CGGTTAAC
850	8mer	0.034454	CGGGAGTT
851	8mer	0.03441	CGCCGCCT
852	8mer	0.034398	CCCTCAGC
853	8mer	0.034398	AGGGCCGG
854	8mer	0.034374	CCCCGGCT
855	8mer	0.034353	AGGGGCGT
856	8mer	0.034352	GGGGGACG
857	8mer	0.034348	CCGTATTA
858	8mer	0.034348	GGATTACA
859	8mer	0.034336	CCCTACCA
860	8mer	0.03431	GAATTTGA
861	8mer	0.034267	GTCGGACC
862	8mer	0.034215	TTTTGTAT
863	8mer	0.034207	CCGGGCAT
864	8mer	0.034187	TTGGAAAG
865	8mer	0.034168	GATGCCCG
866	8mer	0.034158	GTGCGGGG
867	8mer	0.034099	CCGGGGCG
868	8mer	0.034094	GAAGGCGG
869	8mer	0.034065	GCAATTGC
870	8mer	0.034064	TCCCTCTG
871	8mer	0.034041	GCCCAACA
872	8mer	0.034036	GCTGTAGT
873	8mer	0.034008	GGCGGCAC
874	8mer	0.034007	ACCTGGTC
875	8mer	0.033999	AACAAGCG
876	8mer	0.03399	TGCCTAAG
877	8mer	0.033985	TACAAGCG
878	8mer	0.033952	CTGTCTCT
879	8mer	0.033917	GGTCTCGA
880	8mer	0.033911	TATGATAC
881	8mer	0.0339	CCGCGCTT

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	GCCTGGT
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	GCCCGGTG
204	8mer	0.053902	AAAATTAG
205	8mer	0.053867	CGAGCCCG
206	8mer	0.053538	GACCCCTC
207	8mer	0.053492	ATGCAACG
208	8mer	0.053471	TTCAATGC
209	8mer	0.05329	CCTCAGCC
210	8mer	0.053201	CCGCATTG
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
230	8mer	0.05164	GCCGCATT
231	8mer	0.051615	GGGTGAG
232	8mer	0.051532	TCGAGTCG
233	8mer	0.051301	TCCAGCT
234	8mer	0.051286	GAGCGCCA
235	8mer	0.051239	ATTGCTGT
236	8mer	0.051176	CAAGATCG
237	8mer	0.051123	GCGTTTCA
238	8mer	0.051068	GCTCCTCG
239	8mer	0.050908	ACGGGCAG
240	8mer	0.050867	AGATCGCG
241	8mer	0.050733	GATATTGA
242	8mer	0.050698	AGCCCCGA
243	8mer	0.050688	TTTTAGT
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	GGAACCGG
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	CACCTGAC
271	8mer	0.049199	TGCTGCGG
272	8mer	0.049125	CAAGGAGT
273	8mer	0.049102	CATTGCCG

532	8mer	0.0408	GCGCTAGG
533	8mer	0.0408	GTAACGGG
534	8mer	0.0408	GTCGTAA
535	8mer	0.04079	CTGGGCCT
536	8mer	0.04073	CCTGGCGT
537	8mer	0.04071	TGGGACCT
538	8mer	0.04069	CGTAATAC
539	8mer	0.04064	GGCGTTCA
540	8mer	0.04057	TCACGCCA
541	8mer	0.04056	CGAACACC
542	8mer	0.04056	TGCTTACC
543	8mer	0.04048	AACGGGTT
544	8mer	0.04046	CACGCCTC
545	8mer	0.04046	GGGATCGC
546	8mer	0.0404	CACCCCCG
547	8mer	0.0404	CGAACGGG
548	8mer	0.0404	CGTCCAGC
549	8mer	0.04039	TGGCGTGC
550	8mer	0.04038	TGGGACTG
551	8mer	0.04037	CCCCTGAC
552	8mer	0.04037	GCCGGCTG
553	8mer	0.04033	TCTTACCT
554	8mer	0.04033	CGCTGGTA
555	8mer	0.04031	GGGCCTAG
556	8mer	0.04031	CTGACGGC
557	8mer	0.04029	GGATTATA
558	8mer	0.04026	CGACCAAA
559	8mer	0.04025	GAACTGTC
560	8mer	0.04025	TTGGCCAG
561	8mer	0.04024	ACACGGGA
562	8mer	0.04022	AGCCGTTT
563	8mer	0.04021	TGCTCACG
564	8mer	0.04019	AGGAGGGG
565	8mer	0.04019	GGAGCGCT
566	8mer	0.04017	CCAATGCC
567	8mer	0.04017	ATCGCGCC
568	8mer	0.04004	GGGTGTTT
569	8mer	0.04003	GCCACTGC
570	8mer	0.03999	GGGAGGCT
571	8mer	0.03999	TATTGAAG
572	8mer	0.03991	CGAACTCG
573	8mer	0.0399	CGCTAAAT
574	8mer	0.03987	TAATCGAG
575	8mer	0.03987	AACCCGAA
576	8mer	0.03985	CGCGAACG
577	8mer	0.03983	CTGTGGGG
578	8mer	0.03981	CCTCGCGG
579	8mer	0.03981	CGAGATGG
580	8mer	0.03976	GCCGGTCT
581	8mer	0.03963	CATGTAAT
582	8mer	0.0396	TCGCGTCC
583	8mer	0.03955	TCGATGAG
584	8mer	0.03955	GGAGCACT
585	8mer	0.03951	AACGCTCC
586	8mer	0.03949	CGGCGCTG
587	8mer	0.03941	GCGCCGAG
588	8mer	0.0394	CTCGAGAT
589	8mer	0.03938	CGAAATTA
590	8mer	0.03936	CAATGCCG
591	8mer	0.03936	TCGACGCT
592	8mer	0.03927	TATACTCG
593	8mer	0.03922	GGCGCGTG
594	8mer	0.03921	GGTAGCGT
595	8mer	0.03919	GGTCGAGG
596	8mer	0.03918	TTTTTTAG
597	8mer	0.03918	CCGCAAT
598	8mer	0.03915	TGGGGCGG
599	8mer	0.03913	TACCGGCA
600	8mer	0.03909	CATGTACT
601	8mer	0.03906	CCGCAAAA
602	8mer	0.03904	AACTGTAC
603	8mer	0.039	GCAGAAAA
604	8mer	0.039	CACGCAGG
605	8mer	0.039	GCATCGTC
606	8mer	0.03898	GAGGGCGA
607	8mer	0.03897	GGTGAACG
608	8mer	0.03894	CGGATGTC
609	8mer	0.03894	GCGGAGCG
610	8mer	0.03893	AAGCCGCA
611	8mer	0.03885	TGACGCGA
612	8mer	0.03879	CCACCGAT
613	8mer	0.03875	TGCGTGAG
614	8mer	0.03874	TCATTGCC
615	8mer	0.03869	AGCGAGAC
616	8mer	0.03869	TGTTGTTAC
617	8mer	0.03868	CCGTAGCT
618	8mer	0.03866	ACATCCGT
619	8mer	0.03865	CCTCGTCT
620	8mer	0.03865	CTGCGCCG
621	8mer	0.03862	CTATGGGA
622	8mer	0.03862	TGCAACCT
623	8mer	0.03861	GATTAGAG

882	8mer	0.033872	GATTTTAA
883	8mer	0.033869	GCACCCCT
884	8mer	0.033854	CCGTAGAC
885	8mer	0.033778	TCGTGACA
886	8mer	0.033767	GTCCTCTG
887	8mer	0.033762	CGAATTAC
888	8mer	0.03376	GTGCCATA
889	8mer	0.033714	ACTGCGAT
890	8mer	0.033696	GCACCCAG
891	8mer	0.033673	CCGCCGTT
892	8mer	0.033602	CGCAGCGA
893	8mer	0.033572	TGTTACAT
894	8mer	0.033562	TGACGGGG
895	8mer	0.033561	AATACGGA
896	8mer	0.033552	ATAGGTAC
897	8mer	0.033546	TTTGACAG
898	8mer	0.033524	AACCCGCG
899	8mer	0.03352	CGTATGTG
900	8mer	0.033512	CTGCTGTT
901	8mer	0.033509	AATCACAA
902	8mer	0.033465	TCGGGACG
903	8mer	0.033421	AGCGAAAT
904	8mer	0.03342	GCCCTGGC
905	8mer	0.033374	CTCGGTTT
906	8mer	0.033372	AACGGGAC
907	8mer	0.033367	GGTCGCCA
908	8mer	0.033338	CGAATCCC
909	8mer	0.033335	GGCGGGCA
910	8mer	0.033317	GTGCGGTA
911	8mer	0.033292	GAATGCAC
912	8mer	0.033279	CTGTCCGT
913	8mer	0.033276	CGGCCGCG
914	8mer	0.033224	CGTCACTT
915	8mer	0.033193	GCCGTGTC
916	8mer	0.033192	CATCATT
917	8mer	0.033165	TCGGCGCT
918	8mer	0.033164	GGGATATC
919	8mer	0.033131	CCTATATC
920	8mer	0.033087	CAGCGAAG
921	8mer	0.033084	AGGGGCGG
922	8mer	0.033055	GCCGCCCG
923	8mer	0.033036	GGCCGAAT
924	8mer	0.033024	TCGTTGAA
925	8mer	0.033004	GCTCACAC
926	8mer	0.032975	CCGAGTGC
927	8mer	0.032933	CCGCCGCC
928	8mer	0.032904	GACCTCCC
929	8mer	0.032903	GCCGATCG
930	8mer	0.032903	CGACGAAG
931	8mer	0.032903	CCGTATAC
932	8mer	0.032903	ACTCGCGC
933	8mer	0.032903	CCGCTATA
934	8mer	0.032903	CGGTGCTA
935	8mer	0.032903	TTTAGCGC
936	8mer	0.032903	CACGAACG
937	8mer	0.032903	CCGTATCG
938	8mer	0.032903	CTTACGCG
939	8mer	0.032903	CGGTAAAT
940	8mer	0.032903	CGGCTAAC
941	8mer	0.032903	CGGGGTAA
942	8mer	0.032903	CGCGTTGA
943	8mer	0.032903	CGACGATA
944	8mer	0.032903	ATCGGCTA
945	8mer	0.032903	ACGTGCTC
946	8mer	0.032903	CGCTTATG
947	8mer	0.032903	TTCGGTTG
948	8mer	0.032903	ACGACGAT
949	8mer	0.032903	ATCGCGCA
950	8mer	0.032903	TCGTACAC
951	8mer	0.032903	CCGACGAC
952	8mer	0.032903	GGTGCGAC
953	8mer	0.032903	CGTTCCGC
954	8mer	0.032903	TCGAGACG
955	8mer	0.032903	ACGCGTCG
956	8mer	0.032903	CACCCGAT
957	8mer	0.032903	CCGGATTA
958	8mer	0.032903	GTCGGAGT
959	8mer	0.032903	ATAGCGCC
960	8mer	0.032903	TATAGCGC
961	8mer	0.032903	ACGAGCCG
962	8mer	0.032903	CGACTCGA
963	8mer	0.032903	TACGGCGC
964	8mer	0.032903	ACGTAAAT
965	8mer	0.032903	ACCGACGG
966	8mer	0.032903	GTGCGGAC
967	8mer	0.032903	ACGAGTCA
968	8mer	0.032903	CTATCACG
969	8mer	0.032903	ATCTCGAA
970	8mer	0.032903	CGATCCGA
971	8mer	0.032903	CCCGGTTA
972	8mer	0.032903	TGGGAAAC
973	8mer	0.032903	TGCCGCTA

274	8mer	0.048983	GCCAGAGT	624	8mer	0.0386	TCGCCGCC	974	8mer	0.032903	ATCGGTCC
275	8mer	0.0489	GGCGCCCA	625	8mer	0.03858	AGGATGTT	975	8mer	0.032903	CGTACCAA
276	8mer	0.048871	AGCCGCAT	626	8mer	0.03857	CCGCGGAT	976	8mer	0.032903	ATCCCGTA
277	8mer	0.048848	CTCACGGG	627	8mer	0.03857	ATGAGCTA	977	8mer	0.032903	CCCGTACC
278	8mer	0.048843	CTTCCCTA	628	8mer	0.03857	ACTCGTCA	978	8mer	0.032903	CGGAACCC
279	8mer	0.048756	GATCGTCG	629	8mer	0.03856	GCGGAGGT	979	8mer	0.032903	ATCGTCCG
280	8mer	0.048723	GGCTATTG	630	8mer	0.03854	ATAACGAG	980	8mer	0.032903	TTTGTCCG
281	8mer	0.04869	GGGACTGA	631	8mer	0.03854	GAATATCT	981	8mer	0.032903	TCGCGCTA
282	8mer	0.048651	TGAACACG	632	8mer	0.03853	GACACGGG	982	8mer	0.032903	TACCCGCA
283	8mer	0.048619	GGGCGGGG	633	8mer	0.03851	AACATGGT	983	8mer	0.032903	CGACGGAC
284	8mer	0.048591	ACTGTGGG	634	8mer	0.03848	ATAGTGCG	984	8mer	0.032903	ACTCGCCG
285	8mer	0.048586	GACCCGCA	635	8mer	0.03845	GACGGGCG	985	8mer	0.032903	CGGCACGA
286	8mer	0.048551	AGACCTGG	636	8mer	0.03845	TGCAAGTT	986	8mer	0.032903	TGCGACGA
287	8mer	0.048515	GACTCGCG	637	8mer	0.03836	CGCGGATT	987	8mer	0.032903	TGTGCGAC
288	8mer	0.048481	CGCCGCCG	638	8mer	0.03836	GCCCGGAG	988	8mer	0.032903	CGCGGATC
289	8mer	0.048465	CGGGAGCG	639	8mer	0.03834	ACGTCCGG	989	8mer	0.032903	GTCGGCGG
290	8mer	0.048459	CGCCCGGC	640	8mer	0.03833	TAGATCGT	990	8mer	0.032903	CGAGGCGT
291	8mer	0.048409	GGCGCGGC	641	8mer	0.0383	GCGGAGTT	991	8mer	0.032869	CGGGGTGG
292	8mer	0.048382	ATCGTCGT	642	8mer	0.03828	ACGCCACA	992	8mer	0.032866	CCCCGAAG
293	8mer	0.048319	GTTGTACG	643	8mer	0.03827	ATGGGACC	993	8mer	0.032861	CTCACTGC
294	8mer	0.048104	GAACGGCC	644	8mer	0.03825	CGTCATGA	994	8mer	0.032859	GGACGACT
295	8mer	0.048021	GTGGGAGG	645	8mer	0.03825	CATCCGTC	995	8mer	0.032837	AAACCCGT
296	8mer	0.048009	AGGTCCCT	646	8mer	0.03825	GAGGGCGC	996	8mer	0.03282	GACCATTA
297	8mer	0.047939	CGGCCATC	647	8mer	0.03824	TGATTGCG	997	8mer	0.03279	GGTTTCGC
298	8mer	0.047913	CGAGACTG	648	8mer	0.03823	TGCGCTCC	998	8mer	0.032776	AGTCGGGA
299	8mer	0.047904	CCCGAAGG	649	8mer	0.03823	GGGCTCAC	999	8mer	0.032775	TGCCCGTC
300	8mer	0.047877	GTTCTGCA	650	8mer	0.03819	GCACGTGG	1000	Triplet	0.032742	Triplet
301	8mer	0.047858	CCTGTCGG	651	8mer	0.03817	GGTGTCTG				
302	8mer	0.047854	CGGCGGGG	652	8mer	0.03814	CGGTGCGG				
303	8mer	0.047816	ATCGTAGA	653	8mer	0.03812	CTGACGCG				
304	8mer	0.047784	CTCGTCAA	654	8mer	0.03805	TTAAATGA				
305	Triplet	0.047765	Triplet	655	8mer	0.03805	TCACCCGG				
306	8mer	0.047735	AGTCGGAG	656	8mer	0.03802	CTCACATT				
307	8mer	0.047729	CCTCTCAT	657	8mer	0.03795	TGCGCCGA				
308	8mer	0.047708	TGCCCCGA	658	8mer	0.03794	GCAAACGT				
309	8mer	0.047676	ACTTCGAG	659	8mer	0.03794	CCGGGAAC				
310	8mer	0.04766	TGCGGGGG	660	8mer	0.03793	TGGCCAGG				
311	8mer	0.047656	GCCCGAAC	661	8mer	0.03789	CGAATGAT				
312	8mer	0.047521	TGGCGGCC	662	8mer	0.03787	ACGCGACA				
313	8mer	0.047505	ACATGGCA	663	8mer	0.03787	GACGCGAC				
314	Triplet	0.047503	Triplet	664	8mer	0.03787	TGTCGACC				
315	8mer	0.047497	TGCACGCA	665	8mer	0.03787	TAAACCGT				
316	8mer	0.047407	CGCATTCC	666	8mer	0.03787	GCTCGCGT				
317	8mer	0.047393	CACCGGAC	667	8mer	0.03786	CGGCCTCC				
318	8mer	0.047279	CGCTTGAA	668	8mer	0.03783	GACGCTTC				
319	8mer	0.047215	GAGATCCC	669	8mer	0.03783	TTCAATGA				
320	8mer	0.047169	GGTCGAGA	670	8mer	0.03782	CTCGTCGA				
321	8mer	0.047111	AAGTCGGG	671	8mer	0.03776	GGCCGTCT				
322	8mer	0.04707	CGACCCGG	672	8mer	0.03776	TGGGGTAT				
323	8mer	0.047052	GCGCCGTG	673	8mer	0.03776	TGTCGCCG				
324	8mer	0.046985	TGCGTCAA	674	8mer	0.03775	CCTTCACT				
325	8mer	0.04697	TGGACAGC	675	8mer	0.03775	ATATGTGA				
326	8mer	0.046957	TCGTCAAC	676	8mer	0.03774	CCGAACCT				
327	8mer	0.0469	CACGGGAG	677	8mer	0.03768	TGGTTGCT				
328	8mer	0.046795	ACGACCCG	678	8mer	0.03768	CTGCTCAC				
329	8mer	0.046768	TATCCAGT	679	8mer	0.03764	CAGGCGGG				
330	8mer	0.046766	GGTCATCG	680	8mer	0.03761	GCCCGTCG				
331	8mer	0.046758	ACGCAGGG	681	8mer	0.03758	GCGGGCAT				
332	8mer	0.046711	CGCGCCGT	682	8mer	0.03757	CACGAGAC				
333	8mer	0.046694	CTCCCAAA	683	8mer	0.03757	GCAATGCC				
334	8mer	0.046656	CGGGGCGG	684	8mer	0.03756	CTGCTCGG				
335	8mer	0.046639	TTATGTCG	685	8mer	0.03756	CCGAGGGG				
336	8mer	0.046637	GCGCAGCA	686	8mer	0.03754	AGGTCGTC				
337	8mer	0.046537	GACCGCGA	687	8mer	0.03752	GGCTGGTC				
338	8mer	0.046483	AGAAGCGG	688	8mer	0.03751	GAAAGCGA				
339	8mer	0.046434	ATCGCCGA	689	8mer	0.03751	GGGGCTGA				
340	8mer	0.04641	TGTCCGGT	690	8mer	0.0375	GGGTGCTA				
341	8mer	0.046361	AGCAGTCC	691	8mer	0.0375	CTTCATCG				
342	8mer	0.046282	AGCTCGGC	692	8mer	0.03748	CGTGGGTC				
343	8mer	0.046234	TCTGGTAT	693	8mer	0.03746	ATTAGCCG				
344	8mer	0.046221	CGGGCACT	694	8mer	0.03743	GCTGCCCG				
345	8mer	0.046201	GTTAATTT	695	8mer	0.0374	AAGCCCGG				
346	8mer	0.046144	CAGGACGA	696	8mer	0.03739	GGGGGCGG				
347	8mer	0.046108	GGTCGGGC	697	8mer	0.03736	CCCGTCTG				
348	8mer	0.04606	CGGGGCCA	698	8mer	0.03736	CACGGGAT				
349	8mer	0.046033	CGCCCGAG	699	8mer	0.03732	ACCCGAAT				
350	8mer	0.045988	TCGCCCTC	700	8mer	0.03731	TCGCCCGC				

TableS2 The top 1000 features in the optimal feature set of the IncLocPredtp58 according to F-score order

Order	Feature	F-score	kmer	Order	Feature	F-score	kmer	Order	Feature	F-score	kmer
1	5mer	0.118544	GGGGG	351	8mer	0.04707	CGACCCGG	701	8mer	0.037947	TGCGCCGA
2	8mer	0.10931	TCCCAAAG	352	8mer	0.04705	GCGCCGTG	702	8mer	0.037943	GCAAACGT
3	8mer	0.107445	GCGGGGCG	353	8mer	0.04699	TGCGTCAA	703	8mer	0.037942	CCGGGAAC
4	8mer	0.106471	AACGGCCC	354	8mer	0.04697	TGGACAGC	704	8mer	0.037932	TGGCCAGG
5	8mer	0.105616	CGGTGACG	355	8mer	0.04696	TCGTCAAC	705	8mer	0.037889	CGAATGAT
6	8mer	0.098194	CGGGTCAC	356	8mer	0.0469	CACGGGAG	706	8mer	0.037866	ACGCGACA
7	8mer	0.095685	TCACGGGA	357	8mer	0.0468	ACGACCCG	707	8mer	0.037866	GACGCGAC
8	8mer	0.09567	TGCGGTTT	358	8mer	0.04677	TATCCAGT	708	8mer	0.037866	TGTCGACC
9	8mer	0.09254	GGGGGACC	359	8mer	0.04677	GGTCATCG	709	8mer	0.037866	TAAACCCG
10	8mer	0.091784	CCACGACC	360	5mer	0.04676	CCGCA	710	8mer	0.037866	GCTCGCGT

11	8mer	0.091036	CGCCCCGG
12	5mer	0.090816	GCCCG
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	AATCGAGC
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	CCCAAAAGT
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	CTTGATATC
37	8mer	0.07546	CTTAAACC
38	8mer	0.075416	GAGCCCGA
39	8mer	0.07508	CGTAGCTT
40	5mer	0.074635	CCCCG
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
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	GTGACCGC
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	TCCGGGGC
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	TCCGCGCG
79	8mer	0.066489	GGCCCGTG
80	8mer	0.066467	GCGGATCT
81	8mer	0.06636	GGGCCCTT
82	8mer	0.06628	CCTGGGGC
83	8mer	0.066234	GACGAATC
84	8mer	0.066039	AACGTCCG
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	CCTGGGGC
91	8mer	0.065523	TTGTATCC
92	8mer	0.065484	TGTTCTCG
93	8mer	0.065447	CCACTGCA
94	8mer	0.065364	CGACAGGA
95	8mer	0.065247	CATGACAC
96	8mer	0.065246	ATCCGCGG
97	5mer	0.065211	GCGGG
98	8mer	0.064857	TGTCGGTA
99	8mer	0.064394	CTCATTTG
100	8mer	0.06426	GTAACCGC
101	8mer	0.064055	CCGGTTGA
102	5mer	0.063847	CCCCG

361	8mer	0.04676	ACGCAGGG
362	8mer	0.04671	CGCGCCGT
363	8mer	0.04669	CTCCCAAA
364	8mer	0.04666	CGGGGCGG
365	8mer	0.04664	TTATGTCG
366	8mer	0.04664	GCGCAGCA
367	8mer	0.04654	GACCGCGA
368	8mer	0.04648	AGAAGCGG
369	5mer	0.04647	AGGGC
370	8mer	0.04643	ATCGCCGA
371	8mer	0.04641	TGTCCGGT
372	8mer	0.04636	AGCAGTCC
373	8mer	0.04628	AGCTCGGC
374	8mer	0.04623	TCTGGTAT
375	8mer	0.04622	CGGGCACT
376	8mer	0.0462	GTTAATTT
377	8mer	0.04614	CAGGACGA
378	8mer	0.04611	GGTCGGGC
379	8mer	0.04606	CGGGGCCA
380	8mer	0.04603	CGCCCGAG
381	8mer	0.04599	TCGCCCTC
382	8mer	0.04596	GGCGGGGG
383	8mer	0.04584	CCCTGGGC
384	8mer	0.04582	CGCCGGAA
385	8mer	0.04581	ATCTTAAT
386	8mer	0.04574	CCGCGAAC
387	8mer	0.0456	TAAGGGCT
388	8mer	0.04557	CGACCTCC
389	8mer	0.04555	ACGCTCGT
390	8mer	0.04554	GCGGGGCA
391	8mer	0.04554	GCGCCAAT
392	Triplet	0.04548	Triplet
393	8mer	0.04545	GACCCGGC
394	8mer	0.04545	GGCTGATG
395	8mer	0.0454	GCACAAAT
396	5mer	0.04534	CGGCC
397	8mer	0.04529	GGTTGCTA
398	8mer	0.04526	TGTTTACG
399	8mer	0.04524	AGGGCGCA
400	8mer	0.04523	TAGGCATT
401	8mer	0.04521	CCGAAGTC
402	8mer	0.04521	CGTACATT
403	8mer	0.0452	CGCGTACA
404	8mer	0.04518	CCATCGGT
405	8mer	0.04515	CGAGACCA
406	8mer	0.04515	CCGGACTT
407	8mer	0.04512	ATGTCGGC
408	8mer	0.04507	CTAGCGGA
409	8mer	0.04506	TCGACGAA
410	8mer	0.04506	GGGCGCGG
411	8mer	0.04499	CCCCTTAT
412	8mer	0.04495	GGCAGAAA
413	8mer	0.04493	GTTGCTAT
414	8mer	0.04492	GGTTGGTC
415	8mer	0.04491	ACCGTGCC
416	8mer	0.04484	ACGCCTCG
417	8mer	0.04481	CGTGTTC
418	8mer	0.0448	ATGCGCGA
419	8mer	0.04471	CTTTCGTC
420	SC-PseDNC	0.04466	SC-PseDNC
421	8mer	0.04457	GGGGCGCG
422	8mer	0.0445	ACCGCGAA
423	8mer	0.04444	GGGGATAT
424	8mer	0.0444	TTTGTCTC
425	8mer	0.0444	AGATTGCG
426	8mer	0.04432	CGAGAATC
427	5mer	0.0443	GGCGT
428	8mer	0.04426	CAGTTCCG
429	8mer	0.04423	ATCGACGA
430	8mer	0.04422	CACGGCAA
431	5mer	0.0442	TCGCG
432	8mer	0.04419	ACCCCCGA
433	8mer	0.04418	CGAGAGCT
434	8mer	0.04415	GCGACTCG
435	8mer	0.0441	CCGTGGT
436	8mer	0.04408	ACTGCTCA
437	8mer	0.04407	GAGCCCCG
438	8mer	0.04403	CCAAAGTG
439	8mer	0.04392	CCGACGCG
440	8mer	0.0439	CACCGCGC
441	8mer	0.04389	GCCGCCCG
442	8mer	0.04387	GCCCTGTA
443	8mer	0.04382	CAAAGTAC
444	8mer	0.04381	GGAGGGCG
445	8mer	0.04381	GGGGGTCG
446	8mer	0.04379	GCGGACT
447	8mer	0.04379	GCGCCGGA
448	8mer	0.04377	GCAACGCT
449	8mer	0.04377	ACAGTAGC
450	8mer	0.04375	TTTTTGTA
451	8mer	0.04373	AGATACTA
452	8mer	0.04372	AACTCGTC

711	8mer	0.037861	CGGCCTCC
712	8mer	0.037833	GACGCCTC
713	8mer	0.037833	TTCAATGA
714	8mer	0.037822	CTCGTCGA
715	8mer	0.037764	GGCCGTCC
716	8mer	0.037759	TGGGGTAT
717	8mer	0.037755	TGTCCCGC
718	8mer	0.03775	CCTTCACT
719	8mer	0.037746	ATATGTGA
720	8mer	0.037736	CCGAACCT
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	AGGTCTGT
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	CTGGGGTG
739	8mer	0.03746	ATTAGCCG
740	8mer	0.037429	GCTGCCGG
741	8mer	0.037404	AAGCCGGC
742	8mer	0.037394	GGGGCGCG
743	5mer	0.037369	CAGCG
744	8mer	0.037359	CCCGTCTG
745	8mer	0.037355	CACGGGAT
746	8mer	0.037317	ACCCGAAT
747	8mer	0.037314	TCGCCCCG
748	8mer	0.037312	GCCTATTG
749	8mer	0.03727	AATACCGG
750	8mer	0.037266	CCCCGCCC
751	8mer	0.037257	CGGCTCAT
752	8mer	0.037252	GCTGCGCA
753	8mer	0.037222	CCCGGCAC
754	8mer	0.037219	ACCCGGGG
755	8mer	0.037189	AAGCGAAA
756	8mer	0.037161	GAATCGCT
757	8mer	0.037139	GCGCCCGG
758	8mer	0.037131	ACCCGGCC
759	8mer	0.037086	ATTAGGGA
760	8mer	0.037082	CCGCTTAG
761	8mer	0.037036	GGCATTCG
762	8mer	0.037023	GGCATTCC
763	8mer	0.036989	TAAAGTGG
764	8mer	0.036988	TCCTTTCC
765	8mer	0.036971	CCTCCCAA
766	8mer	0.036962	CGGAATAT
767	8mer	0.036961	TTGGGCCC
768	8mer	0.036937	CGCCAATC
769	8mer	0.036924	TGACCTTA
770	8mer	0.036905	GCGGAATT
771	8mer	0.036883	CTTCAAGC
772	8mer	0.036856	AGGCTGGT
773	8mer	0.036827	GCGCGCGG
774	8mer	0.036785	GTGCCGAA
775	8mer	0.036765	TTTCCCGA
776	8mer	0.036755	GAATCCCG
777	8mer	0.036745	CCGACTTT
778	8mer	0.036719	GATGTGGG
779	5mer	0.036696	GAGGG
780	8mer	0.036678	CAAAAAAA
781	8mer	0.036676	GCGACATC
782	8mer	0.036655	ACGGGTTT
783	8mer	0.036648	TTCAAGGT
784	8mer	0.036629	TGTACGTT
785	8mer	0.036619	GTGACCCC
786	8mer	0.036608	GCGGCACT
787	8mer	0.036604	CTGGTATA
788	8mer	0.036599	CGTCAAGG
789	8mer	0.036577	GACTGCGA
790	8mer	0.036576	CAGTAGGA
791	5mer	0.036576	CGGGA
792	8mer	0.036575	CGGGCCAC
793	8mer	0.036574	GAAGTCCG
794	8mer	0.036528	ATGCACGC
795	8mer	0.036518	TAACGAGG
796	8mer	0.036479	TTCCGGGG
797	8mer	0.036471	CGGGTGGC
798	8mer	0.036463	CAATATAC
799	8mer	0.036425	TCCGGTGT
800	8mer	0.03639	CTATGTTG
801	8mer	0.036376	ACGCTCAT
802	8mer	0.036373	GCGTTCTC

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
114	8mer	0.06263	AGGTGAG
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	CGCCG
124	8mer	0.061771	GGCCCTCT
125	8mer	0.061704	CGGGGCGA
126	8mer	0.061609	TCGTCCAA
127	8mer	0.061468	CGGGCGCT
128	8mer	0.061418	CGATGATA
129	8mer	0.06132	GGGCGAGG
130	8mer	0.061146	TCCGTACT
131	8mer	0.061133	CGGTGTGC
132	8mer	0.061067	GGGTTCGA
133	5mer	0.061032	GCCCG
134	8mer	0.060888	TCAAGCGA
135	8mer	0.060681	GGTCTGG
136	8mer	0.06066	GGAGGGGT
137	8mer	0.060619	CCTTCCGC
138	8mer	0.060596	ACTCCCGG
139	8mer	0.060568	CAAGCGAT
140	8mer	0.060416	AGGGGATA
141	8mer	0.060304	CCTCGGCC
142	8mer	0.060137	GCATTCCG
143	8mer	0.060047	GGGCGCCC
144	8mer	0.059996	TTGGCGCG
145	8mer	0.059859	ATGCGGAT
146	8mer	0.059776	GGGCGCTT
147	8mer	0.059758	GCTTCGGC
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	AAACGTGC
161	8mer	0.058738	AGATCCCG
162	8mer	0.058588	TCCGTAGC
163	8mer	0.05853	GGGCGAGC
164	8mer	0.058336	ACGCGGAG
165	8mer	0.058097	TAGGACAA
166	8mer	0.057951	GCCTTGGG
167	5mer	0.057679	GGCCG
168	8mer	0.057646	AATCCCGG
169	8mer	0.057561	CGCCGTGA
170	8mer	0.057561	GTAACCCG
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	AATCCGGC
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

453	8mer	0.04371	GGAACCTA
454	8mer	0.04367	AAGGTGCC
455	8mer	0.04364	TATTAGGG
456	8mer	0.04363	CCCGAACT
457	8mer	0.04362	TTGTACGG
458	8mer	0.04358	CCGTGTCC
459	8mer	0.04352	CGCCTATA
460	8mer	0.04351	TATGTCCG
461	8mer	0.0435	TACTGCCA
462	8mer	0.04347	TATAGCAC
463	8mer	0.04347	GGTTGCAG
464	8mer	0.04346	TTTGGCCA
465	8mer	0.04342	CCGCGATT
466	8mer	0.04336	GGTTGTTT
467	8mer	0.04334	GAGCGGGG
468	8mer	0.04333	GAGGCGGT
469	8mer	0.04332	GTGTCGTA
470	8mer	0.0433	CTGCGAGT
471	8mer	0.0433	CGGCCCTT
472	8mer	0.04328	GCAGTTCG
473	8mer	0.04326	TCCTGTGT
474	8mer	0.04325	GGGCGGTG
475	8mer	0.04325	CCGTTTCC
476	8mer	0.0432	TCGTCCGA
477	8mer	0.04319	TTGCGCCA
478	8mer	0.04314	AGCATTGC
479	8mer	0.04313	ACGCGTAC
480	8mer	0.04313	CTTCCGGG
481	5mer	0.04312	CCCCG
482	8mer	0.0431	GGGCGGTG
483	8mer	0.04309	ACCCGCAC
484	8mer	0.04305	GAGCCCGG
485	8mer	0.04305	ATCCCGAG
486	8mer	0.04304	GAAACGAT
487	8mer	0.04297	CAGTACGT
488	8mer	0.04296	CTTAAGTA
489	8mer	0.04296	GAGATACT
490	8mer	0.04295	CGTCGTCT
491	8mer	0.0429	GACGCGTA
492	8mer	0.04287	GCACTCCA
493	8mer	0.04286	CGCGCGGC
494	8mer	0.04285	TGCTTTTT
495	8mer	0.04278	CCGGACTA
496	8mer	0.04275	ATTGCGCC
497	8mer	0.04274	CGAGGTCT
498	8mer	0.04273	CAGGCCAC
499	8mer	0.04269	GATCCGGG
500	8mer	0.04266	TCCTCCGT
501	8mer	0.04263	CTAAACAT
502	5mer	0.04262	ACGGG
503	8mer	0.04261	CGGCGGAG
504	8mer	0.04258	CACAGATG
505	8mer	0.04258	GTAAGGGC
506	8mer	0.04257	GGACGCAA
507	8mer	0.04255	CCTGCGCC
508	8mer	0.04255	GGCGGTGT
509	8mer	0.04248	TGCACTTG
510	8mer	0.04246	AGGAGGCC
511	Triplet	0.04242	Triplet
512	8mer	0.04237	CCGAAGGG
513	8mer	0.04234	TGAGCCAC
514	8mer	0.04231	TCGGGGGT
515	8mer	0.04227	CCCGGGAG
516	SC-PseDNC	0.04225	SC-PseDNC
517	8mer	0.04224	CAACGTCT
518	5mer	0.04222	CGAGG
519	8mer	0.04222	CCGAAGAC
520	5mer	0.04221	CTCGG
521	8mer	0.04217	TCGAGAGC
522	8mer	0.04215	ACGGCCCA
523	8mer	0.04214	CCGTGGAA
524	8mer	0.04213	AACTGCGG
525	8mer	0.04207	CCAGGCAC
526	8mer	0.04206	AGGTCGTA
527	8mer	0.04206	GCCCGGCA
528	8mer	0.04206	AAGCGAGG
529	8mer	0.04204	TGCTCGCG
530	8mer	0.04204	CGGAGGTT
531	8mer	0.04194	GTTCCCAT
532	8mer	0.04194	AGACGGTC
533	8mer	0.0419	CCATAGTG
534	8mer	0.04188	CGAAGGCG
535	8mer	0.04183	GCCGCCAG
536	8mer	0.04182	AGCTTTTG
537	8mer	0.04181	CCCCCCCC
538	8mer	0.04181	TCGACCTC
539	8mer	0.04173	CCCGGAGG
540	8mer	0.0417	CACCACTT
541	8mer	0.0417	GACCTCGG
542	Triplet	0.04159	Triplet
543	8mer	0.04158	TCCGGTTG
544	8mer	0.04153	CGGAAACG

803	8mer	0.03628	GACGCGCT
804	8mer	0.036274	GAGGTCAT
805	8mer	0.036257	ACGGCCGT
806	8mer	0.036192	GTTATCTG
807	8mer	0.03619	CTACAGTA
808	8mer	0.036189	TATAAGGT
809	8mer	0.036163	CGACTTTG
810	8mer	0.036153	TGAGCCGA
811	8mer	0.036139	CGTAGATA
812	8mer	0.036137	CTAAACCG
813	8mer	0.036098	CTTGCGTC
814	8mer	0.036073	AATTATCC
815	8mer	0.036068	GAGCGACG
816	8mer	0.036049	GTCCAGAC
817	8mer	0.036048	CGGGGGCC
818	8mer	0.036041	ACGCCCGA
819	5mer	0.036039	GGCCC
820	8mer	0.036032	GGCGCCCC
821	8mer	0.036022	GTTTCGCC
822	8mer	0.036005	TCCTGGGG
823	8mer	0.036004	TTGTTAAT
824	8mer	0.03599	CAGCCTCC
825	8mer	0.035982	GGTGCGCG
826	8mer	0.035958	CGTTACGG
827	8mer	0.035936	CGAGTTCT
828	8mer	0.03593	CGCGCGGG
829	5mer	0.035879	CGCGA
830	8mer	0.035852	ACCCGCGC
831	8mer	0.035833	ACCTCGAC
832	8mer	0.035761	CGAACAGA
833	8mer	0.035756	CTGTCCCG
834	8mer	0.035747	CTCCTGTG
835	8mer	0.035746	GGCTCACG
836	8mer	0.035744	GACCTGGG
837	8mer	0.035735	GAACCTGT
838	8mer	0.035724	ACTGCACG
839	8mer	0.035716	GAGGGCCG
840	8mer	0.03568	ACCATTA
841	8mer	0.035668	CAAGCGAA
842	8mer	0.035666	GCGGCCAA
843	8mer	0.035652	GCCATTCT
844	5mer	0.035632	ACCTC
845	8mer	0.035597	GTCTAGGC
846	8mer	0.035581	ATCATTAT
847	8mer	0.035577	TGGGCCCC
848	8mer	0.035562	TAACCGCG
849	8mer	0.035558	GGCTGGGA
850	8mer	0.035557	GTAACAT
851	8mer	0.035551	CGTCCAGG
852	8mer	0.035506	CGGACTAG
853	8mer	0.035482	GGACCTAG
854	8mer	0.035439	TAGTGTCC
855	8mer	0.035432	CACATACG
856	8mer	0.035424	CGCTTCTT
857	8mer	0.035384	TCCCAACG
858	8mer	0.035383	CCGGGGTG
859	8mer	0.035364	CGCGCGCG
860	8mer	0.03536	TGGGATTA
861	8mer	0.035351	AGCCGGGG
862	8mer	0.035334	TTCACTAG
863	8mer	0.03528	CCCTTCGA
864	8mer	0.03522	CTGCGCCT
865	8mer	0.03518	CTCGGCTG
866	8mer	0.035126	CACGCCAC
867	8mer	0.035113	CTCGGGTC
868	8mer	0.035093	CGGCGTGC
869	8mer	0.035069	CTGTCGAC
870	8mer	0.035065	TAGTGGGA
871	8mer	0.035051	CTACAGCT
872	8mer	0.03505	GGCCGCTC
873	8mer	0.035019	GGCCAGGT
874	8mer	0.035018	TTCACGCC
875	8mer	0.035013	CTGGCCAC
876	8mer	0.034942	CTGGGCTG
877	8mer	0.034939	GTCGGAAA
878	8mer	0.034897	GCACTTGT
879	8mer	0.034893	TCGCTTGA
880	5mer	0.034853	GCCGA
881	8mer	0.034842	ACTCAGTG
882	8mer	0.034833	GAACTTAA
883	8mer	0.034822	CCCCAGGC
884	8mer	0.034799	TCCCGCAA
885	8mer	0.034779	TCITTAAGT
886	5mer	0.034747	CGCAC
887	8mer	0.034736	TGTGGCCG
888	8mer	0.034735	CCCTGATT
889	8mer	0.03473	CGCAAAAG
890	8mer	0.034724	ACAAAGCA
891	8mer	0.034711	CGGGCGGT
892	8mer	0.034709	GAGGGCCT
893	8mer	0.034708	GGCGTCGG
894	8mer	0.034707	ACCAGGGG

195	8mer	0.056455	CGGGATTG
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	GTTCAATTG
212	8mer	0.055039	CGGCGATG
213	8mer	0.054965	TGAGTGTG
214	8mer	0.054878	TTCACAGC
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	CCGCG
222	8mer	0.054301	CGGTAACC
223	8mer	0.054267	TAGAGGGC
224	8mer	0.054134	CGGACTTT
225	5mer	0.05409	CCGAA
226	8mer	0.053955	GCCCCGTG
227	8mer	0.053902	AAAAATTAG
228	8mer	0.053867	CGAGCCCC
229	8mer	0.053538	GACCCCTC
230	8mer	0.053492	ATGCAACG
231	8mer	0.053471	TCCAATGC
232	8mer	0.05329	CCTCAGCC
233	8mer	0.053201	CCGCATTG
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	CGGCCCCC
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	GGGTGCGAG
257	8mer	0.051532	TCGAGTCG
258	8mer	0.051301	TCCAGAGT
259	8mer	0.051286	GAGCGCCA
260	8mer	0.051239	ATTGCTTG
261	8mer	0.051176	CAAGATCG
262	8mer	0.051123	GCGTTCAG
263	8mer	0.051068	GCTCTCTG
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	CCGCCACG
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	GGCGGGCG
278	8mer	0.050237	TGCCGGTC
279	8mer	0.0502	TAATACCG
280	8mer	0.050185	ATTACTGC
281	8mer	0.050111	CCTGAGCA
282	8mer	0.050078	GAGTGTAT
283	8mer	0.050071	GGCGTGCT
284	8mer	0.049934	GGAACCGG
285	8mer	0.049918	GCGGTTCG
286	5mer	0.049887	TCGGC

545	8mer	0.04145	ACAATGAG
546	8mer	0.04145	CCCTCGTC
547	8mer	0.04144	AGGTTCAA
548	8mer	0.04144	GCCCCCCC
549	8mer	0.04142	CCGGGCCC
550	8mer	0.04142	GGGAGATC
551	8mer	0.04138	TACTCGGG
552	8mer	0.04134	TACGTAAA
553	8mer	0.0413	TCGCGGGC
554	8mer	0.04129	CCCGTAIT
555	8mer	0.04124	GACCTGGC
556	8mer	0.04118	GCCCGAGG
557	8mer	0.04114	GCATTCCG
558	8mer	0.04113	CATTTCGAT
559	8mer	0.04111	GGGGCGGG
560	8mer	0.04108	AAGGGTCC
561	8mer	0.04108	TTGCGGAA
562	5mer	0.04104	GCGCG
563	8mer	0.04101	CTAATGCA
564	8mer	0.04095	CCCGTCTC
565	5mer	0.04091	GCCCC
566	8mer	0.04083	CATCGCCG
567	8mer	0.04082	CCGCGGGC
568	8mer	0.0408	GGGACCGC
569	8mer	0.0408	TAACGGGA
570	8mer	0.0408	CGCGATGA
571	8mer	0.0408	ACGCGCTA
572	8mer	0.0408	GCGCTAGG
573	8mer	0.0408	GTAACGGG
574	8mer	0.0408	GTCGTAAT
575	8mer	0.04079	CTGGGCCT
576	8mer	0.04073	CCTGGCGT
577	8mer	0.04071	TGGGACCT
578	8mer	0.04069	CGTAATAC
579	8mer	0.04064	GGCGTTCA
580	8mer	0.04057	TCACGCCA
581	8mer	0.04056	CGAACACC
582	8mer	0.04056	TGCTTACC
583	8mer	0.04048	AACGGGTT
584	8mer	0.04046	CACGCCTC
585	8mer	0.04046	GGGATCGC
586	8mer	0.0404	CACCCCCG
587	8mer	0.0404	CGAACGGG
588	8mer	0.0404	CGTCCAGC
589	8mer	0.04039	TGGCGTGC
590	8mer	0.04038	TGGGACTG
591	8mer	0.04037	CCCCTGAC
592	8mer	0.04037	GCCGGCTG
593	8mer	0.04033	TCCTACCT
594	8mer	0.04033	CGCTGGTA
595	8mer	0.04031	GGGCTAG
596	8mer	0.04031	CTGACGGC
597	8mer	0.04029	GGATTATA
598	8mer	0.04026	GCACCAAA
599	8mer	0.04025	GAACGTGC
600	8mer	0.04025	TTGGCCAG
601	8mer	0.04024	ACACGGGA
602	8mer	0.04022	AGCCGTTT
603	8mer	0.04021	TGCTCAGC
604	8mer	0.04019	AGGAGGGG
605	8mer	0.04019	GGAGCGCT
606	8mer	0.04017	CCAATGCC
607	8mer	0.04017	ATCGCGCC
608	8mer	0.04004	GGGTGTTT
609	8mer	0.04003	GCCACTCG
610	8mer	0.03999	GGGAGGCT
611	8mer	0.03999	TATTGAAG
612	8mer	0.03991	CGAATCAG
613	8mer	0.0399	CGCTAAAT
614	8mer	0.03987	TAATCGAG
615	8mer	0.03987	AACCCGAA
616	8mer	0.03985	CGCGAACG
617	8mer	0.03983	CTGTGGGG
618	8mer	0.03981	CCTCGCGG
619	8mer	0.03981	CGAGATGG
620	8mer	0.03976	GCCGGTCT
621	8mer	0.03963	CATGTAAT
622	8mer	0.0396	TCGCGTCC
623	8mer	0.03955	TCGATGAG
624	8mer	0.03955	GGAGCACT
625	8mer	0.03951	AACGCCTG
626	8mer	0.03949	CGCGCGTG
627	8mer	0.03941	GCGCCGAG
628	8mer	0.0394	CTCGAGAT
629	8mer	0.03938	CGAAATTA
630	8mer	0.03936	CAATGCCC
631	8mer	0.03936	TCGACCCG
632	5mer	0.03936	GGTCG
633	8mer	0.03927	TATACTCG
634	8mer	0.03922	GGCGCGTG
635	8mer	0.03921	GGTAGCGT
636	8mer	0.03919	GGTGCGAG

895	8mer	0.0347	AGGCACCT
896	8mer	0.034691	CTAATGGT
897	8mer	0.034684	ACCGCTGG
898	8mer	0.034665	CATGGCAT
899	8mer	0.034628	GCGTTGTT
900	8mer	0.034567	CACCGAAG
901	8mer	0.034557	CTGAAAAA
902	8mer	0.034523	CGCGGGAG
903	8mer	0.034485	CGGTTAAC
904	8mer	0.034454	CGGGAGTT
905	8mer	0.03441	CGCCGCCCT
906	8mer	0.034398	CCCTCACG
907	5mer	0.034398	GGCCT
908	8mer	0.034398	AGGGCGCG
909	5mer	0.034375	CGGCA
910	8mer	0.034374	CCCGGGCT
911	8mer	0.034353	AGGGGCGGT
912	8mer	0.034352	GGGGGACG
913	8mer	0.034348	CCGTATTA
914	8mer	0.034348	GGATTACA
915	8mer	0.034336	CCCTACCA
916	8mer	0.03431	GAAATTGA
917	8mer	0.034267	GTCGGACC
918	8mer	0.034215	TTTTGTAT
919	8mer	0.034207	TAACGGCAT
920	8mer	0.034187	TTGGAAAG
921	8mer	0.034168	GATGCCCC
922	8mer	0.034158	GTCGGCGG
923	8mer	0.034099	CCGGGGCG
924	8mer	0.034094	GAAGCGCG
925	8mer	0.034065	CGCAATGC
926	8mer	0.034064	TCCCTCTG
927	8mer	0.034041	GCCACAAA
928	8mer	0.034036	GCTGCTAG
929	8mer	0.034008	GGCGGCAC
930	8mer	0.034007	ACCTGGTC
931	8mer	0.033999	AACAAGCG
932	8mer	0.03399	TGCCTAAG
933	8mer	0.033985	TACAGCGC
934	8mer	0.033952	CTGGTCTC
935	8mer	0.033917	GGTCTCGA
936	8mer	0.033911	TATGATAC
937	8mer	0.0339	CCGCGCTT
938	8mer	0.033872	GATTTTAA
939	8mer	0.033869	GCACCCCT
940	8mer	0.033854	CCGTAGAC
941	8mer	0.033778	TCGTGACA
942	8mer	0.033767	GTCCTCTG
943	8mer	0.033762	CGAATTAC
944	8mer	0.03376	GTGCTCAA
945	5mer	0.033727	GCGTG
946	8mer	0.033714	ACTGCGAT
947	8mer	0.033696	GCACCCAG
948	8mer	0.033673	CCCGCGTT
949	8mer	0.033602	GCGAGCGA
950	8mer	0.033572	TGTTACAT
951	8mer	0.033562	TGACGGGG
952	8mer	0.033561	AATACGGA
953	8mer	0.033552	ATAGGTAC
954	8mer	0.033546	TTTGACAG
955	8mer	0.033524	AACCCGCG
956	8mer	0.03352	CGTATGTG
957	8mer	0.033512	CTGCTGTT
958	8mer	0.033509	AATCACAA
959	5mer	0.033471	CTGGG
960	8mer	0.033465	TCGGGACG
961	8mer	0.033421	AGCGAAAT
962	8mer	0.03342	GCCCTGGC
963	8mer	0.033374	CTCGGTTC
964	8mer	0.033372	AACGGACG
965	8mer	0.033367	GGTCGCCA
966	8mer	0.033338	CGAATCCC
967	8mer	0.033335	GGCGGGCA
968	8mer	0.033317	GTCGGGTA
969	8mer	0.033292	GAATGCAC
970	8mer	0.033279	CTGTCTGT
971	8mer	0.033276	CGGGCGGG
972	8mer	0.033224	CGTCACCT
973	8mer	0.033193	GCCGTGTC
974	8mer	0.033192	CATCATTA
975	8mer	0.033165	TCGGCCTC
976	8mer	0.033164	GGGATATC
977	8mer	0.033131	CCTATATC
978	8mer	0.033087	CAGCCAAG
979	8mer	0.033084	AGGGGCGG
980	8mer	0.033055	GCCGGCCG
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

287	8mer	0.049887	CCTCCGTA	637	8mer	0.03918	TTTTTAT	987	8mer	0.032933	CCGCCGCC
288	8mer	0.049846	GATCGCGC	638	8mer	0.03918	CCGCCAAT	988	8mer	0.032904	GACCTCCC
289	8mer	0.049818	TTGGTGCT	639	8mer	0.03915	TGGGGCGG	989	8mer	0.032903	GCCGATCG
290	8mer	0.049787	CGGAGGCG	640	8mer	0.03913	TACCGGCA	990	8mer	0.032903	CGACGAAG
291	8mer	0.049757	CGAGATCG	641	8mer	0.03909	CATGCTAC	991	8mer	0.032903	CCGTATAC
292	8mer	0.049721	TGAACCCG	642	8mer	0.03906	CCCCCAAA	992	8mer	0.032903	ACTCGCGC
293	8mer	0.049702	AATAGCCA	643	5mer	0.03904	AGCCG	993	8mer	0.032903	CCGTATATA
294	8mer	0.049633	AGCGGGGG	644	8mer	0.03904	AACTGTAC	994	8mer	0.032903	CGGTGCTA
295	8mer	0.049554	GGTGAACC	645	8mer	0.039	GCAGAAAA	995	8mer	0.032903	TTTAGCGC
296	8mer	0.049369	ATAAATGG	646	8mer	0.039	CACGCAGG	996	8mer	0.032903	CACGAACG
297	8mer	0.049266	CACTGCAC	647	8mer	0.039	GCATCGTC	997	8mer	0.032903	CCGTATCG
298	8mer	0.049199	TGCTGCG	648	8mer	0.03898	GAGGGCGA	998	8mer	0.032903	CTTACGCG
299	8mer	0.049125	CAAGGAGT	649	8mer	0.03897	GGTGAAC	999	8mer	0.032903	CGGTAATG
300	8mer	0.049102	CATTGCCG	650	8mer	0.03894	CGGATGTC	1000	8mer	0.032903	CGGCTAAC
301	5mer	0.049087	CCGGG	651	8mer	0.03894	GCGGAGCG				
302	8mer	0.048983	GCCAGAGT	652	8mer	0.03893	AAGCCGCA				
303	8mer	0.0489	GGCGCCCA	653	8mer	0.03885	TGACGCGA				
304	8mer	0.048871	AGCCGCAT	654	8mer	0.03879	CCACCGAT				
305	8mer	0.048848	CTCACGGG	655	8mer	0.03875	TGCGTGAG				
306	8mer	0.048843	CTTCCCTA	656	8mer	0.03874	TCATTGCC				
307	8mer	0.048756	GATCGTCG	657	8mer	0.03869	AGCGAGAC				
308	8mer	0.048723	GGCTATTG	658	8mer	0.03869	TGTGTTAC				
309	8mer	0.04869	GGGACTGA	659	8mer	0.03868	CCGTAGCT				
310	8mer	0.048651	TGAACACG	660	8mer	0.03866	ACATCCGT				
311	8mer	0.048619	GGGCGGGG	661	8mer	0.03865	CCTCGTCG				
312	8mer	0.048591	ACTGTGGG	662	8mer	0.03865	CTGCGCCG				
313	8mer	0.048586	GACCCGCA	663	8mer	0.03862	CTATGGGA				
314	8mer	0.048551	AGACCTGG	664	8mer	0.03862	TGCAACCT				
315	8mer	0.048515	GACTCGCG	665	8mer	0.03861	GATTAGAG				
316	8mer	0.048481	CGCCGCCG	666	8mer	0.0386	TCGCCGCC				
317	8mer	0.048465	CGGGAGCG	667	8mer	0.03858	AGGATGTT				
318	8mer	0.048459	CGCCCGGC	668	8mer	0.03857	CCGCGGAT				
319	8mer	0.048409	GGCGCGGC	669	8mer	0.03857	ATGAGCTA				
320	8mer	0.048382	ATCGTCGT	670	8mer	0.03857	ACTCGTCA				
321	8mer	0.048319	GTTGTACG	671	8mer	0.03856	GCGGAGGT				
322	8mer	0.048104	GAACGGCC	672	8mer	0.03854	ATAACGAG				
323	8mer	0.048021	GTGGGAGG	673	8mer	0.03854	GAATATCT				
324	5mer	0.048021	CCGAG	674	8mer	0.03853	GACACGGG				
325	8mer	0.048009	AGGTCCCT	675	8mer	0.03851	AACATGGT				
326	8mer	0.047939	CGGCCATC	676	8mer	0.03848	ATAGTGCG				
327	8mer	0.047913	CGAGACTG	677	8mer	0.03845	GACGGGCG				
328	8mer	0.047904	CCGAAGG	678	8mer	0.03845	TGCAAGTT				
329	8mer	0.047877	GTTCTGCA	679	8mer	0.03836	CGCGGATT				
330	8mer	0.047858	CCTGTCGG	680	8mer	0.03836	GCCCGGAG				
331	8mer	0.047854	CGGCGGGG	681	8mer	0.03834	ACGTCGCG				
332	8mer	0.047816	ATCGTAGA	682	8mer	0.03833	TAGATCGT				
333	8mer	0.047784	CTCGTCAA	683	5mer	0.03832	TGGTA				
334	Triplet	0.047765	Triplet	684	8mer	0.0383	GCGGAGTT				
335	8mer	0.047735	AGTCGGAG	685	8mer	0.03828	ACGCCACA				
336	8mer	0.047729	CCTCTCAT	686	8mer	0.03827	ATGGGACC				
337	8mer	0.047708	TGCCCCGA	687	8mer	0.03825	CGTCATGA				
338	8mer	0.047676	ACTTCGAG	688	8mer	0.03825	CATCCGTC				
339	8mer	0.04766	TGCGGGGG	689	8mer	0.03825	GAGGGCGC				
340	8mer	0.047656	GCCCGAAC	690	8mer	0.03824	TGATTTCG				
341	8mer	0.047521	TGGGCGCC	691	8mer	0.03823	TGCGCTCC				
342	8mer	0.047505	ACATGGCA	692	8mer	0.03823	GGGCTCAC				
343	Triplet	0.047503	Triplet	693	5mer	0.0382	GGCGG				
344	8mer	0.047497	TGCACGCA	694	8mer	0.03819	GCACGTGG				
345	8mer	0.047407	CGCATTC	695	8mer	0.03817	GGTGTCTG				
346	8mer	0.047393	CACCGGAC	696	8mer	0.03814	CGGCTGGG				
347	8mer	0.047279	CGCTTGAA	697	8mer	0.03812	CTGACGCG				
348	8mer	0.047215	GAGATCCC	698	8mer	0.03805	TAAATGA				
349	8mer	0.047169	GGTCGAGA	699	8mer	0.03805	TCACGCGC				
350	8mer	0.047111	AAGTCGGG	700	8mer	0.03802	CTCACATC				

TableS3 The top 1000 features in the optimal feature set of the IncLocPredtp68 according to F-score order

Order	Feature	F-score	kmer	Order	Feature	F-score	kmer	Order	Feature	F-score	kmer
1	8mer	0.10931	TCCCAAAG	351	8mer	0.04717	GGTCGAGA	701	8mer	0.038531	GACACGGG
2	8mer	0.107445	GCGGGGCG	352	8mer	0.04711	AAGTCGGG	702	6mer	0.038521	CCCCGC
3	8mer	0.106471	AACGGCCC	353	8mer	0.04707	CGACCCGG	703	8mer	0.038508	AACATGGT
4	8mer	0.105616	CGGTGACG	354	8mer	0.04705	GCGCGGTG	704	8mer	0.038481	ATAGTGCG
5	8mer	0.098194	CGGGTCAC	355	8mer	0.04699	TGCGTCAA	705	6mer	0.038469	GCGGAG
6	8mer	0.095685	TCACGGGA	356	8mer	0.04697	TGGACAGC	706	8mer	0.038451	GACGGGCG
7	8mer	0.09567	TGCGGTTT	357	8mer	0.04696	TCGTCAAC	707	8mer	0.038446	TGCAAGTT
8	8mer	0.09254	GGGGGACC	358	8mer	0.0469	CACGGGAG	708	8mer	0.038361	CGCGGATT
9	8mer	0.091784	CCACGACC	359	8mer	0.0468	ACGACCCG	709	8mer	0.03836	GCCCGGAG
10	8mer	0.091036	CGCCCCGG	360	8mer	0.04677	TATCCAGT	710	8mer	0.038343	ACGTCGCG
11	8mer	0.09051	TCAACGTC	361	8mer	0.04677	GGTCATCG	711	8mer	0.038332	TAGATCGT
12	8mer	0.088855	CGAAGTCG	362	8mer	0.04676	ACGCAGGG	712	8mer	0.038297	GCGGAGTT
13	8mer	0.088667	CGGGGGAC	363	8mer	0.04671	CGCGCCGT	713	8mer	0.038275	ACGCCACA
14	6mer	0.088041	GGGGCG	364	8mer	0.04669	CTCCCAAA	714	8mer	0.03827	ATGGGACC
15	8mer	0.08771	ATGGGGCG	365	8mer	0.04666	CGGGGCGG	715	8mer	0.038252	CGTCATGA
16	8mer	0.087693	TCGGCCCC	366	8mer	0.04664	TTATGTCT	716	8mer	0.038251	CATCCGTC
17	8mer	0.087595	AGCGGAGC	367	8mer	0.04664	GCGCAGCA	717	8mer	0.038249	GAGGGCGC
18	8mer	0.087322	AATCGACG	368	6mer	0.04658	CGGGCG	718	8mer	0.038244	TGATTTCG
19	8mer	0.085901	GCTCACGG	369	8mer	0.04654	GACCGCGA	719	8mer	0.038228	TGCGCTCC
20	6mer	0.084846	GCGGGG	370	8mer	0.04648	AGAAAGCG	720	8mer	0.038226	GGGCTCAC
21	8mer	0.084716	GTAGGGCG	371	6mer	0.04647	ACCCGC	721	8mer	0.038193	GCACGTGG
22	8mer	0.084641	CGCAGTTC	372	6mer	0.04646	CCGAGG	722	8mer	0.038168	GGTGTCTG
23	8mer	0.083424	CCGCAGTT	373	8mer	0.04643	ATCGCCGA	723	8mer	0.038135	CGGCTGGG



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	8mer	0.078261	GGGATGCG
30	8mer	0.077741	GATCCCGA
31	6mer	0.077289	GGGCGC
32	8mer	0.077287	GCGTGAGC
33	8mer	0.077177	CGGGCGGG
34	8mer	0.077057	GCGGGGGA
35	6mer	0.076967	GGGCCT
36	8mer	0.076646	AGACGGCG
37	8mer	0.075641	CTTGATATC
38	8mer	0.07546	CTTAAACC
39	8mer	0.075416	GAGCCCGA
40	8mer	0.07508	CGTAGCTT
41	8mer	0.074414	CCCCCGAA
42	6mer	0.073596	ATCCGC
43	8mer	0.073205	GGCGAGGT
44	8mer	0.073205	CCGCACCG
45	8mer	0.072737	CGCAGCGT
46	8mer	0.072567	CATCGCGG
47	8mer	0.072132	TGGGCCCC
48	6mer	0.072091	CGGGCG
49	8mer	0.072036	CCCGTTTC
50	8mer	0.072034	GTGCGTGA
51	8mer	0.071804	GGATGCCG
52	6mer	0.070845	CGTCCA
53	8mer	0.070518	AAGGGGCG
54	6mer	0.070402	AGCCGG
55	8mer	0.07038	CCGAAGGC
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	ACGTGCCG
63	8mer	0.068033	GTGCGACT
64	8mer	0.067987	GACCATCG
65	8mer	0.067781	ACGAATCA
66	8mer	0.067781	TCGTGCAC
67	8mer	0.067781	TAACCGCC
68	8mer	0.067781	GTGACACG
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	TCCGGGGG
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	GGCCCCGT
83	8mer	0.066467	GCGGATCT
84	8mer	0.06636	GGGGCCCT
85	8mer	0.06628	CCTGGGGC
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	CCTGGGGC
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	CCGGTTGA
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	CCCGACGG
110	8mer	0.062931	ATACCGCC
111	8mer	0.062913	TCCAGGCG
112	8mer	0.062893	TCCCGCAC
113	8mer	0.062656	ACCGGACT
114	8mer	0.06263	AGGTGCGA
115	8mer	0.062472	GAAACCCG

374	8mer	0.04641	TGTCCGGT
375	8mer	0.04636	AGCAGTCC
376	8mer	0.04628	AGCTCGGC
377	6mer	0.04625	GTCGTG
378	8mer	0.04623	TCTGGTAT
379	8mer	0.04622	CGGGCACT
380	8mer	0.0462	GTTAATTT
381	8mer	0.04614	CAGGACGA
382	8mer	0.04611	GGTCGGGC
383	6mer	0.04608	CGCCGG
384	8mer	0.04606	CGGGGCCA
385	8mer	0.04603	CGCCCGAG
386	8mer	0.04599	TCGCCCTC
387	8mer	0.04596	GGCGGGGC
388	6mer	0.04585	CGCCCG
389	8mer	0.04584	CCCTGGGC
390	8mer	0.04582	CGCCGGAA
391	8mer	0.04581	ATCTTAAT
392	6mer	0.04578	CGCGCT
393	8mer	0.04574	CCGCGAAC
394	8mer	0.0456	TAAGGGCT
395	8mer	0.04557	CGACCTCC
396	8mer	0.04555	ACGCTCGT
397	8mer	0.04554	CGCGGGCA
398	8mer	0.04554	CGCCCAAT
399	Triplet	0.04548	Triplet
400	8mer	0.04545	GACCCGGC
401	8mer	0.04545	GGCTGATG
402	8mer	0.0454	GCACAAAT
403	8mer	0.04529	GGTTGCTA
404	8mer	0.04526	TGTTIACG
405	8mer	0.04524	AGGGCGCA
406	8mer	0.04523	TAGGCATT
407	8mer	0.04521	CCGAAGTC
408	8mer	0.04521	CGTACATT
409	8mer	0.0452	CGCGTACA
410	8mer	0.04518	CCATCGGT
411	8mer	0.04515	CGGACACCA
412	8mer	0.04515	CCGGACTT
413	8mer	0.04512	ATGTCGGC
414	8mer	0.04507	CTAGCGGA
415	8mer	0.04506	TCCAGCAA
416	8mer	0.04506	GGGCGCGG
417	8mer	0.04499	CCCCTTAT
418	6mer	0.04497	TCGGCC
419	8mer	0.04495	GGCAGAAA
420	6mer	0.04493	CCCGAA
421	8mer	0.04493	GTTGCTAT
422	8mer	0.04492	GGTTGGTC
423	8mer	0.04491	ACCGTGCC
424	6mer	0.04489	CCCGCC
425	8mer	0.04484	ACGCCTCG
426	8mer	0.04481	CGTGTIAC
427	8mer	0.0448	ATGCGCGA
428	8mer	0.04471	CTTTCGTC
429	SC-PseDNC	0.04466	SC-PseDNC
430	8mer	0.04457	GGGGCGCG
431	6mer	0.04457	ACGCGA
432	8mer	0.0445	ACCGCGAA
433	8mer	0.04444	GGGGATAT
434	8mer	0.0444	TTTGCTCT
435	8mer	0.0444	AGATTGCG
436	8mer	0.04432	CGAGAATC
437	8mer	0.04426	CAGTTCGG
438	8mer	0.04423	ATCGACGA
439	8mer	0.04422	CACGGCAA
440	8mer	0.04419	ACCCCGCA
441	8mer	0.04418	CGAGAGCT
442	6mer	0.04416	CTGGGC
443	6mer	0.04416	GGCGGG
444	8mer	0.04415	GCGACTCG
445	8mer	0.0441	CCGCTGGT
446	8mer	0.04408	ACTGCTCA
447	8mer	0.04407	GAGCCCGC
448	8mer	0.04403	CCAAAGTG
449	6mer	0.04402	GGGCGA
450	6mer	0.04401	CGGGTA
451	8mer	0.04392	CCGCAGCG
452	8mer	0.0439	CACCGCGC
453	8mer	0.04389	GCCGCCGC
454	8mer	0.04387	GCCCTCGA
455	8mer	0.04382	CAAAAGTC
456	8mer	0.04381	GGAGGGCG
457	8mer	0.04381	GGGGGTCG
458	8mer	0.04379	GCCGCGAT
459	8mer	0.04379	GCGCGGGA
460	8mer	0.04377	GCAACGTC
461	8mer	0.04377	ACCTAGAC
462	6mer	0.04376	CGCGCC
463	8mer	0.04375	TTTTTGTA
464	8mer	0.04373	AGATACTA
465	8mer	0.04372	AACTCGTC

724	8mer	0.038124	CTGACGCG
725	8mer	0.03805	TTAAATGA
726	8mer	0.038048	TCACCGCG
727	8mer	0.038015	CTCAATATC
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	TAAACCCG
737	8mer	0.037866	GCTCGCGT
738	8mer	0.037861	CGGCCCTCC
739	8mer	0.037833	GACGCCCTC
740	8mer	0.037833	TTCATATGA
741	8mer	0.037822	CTCGTCTGA
742	6mer	0.037789	TCCGCG
743	8mer	0.037764	GGCCGTCTG
744	8mer	0.037759	TGGGGTAT
745	8mer	0.037755	TGTCGCCG
746	8mer	0.03775	CCTTCACT
747	8mer	0.037746	ATATGTGA
748	8mer	0.037736	CCGAACCTC
749	8mer	0.037682	TGGTTGCT
750	8mer	0.037677	CTGCTCAC
751	8mer	0.037644	CAGGCGGG
752	8mer	0.03761	GCCCGTCTG
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	AGGTCTGT
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	AAGCGCGG
768	8mer	0.037394	GGGGCGCG
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	CCCGCCCC
776	8mer	0.037257	CGGCCCTAT
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	GCGGCCGG
784	8mer	0.037131	ACCCGGCC
785	8mer	0.037086	ATTAGGGA
786	8mer	0.037082	CCGCTTAG
787	6mer	0.037077	GGCCCG
788	8mer	0.037036	GGCATTCTG
789	8mer	0.037023	GGCATTC
790	8mer	0.036989	TAAATATGG
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	CGCCAACT
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	GTGCGCAA
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	ACGGGTTT
813	8mer	0.036648	TTCAGGTT
814	8mer	0.036629	TGTACGTT
815	8mer	0.036619	GTCGACCC

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	GGCCCTTG
123	8mer	0.061704	CGGGGCGA
124	8mer	0.061609	TCGTCCAA
125	8mer	0.061468	CGGGGCCT
126	8mer	0.061418	CGATGATA
127	8mer	0.06132	GGGCGAGG
128	6mer	0.061176	CCCCCG
129	8mer	0.061146	TCCGTACT
130	8mer	0.061133	CGGTGTGC
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	ACTCCCGC
138	8mer	0.060568	CAAGCGAT
139	8mer	0.060416	AGGGGATA
140	8mer	0.060304	CCTCGGGC
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	GGGGCCCT
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	AAACGTGC
160	8mer	0.058738	AGATCCCG
161	8mer	0.058588	TCCGTAGC
162	8mer	0.05853	GGCGCAGC
163	8mer	0.058336	ACGGCGAG
164	8mer	0.058097	TAGGACAA
165	8mer	0.057951	GCCTTGGG
166	8mer	0.057646	AATCCCGC
167	8mer	0.057561	CGCCGTTA
168	8mer	0.057561	GTAACCGC
169	8mer	0.057514	GGCCTATG
170	8mer	0.057493	CGGACGTG
171	8mer	0.057406	GCCGGAAG
172	8mer	0.057352	TTCCTTAA
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	GGAGCCGG
191	8mer	0.056504	ATAGCTAA
192	8mer	0.056455	CGGGATTG
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	TCGCGCGC
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

466	8mer	0.04371	GGAACCTA
467	8mer	0.04367	AAGGTGCC
468	8mer	0.04364	TATTAGGG
469	8mer	0.04363	CCCGAACT
470	8mer	0.04362	TTGTACGG
471	6mer	0.04358	GATGCG
472	8mer	0.04358	CCGTGTCC
473	8mer	0.04352	CGCCTATA
474	8mer	0.04351	TATGTCGG
475	8mer	0.0435	TACTGCCA
476	8mer	0.04347	TATAGCAC
477	8mer	0.04347	GGTTGCAG
478	8mer	0.04346	TTTGCCCA
479	8mer	0.04342	CCGCGATT
480	8mer	0.04336	GGTTGTTC
481	8mer	0.04334	GAGCGGGG
482	8mer	0.04333	GAGGCCGT
483	8mer	0.04332	GTGTCGTA
484	8mer	0.0433	CTGCGAGT
485	8mer	0.0433	CGGCCCTT
486	8mer	0.04328	GCAGTTCG
487	8mer	0.04326	TCCTGTGT
488	8mer	0.04325	GGGCGGTG
489	8mer	0.04325	CCGTTTCC
490	6mer	0.04324	GCCGGG
491	8mer	0.0432	TCGTGCAA
492	8mer	0.04319	TTGCGCCA
493	8mer	0.04314	AGCATTGC
494	8mer	0.04313	ACGCGTAC
495	8mer	0.04313	CTTCCGGT
496	8mer	0.0431	GGGGCGGT
497	8mer	0.04309	ACCCGCAC
498	8mer	0.04305	GAGCCCGG
499	8mer	0.04305	ATCCCGAC
500	8mer	0.04304	GAACGAAT
501	8mer	0.04297	CAGTACGT
502	6mer	0.04296	GCCGGC
503	8mer	0.04296	CTTAAAGT
504	8mer	0.04296	GAGATACT
505	8mer	0.04295	CGTCGTCT
506	8mer	0.0429	GACGCGTA
507	8mer	0.04287	GCACTCCA
508	8mer	0.04286	CGCGCGGC
509	8mer	0.04285	TGCTTTTT
510	6mer	0.04284	CCGGGC
511	8mer	0.04278	CCGGACTA
512	8mer	0.04275	ATTGCGCC
513	8mer	0.04274	CGAGGTCT
514	8mer	0.04273	CAGGCCAC
515	6mer	0.04273	CGTCGT
516	8mer	0.04269	GATCCGGG
517	8mer	0.04266	TCCTCCGT
518	8mer	0.04263	CTAAACAT
519	8mer	0.04261	CGGCGCAG
520	6mer	0.04258	GTTTCG
521	8mer	0.04258	CACAGATG
522	8mer	0.04258	GTAAGGGC
523	8mer	0.04257	GGACGCAA
524	8mer	0.04255	CCTGCGCC
525	8mer	0.04255	GGCGGTGT
526	6mer	0.04251	ACCGCG
527	8mer	0.04248	TGCACTTG
528	8mer	0.04246	AGGAGGCC
529	Triplet	0.04242	Triplet
530	8mer	0.04237	CCGAAGGG
531	8mer	0.04234	TGAGCCAC
532	8mer	0.04231	TCGGGGGT
533	8mer	0.04227	CCCCGGAG
534	SC-PseDNC	0.04225	SC-PseDNC
535	8mer	0.04224	CAACGTCT
536	8mer	0.04222	CCGAAGAC
537	8mer	0.04217	TCGAGAGC
538	8mer	0.04215	ACGGCCCA
539	8mer	0.04214	CGTGCGAA
540	8mer	0.04213	AACTGCGG
541	8mer	0.04207	CCAGGCAC
542	8mer	0.04206	AGGTCTGA
543	8mer	0.04206	GCCCGGCA
544	8mer	0.04206	AAGCGAGG
545	8mer	0.04204	TGCTCGCG
546	8mer	0.04204	CGGAGGTT
547	8mer	0.04194	GTTCCCAT
548	8mer	0.04194	AGACGGTC
549	8mer	0.0419	CCATAGTG
550	8mer	0.04188	CGAAGGCG
551	8mer	0.04183	GCGGCCAG
552	8mer	0.04182	AGCTTTTG
553	8mer	0.04181	CCCCCCTT
554	8mer	0.04181	TCGACCTC
555	8mer	0.04173	CCCCGAGG
556	8mer	0.0417	CACCACTT
557	8mer	0.0417	GACCTCGG

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	CGGCCCCAC
822	8mer	0.036574	GAAGTCGG
823	8mer	0.036528	ATGCACGC
824	8mer	0.036518	TAACGAGG
825	8mer	0.036479	TTCGGGGG
826	8mer	0.036471	CGGGTGGC
827	8mer	0.036463	CAATATAC
828	8mer	0.036425	TCCGCTGT
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	ACGCCCGA
851	8mer	0.036032	GCGCGCCG
852	8mer	0.036022	GTTTCGCC
853	8mer	0.036005	TCCTGGGG
854	8mer	0.036004	TTGTTAAT
855	8mer	0.03599	CAGCCTCC
856	8mer	0.035982	GGTGCGCG
857	8mer	0.035958	CGTTACAG
858	6mer	0.035953	CCAACG
859	8mer	0.035936	CGAGTTCT
860	8mer	0.03593	GCGCGCGG
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	GAACCTGT
870	8mer	0.035724	ACTGCACG
871	8mer	0.035716	GAGGGCCG
872	8mer	0.03568	ACCATTA
873	6mer	0.03567	CGGAGC
874	8mer	0.035668	CAAGCGAA
875	8mer	0.035666	GCGGCGCA
876	8mer	0.035652	GCCATCTA
877	8mer	0.035597	GTCTAGGC
878	8mer	0.035581	ATCATTTAT
879	8mer	0.035577	TGGGCCCC
880	8mer	0.035562	TAACCGCG
881	8mer	0.035558	GGCTGGGA
882	8mer	0.035557	GTAACAT
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	GCGGCGCG
895	8mer	0.03536	TGGGATTA
896	8mer	0.035351	AGCCGGGG
897	8mer	0.035334	TTCATAGT
898	8mer	0.03528	CCCTTCGA
899	6mer	0.035237	GGCCCG
900	8mer	0.03522	CTGCGCCT
901	6mer	0.035216	CCAGG
902	8mer	0.03518	CTCGGCTG
903	8mer	0.035126	CACGCCAC
904	8mer	0.035113	CTCGGGTC
905	8mer	0.035093	CGCGGTGC
906	6mer	0.035071	CCGGCA
907	8mer	0.035069	CTGTCGAC

208	8mer	0.0552	GCCCTGGT
209	8mer	0.055116	GTCCAATG
210	8mer	0.055039	CGGCGATG
211	8mer	0.054965	TGAGTGTG
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	TAGAGGGC
220	8mer	0.054134	CGGACTTT
221	8mer	0.053955	GCCCCGTG
222	8mer	0.053902	AAAATTAG
223	8mer	0.053867	CGAGCCCG
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	CCGCAATC
230	8mer	0.052943	ACGGAGGG
231	8mer	0.052902	GCTGCTCG
232	8mer	0.052867	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	TGGCCGTG
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	GGGTTCGAG
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	ACGGGCGAG
263	8mer	0.050867	AGATCGCG
264	6mer	0.050856	ACGGGA
265	8mer	0.050733	GATATTGA
266	8mer	0.050698	AGCCCCGA
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	GGAACACG
285	8mer	0.049918	GCGGTTCC
286	8mer	0.049887	CCTCCGTA
287	8mer	0.049846	GATCGCGC
288	8mer	0.049818	TTGGTGCT
289	8mer	0.049787	CGGAGGGC
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	CACCTGCAC
297	8mer	0.049199	TGCCTGCG
298	8mer	0.049125	CAAGGAGT
299	8mer	0.049102	CATTGCCG

558	6mer	0.04166	TCGCCC
559	Triplet	0.04159	Triplet
560	8mer	0.04158	TCCGGTTG
561	8mer	0.04153	CGGAAACG
562	6mer	0.04151	CTCACG
563	8mer	0.04145	ACAATGAG
564	8mer	0.04145	CCCTCGTC
565	8mer	0.04144	AGGTTCAA
566	8mer	0.04144	GCCCCCGC
567	8mer	0.04142	CCGGGCCC
568	8mer	0.04142	GGGAGATC
569	8mer	0.04138	TACTCGGG
570	8mer	0.04134	TACGTAAA
571	8mer	0.0413	TCGGGGGC
572	8mer	0.04129	CCCCGATT
573	8mer	0.04124	GACCTGGC
574	6mer	0.04121	TTCGCC
575	6mer	0.04121	AGGCAC
576	8mer	0.04118	GCCCGAGG
577	8mer	0.04114	GCATTCCG
578	8mer	0.04113	CATTGATC
579	8mer	0.04111	GGGGCGGG
580	8mer	0.04108	AAGGGTCC
581	8mer	0.04108	TTGCGGAA
582	8mer	0.04101	CTAATGCA
583	8mer	0.04095	CCCCGTCT
584	6mer	0.04088	CGACGA
585	8mer	0.04083	CATCGCCG
586	8mer	0.04082	CCGGCCGG
587	8mer	0.0408	GGGACCGC
588	8mer	0.0408	TAACGGGA
589	8mer	0.0408	CGCGATGA
590	8mer	0.0408	ACGCGCTA
591	8mer	0.0408	GCGCTAGG
592	8mer	0.0408	GTAACGGG
593	8mer	0.0408	GTCGTAAT
594	8mer	0.04079	CTGGGCCT
595	8mer	0.04073	CCTGGCGT
596	6mer	0.04072	GAGGGG
597	8mer	0.04071	TGGGACCT
598	8mer	0.04069	CGTAATAC
599	6mer	0.04065	CTCGGC
600	8mer	0.04064	GGCGTTCA
601	8mer	0.04057	TCACGCCA
602	6mer	0.04057	TCCGCC
603	8mer	0.04056	CGAACACC
604	8mer	0.04056	TGCTIACC
605	8mer	0.04048	AACGGGTT
606	8mer	0.04046	CACGCCCT
607	8mer	0.04046	GGGATCCG
608	8mer	0.0404	CACCCCGC
609	8mer	0.0404	CGAACGGG
610	8mer	0.0404	CGTCCAGC
611	8mer	0.04039	TGGCGTGC
612	8mer	0.04038	TGGGACTG
613	8mer	0.04037	CCCCTGAC
614	8mer	0.04037	GCCGCGCT
615	8mer	0.04033	TCCTACCT
616	8mer	0.04033	CGCTGGTA
617	8mer	0.04031	GCGCCTAG
618	8mer	0.04031	CTGACGGC
619	8mer	0.04029	GGATTATA
620	8mer	0.04026	CGACCAAA
621	8mer	0.04025	GAACGTGC
622	8mer	0.04025	TTGGCCAG
623	8mer	0.04024	ACACGGGA
624	8mer	0.04022	AGCCGTTT
625	8mer	0.04021	TGCTCACG
626	8mer	0.04019	AGGAGGGG
627	8mer	0.04019	GGAGCGCT
628	8mer	0.04017	CCAATGCC
629	8mer	0.04017	ATCGCGCC
630	6mer	0.04012	GCTCAC
631	8mer	0.04004	GGGTGTTT
632	8mer	0.04003	GCCACTGC
633	8mer	0.03999	GGGAGGCT
634	8mer	0.03999	TATTGAAG
635	6mer	0.03993	GCGTTC
636	8mer	0.03991	CGAACTCG
637	8mer	0.0399	CGCTAAAT
638	8mer	0.03987	TAATCGAG
639	8mer	0.03987	AACCCGAA
640	8mer	0.03985	CGCGAACG
641	8mer	0.03983	CTGTGGGG
642	8mer	0.03981	CCTCGCGG
643	8mer	0.03981	CGAGATGG
644	6mer	0.03977	GCGAGA
645	8mer	0.03976	GCCGGTCT
646	8mer	0.03963	CATGTAAT
647	8mer	0.0396	TCGCGTCC
648	8mer	0.03955	TCGATGAG
649	8mer	0.03955	GGAGCACT

908	8mer	0.035065	TAGTGGGA
909	8mer	0.035051	CTCAGCCT
910	8mer	0.03505	GGCCGCTC
911	8mer	0.035019	GGCCAGGT
912	8mer	0.035018	TTCACGCC
913	8mer	0.035013	CTGGCCAC
914	8mer	0.034942	CTGGGCTG
915	8mer	0.034939	GTCGGAAA
916	8mer	0.034897	GCACTTGT
917	8mer	0.034893	TCGCTTGA
918	8mer	0.034842	ACTCAGTG
919	8mer	0.034833	GAACCTAA
920	8mer	0.034822	CCCCAGGC
921	8mer	0.034799	TCCCGCAA
922	8mer	0.034779	TCTTAAGT
923	8mer	0.034736	TGTGGCCG
924	8mer	0.034735	CCCTGATT
925	8mer	0.03473	CGCAAAAG
926	8mer	0.034724	ACAAGCGA
927	8mer	0.034711	CGGGCGTT
928	8mer	0.034709	GAGGGCCT
929	8mer	0.034708	GGCGTCCG
930	8mer	0.034707	ACCAAGGG
931	8mer	0.0347	AGGCACCT
932	8mer	0.034691	CTAATGTT
933	8mer	0.034684	ACCGCTGG
934	6mer	0.034677	GGGACG
935	6mer	0.034677	AGGCCG
936	8mer	0.034665	CATGGCAT
937	8mer	0.034628	GCGTGTAT
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	CGCCGCTT
947	8mer	0.034398	CCCTCACG
948	8mer	0.034398	AGGGCCGG
949	8mer	0.034374	CCCCGGCT
950	8mer	0.034353	AGGGGCGT
951	8mer	0.034352	GGGGGACG
952	8mer	0.034348	CGGTATTA
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	GTCGCCGG
964	8mer	0.034099	CCGGGGCG
965	8mer	0.034094	GAAGGCGG
966	8mer	0.034065	GCAATTGC
967	8mer	0.034064	TCCCTCTG
968	8mer	0.034041	GCCACAAA
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	GTCCTTGG
986	8mer	0.033762	CGAATTAC
987	8mer	0.03376	GTCGCTAA
988	8mer	0.033714	ACTGCGAT
989	8mer	0.033696	GCACACCG
990	8mer	0.033673	CCGCGGTT
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	TTTGACGC
998	8mer	0.033524	AACCCGCG
999	8mer	0.03352	CGTATGTG

300	8mer	0.048983	GCCAGAGT	650	8mer	0.03951	AACGCGCTG	1000	8mer	0.033512	CTGCTGTT
301	8mer	0.0489	GGCGCCCA	651	8mer	0.03949	CGGCGCTG				
302	8mer	0.048871	AGCCGCAT	652	8mer	0.03941	GCGCCGAG				
303	8mer	0.048848	CTCACGGG	653	8mer	0.0394	CTCGAGAT				
304	8mer	0.048843	CTTCCCTA	654	8mer	0.03938	CGAAATTA				
305	6mer	0.048775	CCCGGG	655	8mer	0.03936	CAATGCC				
306	8mer	0.048756	GATCGTCG	656	8mer	0.03936	TCGACCCG				
307	8mer	0.048723	GGCTATTG	657	8mer	0.03927	TATACTCG				
308	6mer	0.048722	CCGGTC	658	8mer	0.03922	GGCGCGTG				
309	8mer	0.04869	GGGACTGA	659	8mer	0.03921	GGTAGCGT				
310	8mer	0.048651	TGAACACG	660	8mer	0.03919	GGTCGAGG				
311	8mer	0.048619	GGGCGGGG	661	8mer	0.03918	TTTTTTAG				
312	8mer	0.048591	ACTGTGGG	662	8mer	0.03918	CCGCCAAT				
313	8mer	0.048586	GACCCGCA	663	8mer	0.03915	TGGGGCGG				
314	8mer	0.048551	AGACCTGG	664	8mer	0.03913	TACCGGCA				
315	8mer	0.048515	GACTCGCG	665	8mer	0.03909	CATGCTAC				
316	8mer	0.048481	CGCCGCCG	666	8mer	0.03906	CCCGCAAA				
317	8mer	0.048465	CGGGAGCG	667	8mer	0.03904	AACTGTAC				
318	8mer	0.048459	CGCCCGGC	668	8mer	0.039	GCAGAAAA				
319	8mer	0.048409	GGCGCGGC	669	8mer	0.039	CACGCAAG				
320	8mer	0.048382	ATCGTCGT	670	6mer	0.039	GCCCCG				
321	8mer	0.048319	GTTGTACG	671	8mer	0.039	GCATCGTC				
322	6mer	0.048125	CCCGCA	672	8mer	0.03898	GAGGGCGA				
323	8mer	0.048104	GAACGGCC	673	8mer	0.03897	GGTAAAC				
324	8mer	0.048021	GTGGGAGG	674	8mer	0.03894	CGGATGTC				
325	8mer	0.048009	AGGTCCCT	675	8mer	0.03894	GCGGAGCG				
326	8mer	0.047939	CGGCCATC	676	8mer	0.03893	AAGCCGCA				
327	8mer	0.047913	CGAGACTG	677	6mer	0.0389	CGCGGC				
328	8mer	0.047904	CCCGAAGG	678	8mer	0.03885	TGACGCGA				
329	8mer	0.047877	GTTCTGCA	679	8mer	0.03879	CCACCGAT				
330	8mer	0.047858	CCTGTCCG	680	6mer	0.03876	GCGCCC				
331	8mer	0.047854	CGGCGGGG	681	8mer	0.03875	TGCGTGAG				
332	6mer	0.047844	TGAGCC	682	8mer	0.03874	TCATTGCC				
333	8mer	0.047816	ATCGTAGA	683	8mer	0.03869	AGCGAGAC				
334	8mer	0.047784	CTCGTCAA	684	8mer	0.03869	TGTGTTAC				
335	Triplet	0.047765	Triplet	685	8mer	0.03868	CCGTAGCT				
336	8mer	0.047735	AGTCGGAG	686	8mer	0.03866	ACATCCGT				
337	8mer	0.047729	CCTCTCAT	687	8mer	0.03865	CCTCGTCC				
338	8mer	0.047708	TGCCCCGA	688	8mer	0.03865	CTGCGCCG				
339	8mer	0.047676	ACTTCGAG	689	8mer	0.03862	CTATGGGA				
340	8mer	0.04766	TGCGGGGG	690	8mer	0.03862	TGCAACCT				
341	8mer	0.047656	GCCCGAAC	691	8mer	0.03861	GATTAGAG				
342	8mer	0.047521	TGGGCGCC	692	8mer	0.0386	TCGCCGCC				
343	8mer	0.047505	ACATGGCA	693	8mer	0.03858	AGGATGTT				
344	Triplet	0.047503	Triplet	694	8mer	0.03857	CCGCGGAT				
345	8mer	0.047497	TGCACGCA	695	8mer	0.03857	ATGAGCTA				
346	8mer	0.047407	CGCATTCC	696	8mer	0.03857	ACTCGTCA				
347	8mer	0.047393	CACCGGAC	697	8mer	0.03856	GCGGAGGT				
348	6mer	0.047356	GAGGGC	698	6mer	0.03855	AGGGCG				
349	8mer	0.047279	CGCTTGAA	699	8mer	0.03854	ATAACGAG				
350	8mer	0.047215	GAGATCCC	700	8mer	0.03854	GAATATCT				

TableS4 The top 1000 features in the optimal feature set of the IncLocPredtp568 according to F-score order

Order	Feature	F-score	kmer	Order	Feature	F-score	kmer	Order	Feature	F-score	kmer
1	5mer	0.118544	GGGCG	351	8mer	0.0481	GAACGGCC	701	8mer	0.039192	GGTCGAGG
2	8mer	0.10931	TCCCAAAG	352	8mer	0.04802	GTGGGAGG	702	8mer	0.039183	TTTTTTAG
3	8mer	0.107445	GCGGGGGG	353	5mer	0.04802	CCGAG	703	8mer	0.039178	CCGCCAAT
4	8mer	0.106471	AACGGGCC	354	8mer	0.04801	AGGTCCTC	704	8mer	0.039146	TGGGGCGG
5	8mer	0.105616	CGGTGACG	355	8mer	0.04794	CGGCCATC	705	8mer	0.039133	TACCGGCA
6	8mer	0.098194	CGGGTCAC	356	8mer	0.04791	CGAGACTG	706	8mer	0.039092	CATGCTAC
7	8mer	0.095685	TCACGGGA	357	8mer	0.0479	CCCGAAGG	707	8mer	0.03906	CCCGCAAA
8	8mer	0.09567	TGCGGTTT	358	8mer	0.04788	GTTCTGCA	708	5mer	0.039041	AGCCG
9	8mer	0.09254	GGGGGACC	359	8mer	0.04786	CCTGTCCG	709	8mer	0.039038	AACTGTAC
10	8mer	0.091784	CCACGACC	360	8mer	0.04785	CGGCGGGG	710	8mer	0.039002	GCAGAAAA
11	8mer	0.091036	CGCCCCGG	361	6mer	0.04784	TGAGCC	711	8mer	0.039	CACGCAAG
12	5mer	0.090816	GCCCG	362	8mer	0.04782	ATCGTAGA	712	6mer	0.039	GCCCCG
13	8mer	0.09051	TCAACGTC	363	8mer	0.04778	CTCGTCAA	713	8mer	0.038999	GCATCGTC
14	8mer	0.088855	CGAAGTCG	364	Triplet	0.04777	Triplet	714	8mer	0.038984	GAGGGCGA
15	8mer	0.088667	CGGGGGAC	365	8mer	0.04774	AGTCGGAG	715	8mer	0.038973	GGTAAAC
16	6mer	0.088041	GGGGCG	366	8mer	0.04773	CCTCTCAT	716	8mer	0.038944	CGGATGTC
17	8mer	0.08771	ATGGGGCG	367	8mer	0.04771	TGCCCCGA	717	8mer	0.038939	GCGGAGCG
18	8mer	0.087693	TCGGCCCC	368	8mer	0.04768	ACTTCGAG	718	8mer	0.03893	AAGCCGCA
19	8mer	0.087595	AGCGGAGC	369	8mer	0.04766	TGCGGGGG	719	6mer	0.038902	CGCGGC
20	8mer	0.087322	AATCGACG	370	8mer	0.04766	GCCCGAAC	720	8mer	0.03885	TGACGCGA
21	8mer	0.085901	GCTCACGG	371	8mer	0.04752	TGGGCGCC	721	8mer	0.038789	CCACCGAT
22	6mer	0.084846	GCGGGG	372	8mer	0.04751	ACATGGCA	722	6mer	0.038761	GCGCCC
23	8mer	0.084716	GTAGGGCG	373	Triplet	0.0475	Triplet	723	8mer	0.038745	TGCGTGAG
24	8mer	0.084641	CGCAGTTC	374	8mer	0.0475	TGCACGCA	724	8mer	0.038735	TCATTGCC
25	8mer	0.083424	CCGCAGTT	375	8mer	0.04741	CGCATTCC	725	8mer	0.03869	AGCGAGAC
26	8mer	0.082921	GCGGAGCC	376	8mer	0.04739	CACCGGAC	726	8mer	0.038687	TGTGTTAC
27	8mer	0.081369	CATCGGCT	377	6mer	0.04736	GAGGGC	727	8mer	0.038675	CCGTAGCT
28	8mer	0.079854	AGTAGGAC	378	8mer	0.04728	CGCTTGAA	728	8mer	0.038664	ACATCCGT
29	8mer	0.079681	CCCAAAGT	379	8mer	0.04722	GAGATCCC	729	8mer	0.03865	CCTCGTCC
30	8mer	0.079471	AAGGGCTG	380	8mer	0.04717	GGTCGAGA	730	8mer	0.038649	CTGCGCCG
31	5mer	0.078959	CGCCG	381	8mer	0.04711	AAGTCGGG	731	8mer	0.038618	CTATGGGA
32	8mer	0.078261	GGGATGCG	382	8mer	0.04707	GCAGCCCG	732	8mer	0.038616	TGCAACCT
33	8mer	0.077741	GATCCCGA	383	8mer	0.04705	GCGCGCTG	733	8mer	0.038607	GATTAGAG
34	6mer	0.077289	GGGCGC	384	8mer	0.04699	TGCGTCAA	734	8mer	0.0386	TCGCCGCC
35	8mer	0.077287	GCGTGAGC	385	8mer	0.04697	TGGACAGC	735	8mer	0.038583	AGGATGTT
36	8mer	0.077177	CGGCGCGG	386	8mer	0.04696	TCGTCAAC	736	8mer	0.038573	CCGCGGAT

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	GAGCCCCA
43	8mer	0.07508	CGTAGCTT
44	5mer	0.074635	CCCCG
45	8mer	0.074414	CCCCCGAA
46	5mer	0.073675	CGCC
47	6mer	0.073596	ATCGCG
48	8mer	0.073205	GGCGAGGT
49	8mer	0.073205	CCGCACCG
50	8mer	0.072737	CGCAGCGT
51	8mer	0.072567	CATCGCGG
52	8mer	0.072132	TGGGCCCG
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	0.071377	CGTCCA
59	8mer	0.070518	AAGGGGCG
60	6mer	0.070402	AGCCCG
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	AGACCCCT
68	8mer	0.068069	ACGTGCGG
69	8mer	0.068033	GTGCGACT
70	8mer	0.067987	GACCATCG
71	8mer	0.067781	ACGAATCA
72	8mer	0.067781	TCGTGCGAC
73	8mer	0.067781	TAACCGCC
74	8mer	0.067781	GTGCAACG
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	CGCGCG
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	GGCCCCGT
90	8mer	0.066467	GCGGATCT
91	8mer	0.06636	GGGCCCTC
92	8mer	0.06628	CCTGGGCG
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
107	8mer	0.065246	ATCCGCCG
108	5mer	0.065211	GCGGG
109	8mer	0.064857	TGTGCGTA
110	8mer	0.064394	CTCATTCG
111	8mer	0.06426	GTAACCGC
112	8mer	0.064055	CCGGTTGA
113	5mer	0.063847	CCCCG
114	8mer	0.063797	TGTGCGCC
115	8mer	0.063633	TGCGCGGT
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	TCCAGGGG
123	8mer	0.062893	TCCCGCAC
124	8mer	0.062656	ACCGGAGT
125	8mer	0.06263	AGGTGCGA
126	8mer	0.062472	GAAACCCG
127	8mer	0.062448	TTGCCGGT
128	8mer	0.06233	TCAAGCAC

387	8mer	0.0469	CACGGGAG
388	8mer	0.0468	ACGACCCG
389	8mer	0.04677	TATCCAGT
390	8mer	0.04677	GGTCATCG
391	5mer	0.04676	CCGCA
392	8mer	0.04676	ACGCAGGG
393	8mer	0.04671	CGCGCCGT
394	8mer	0.04669	CTCCCAAA
395	8mer	0.04666	CGGGGCGG
396	8mer	0.04664	TTATGTGC
397	8mer	0.04664	GCGCAGCA
398	6mer	0.04658	CGGGCG
399	8mer	0.04654	GACCGCGA
400	8mer	0.04648	AGAAGCGG
401	6mer	0.04647	ACCCGC
402	5mer	0.04647	AGGGC
403	6mer	0.04646	CCGAGG
404	8mer	0.04643	ATCGCCGA
405	8mer	0.04641	TGTCCGGT
406	8mer	0.04636	AGCAGTCC
407	8mer	0.04628	AGCTCGGC
408	6mer	0.04625	GTCGTG
409	8mer	0.04623	TCTGGTAT
410	8mer	0.04622	CGGGCACT
411	8mer	0.0462	GTTAATTT
412	8mer	0.04614	CAGGACGA
413	8mer	0.04611	GGTCGGGC
414	6mer	0.04608	CGCCGG
415	8mer	0.04606	CGGGGCCA
416	8mer	0.04603	CGCCCGAG
417	8mer	0.04599	TCGCCCTC
418	8mer	0.04596	GGCGGGGG
419	6mer	0.04585	CGCCCG
420	8mer	0.04584	CCCTGGGC
421	8mer	0.04582	CGCCGGAA
422	8mer	0.04581	ATCTTAAT
423	6mer	0.04578	CCGGCT
424	8mer	0.04574	CCGCGAAT
425	8mer	0.0456	TAAGGGCT
426	8mer	0.04557	CGACCTCC
427	8mer	0.04555	ACGCTCGT
428	8mer	0.04554	CGCGGGCA
429	8mer	0.04554	GCGCCAAT
430	Triplet	0.04548	Triplet
431	8mer	0.04545	GACCCGGC
432	8mer	0.04545	GGCTGATG
433	8mer	0.0454	GCACAATT
434	5mer	0.04534	CGGCC
435	8mer	0.04529	GGTTGCTA
436	8mer	0.04526	TGTTTACG
437	8mer	0.04524	AGGGCGCA
438	8mer	0.04523	TAGGCATT
439	8mer	0.04521	CCGAAGTC
440	8mer	0.04521	CGTACATT
441	8mer	0.0452	GCGGTACA
442	8mer	0.04518	CCATCGGT
443	8mer	0.04515	CGAGACCA
444	8mer	0.04515	CCGGACTT
445	8mer	0.04512	ATGTCGGG
446	8mer	0.04507	CTAGCGGA
447	8mer	0.04506	TCGACGAA
448	8mer	0.04506	GGGCGCGG
449	8mer	0.04499	CCCTTTAT
450	6mer	0.04497	TCGGCC
451	8mer	0.04495	GGCAGAAA
452	6mer	0.04493	CCCGAA
453	8mer	0.04493	GTTGCTAT
454	8mer	0.04492	GGTTGGTC
455	8mer	0.04491	ACCGTGCC
456	6mer	0.04489	CCCGCC
457	8mer	0.04484	ACGCCTCG
458	8mer	0.04481	CGTGTTAC
459	8mer	0.0448	ATGCGCGA
460	8mer	0.04471	CTTTCGTC
461	SC-PseDNC	0.04466	SC-PseDNC
462	8mer	0.04457	GGGGCGCG
463	6mer	0.04457	ACGCGA
464	8mer	0.0445	ACCGCGAA
465	8mer	0.04444	GGGGATAT
466	8mer	0.0444	TTTGTCTC
467	8mer	0.0444	AGATTGCG
468	8mer	0.04432	CGAGAATC
469	5mer	0.0443	GGCGT
470	8mer	0.04426	CAGTTCCG
471	8mer	0.04423	ATCGACGA
472	8mer	0.04422	CACGGCAA
473	5mer	0.0442	TCGGC
474	8mer	0.04419	ACCCCGCA
475	8mer	0.04418	CGAGAGCT
476	6mer	0.04416	CTGGGC
477	6mer	0.04416	GGCGGG
478	8mer	0.04415	GCGACTCG

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	CGTACATG
759	8mer	0.038251	CATCCGCT
760	8mer	0.038249	GAGGGCGC
761	8mer	0.038244	TGATTCCG
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	CTCACATG
772	8mer	0.037947	TGCGCCGA
773	8mer	0.037943	GCAAACGT
774	8mer	0.037942	CCGGGAAC
775	8mer	0.037932	TGGCCAGG
776	8mer	0.037889	CGAATGAT
777	8mer	0.037866	ACCGGACA
778	8mer	0.037866	GACGCGAC
779	8mer	0.037866	TGTCGACC
780	8mer	0.037866	TAAACCGT
781	8mer	0.037866	GCTCGCGC
782	8mer	0.037861	CGGCCTCC
783	8mer	0.037833	GACGGCTC
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	TGTCCCCG
790	8mer	0.03775	CCTTCACT
791	8mer	0.037746	ATATGTGA
792	8mer	0.037736	CCGAACCT
793	5mer	0.037698	TGCC
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	CCGAGGGG
803	8mer	0.037538	AGGTGCTG
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	AAGCCGCG
814	8mer	0.037394	GGGGGCGG
815	5mer	0.037369	CAGGC
816	8mer	0.037359	CCCGTCTG
817	8mer	0.037355	CACGGGAT
818	8mer	0.037317	ACCGGAAT
819	8mer	0.037314	TCGCGCCG
820	8mer	0.037312	GCCTATGG
821	8mer	0.03727	AATACCGG
822	8mer	0.037266	CCCCGCCG
823	8mer	0.037257	CGGCCTAT
824	8mer	0.037252	GCTGCGCA
825	8mer	0.037222	CCCGGCAC
826	8mer	0.037219	ACCGGGGG
827	8mer	0.037189	AAGCGAAA
828	6mer	0.037183	CCCCGC

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	GGCCCCCTG
136	8mer	0.061704	CGGGGCGA
137	8mer	0.061609	TCGTCCAA
138	8mer	0.061468	CGGGGCGCT
139	8mer	0.061418	CGATGATA
140	8mer	0.06132	GGGCGAGG
141	6mer	0.061176	CCCCCG
142	8mer	0.061146	TGGGTACT
143	8mer	0.061133	CGGTGTGC
144	8mer	0.061067	GGGGTCGA
145	5mer	0.061032	GCCCC
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	CCTCGGGC
155	8mer	0.060137	GCATTCCG
156	6mer	0.060056	GGCGCG
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	GTGTACT
163	8mer	0.059509	CCAACATG
164	5mer	0.059434	GGCCG
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	AAACGTGC
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	GGCCG
183	8mer	0.057646	AATCCCCG
184	8mer	0.057561	CGCCGTTA
185	8mer	0.057561	GTAACCG
186	8mer	0.057514	GGCCTATG
187	8mer	0.057493	CGGACGTG
188	8mer	0.057406	GCCGGAAG
189	8mer	0.057352	TTCCTTAA
190	8mer	0.057325	CGCGGGAA
191	8mer	0.057279	GGAATATC
192	8mer	0.057262	GGCGCGTC
193	8mer	0.057246	GGTAACCG
194	8mer	0.057244	AAATCCGC
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	TCGTATGG
203	5mer	0.057046	GGGCC
204	8mer	0.056971	CGGGCATC
205	5mer	0.056908	CCCGA
206	8mer	0.056836	CGCCGTGT
207	6mer	0.056774	GCGCGC
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	AGCTACCG
217	8mer	0.056053	ATCGCGGG
218	6mer	0.056047	TCCCGA
219	8mer	0.055956	TCGCGCCG
220	8mer	0.055933	GATTCAAC

479	8mer	0.0441	CCGCTGGT
480	8mer	0.04408	ACTGCTCA
481	8mer	0.04407	GAGCCCCG
482	8mer	0.04403	CCAAAGTG
483	6mer	0.04402	GGGCGA
484	6mer	0.04401	GCGGTA
485	8mer	0.04392	CCGACGCG
486	8mer	0.0439	CACCGCGC
487	8mer	0.04389	GCCGCGCG
488	8mer	0.04387	GCCCTCGA
489	8mer	0.04382	CAAAGTAC
490	8mer	0.04381	GGAGGGCG
491	8mer	0.04381	GGGGGTCT
492	8mer	0.04379	GCCGGACT
493	8mer	0.04379	GCGCCGGA
494	8mer	0.04377	GCAACGTC
495	8mer	0.04377	ACCTAGAC
496	6mer	0.04376	GCGCCG
497	8mer	0.04375	TTTTTGTA
498	8mer	0.04373	AGATACTA
499	8mer	0.04372	AACTCGTC
500	8mer	0.04371	GGAACCTA
501	8mer	0.04367	AAGGTGCC
502	8mer	0.04364	TATTAGGG
503	8mer	0.04363	CCCGAACT
504	8mer	0.04362	TTGTACGG
505	6mer	0.04358	GATGCG
506	8mer	0.04358	CCGTGTC
507	8mer	0.04352	CGCCTATA
508	8mer	0.04351	TATGTCGG
509	8mer	0.0435	TACTGCCA
510	8mer	0.04347	TATAGCAC
511	8mer	0.04347	GGTTGCAG
512	8mer	0.04346	TTTGCCCA
513	8mer	0.04342	CCGCGATT
514	8mer	0.04336	GTTTGTTC
515	8mer	0.04334	GAGCGGGG
516	8mer	0.04333	GAGGCGGT
517	8mer	0.04332	GTCGCTGA
518	8mer	0.0433	CTGCGAGT
519	8mer	0.0433	CGGCCCCT
520	8mer	0.04328	GCAGTTCT
521	8mer	0.04326	TCCTGTGT
522	8mer	0.04325	GGGCGGTG
523	8mer	0.04325	CCGTTTCC
524	6mer	0.04324	GCCGGG
525	8mer	0.0432	TCGTGCGA
526	8mer	0.04319	TGCGCCCA
527	8mer	0.04314	AGCATTGC
528	8mer	0.04313	ACGCGTAC
529	8mer	0.04313	CTCCGCGG
530	5mer	0.04312	CCCCG
531	8mer	0.0431	GGGGCGGT
532	8mer	0.04309	ACCCGCAC
533	8mer	0.04305	GAGCCCCG
534	8mer	0.04305	ATCCCCGAC
535	8mer	0.04304	GAACGAAT
536	8mer	0.04297	CAGTACGT
537	6mer	0.04296	GCCGGC
538	8mer	0.04296	CTTAAGTA
539	8mer	0.04296	GAGATACT
540	8mer	0.04295	CGTCGTCT
541	8mer	0.0429	GACGCGTA
542	8mer	0.04287	GCATCCCA
543	8mer	0.04286	CGCGCGGC
544	8mer	0.04285	TGCTTTTT
545	6mer	0.04284	CCGGGC
546	8mer	0.04278	CCGGACTA
547	8mer	0.04275	ATTGCGCC
548	8mer	0.04274	CGAGGTCT
549	8mer	0.04273	CAGGCCAC
550	6mer	0.04273	CGTCGT
551	8mer	0.04269	GATCCGGG
552	8mer	0.04266	TCCTCCGT
553	8mer	0.04263	CTAAACAT
554	5mer	0.04262	ACGGG
555	8mer	0.04261	CGGGCGAG
556	6mer	0.04258	GTTTCG
557	8mer	0.04258	CACAGATG
558	8mer	0.04258	GTAAGGGC
559	8mer	0.04257	GGACGCAA
560	8mer	0.04255	CCTGCGCC
561	8mer	0.04255	GGCGGTGT
562	6mer	0.04251	ACCGCG
563	8mer	0.04248	TGCATTGT
564	8mer	0.04246	AGGAGGCC
565	Triplet	0.04242	Triplet
566	8mer	0.04237	CCGAAGGG
567	8mer	0.04234	TGAGCCAC
568	8mer	0.04231	TCGGGGGT
569	8mer	0.04227	CCGCGGAG
570	SC-PseDNC	0.04225	SC-PseDNC

829	8mer	0.037161	GAATCGCT
830	8mer	0.037139	GCGCCCCG
831	8mer	0.037131	ACCCGGCC
832	8mer	0.037086	ATTAGGGA
833	8mer	0.037082	CCGCTTAG
834	6mer	0.037077	GGCCCG
835	8mer	0.037036	GGCATTCT
836	8mer	0.037023	GGCATTCC
837	8mer	0.036989	TAAATATG
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	TCGACCTT
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	ACGGGTTT
861	8mer	0.036648	TTCAGGTT
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	GACTCGCA
868	8mer	0.036576	CAGTAGGA
869	5mer	0.036576	CGGGA
870	8mer	0.036575	CGGCCAC
871	8mer	0.036574	GAAGTCGG
872	8mer	0.036528	ATGCACCG
873	8mer	0.036518	TAAAGGAG
874	8mer	0.036479	TTCCGGGG
875	8mer	0.036471	CGGGTGGC
876	8mer	0.036463	CAATATAC
877	8mer	0.036425	TCGGCTGT
878	8mer	0.03639	CTATGTTG
879	6mer	0.036385	TCCGCT
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	ACGGCCCT
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	CTGCGCTC
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	CGTTACAG
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	GAATCTGT

221	6mer	0.055895	CGCCGC
222	5mer	0.055828	CGGAG
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	TGAGTGTG
232	8mer	0.054878	TTCGAAGC
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	CCGCG
240	8mer	0.054301	CGGTAAAC
241	8mer	0.054267	TAGAGGCG
242	8mer	0.054134	CGGACTTT
243	5mer	0.05409	CCGAA
244	8mer	0.053955	CCCCCGTG
245	8mer	0.053902	AAAAATTAG
246	8mer	0.053867	CGAGCCGC
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	CCGCACTT
253	5mer	0.053022	GGGCG
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	TTTGCGGA
268	8mer	0.052338	CCGCCCCC
269	8mer	0.052314	CTGACGAG
270	5mer	0.052297	CGCGG
271	8mer	0.052189	GGGGCCTA
272	8mer	0.052127	GCAGAGGG
273	8mer	0.052092	CGTCGGGG
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	GGGTGCGA
280	8mer	0.051532	TCGAGTCG
281	8mer	0.051301	TCCAGGCT
282	8mer	0.051286	GAGCGCCA
283	8mer	0.051239	ATTGCCTG
284	8mer	0.051176	CAAGATCG
285	8mer	0.051123	GCGTTTCA
286	8mer	0.051068	GCTCTTCG
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	AGCCCCGA
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	GGGCGGGC
303	8mer	0.050237	TGCCGGTC
304	8mer	0.0502	TAATACCG
305	8mer	0.050185	ATTACTGC
306	8mer	0.050111	CTGAGCA
307	8mer	0.050078	GACTGTAT
308	8mer	0.050071	GGCGTGCT
309	6mer	0.049974	CGAGAC
310	8mer	0.049934	GGAACCGG
311	8mer	0.049918	GCGGTTCC
312	5mer	0.049887	TCGGC

571	8mer	0.04224	CAACGTCT
572	5mer	0.04222	CGAGG
573	8mer	0.04222	CCGAAGAC
574	5mer	0.04221	CTCGG
575	8mer	0.04217	TCGAGAGC
576	8mer	0.04215	ACGGCCCC
577	8mer	0.04214	CCGTGGAA
578	8mer	0.04213	AACTGCGG
579	8mer	0.04207	CCAGGCAC
580	8mer	0.04206	AGGTCGTA
581	8mer	0.04206	GCCCGGCA
582	8mer	0.04206	AAGCGAGG
583	8mer	0.04204	TGCTGCGC
584	8mer	0.04204	CGGAGGTT
585	8mer	0.04194	GTTCCAC
586	8mer	0.04194	AGACGGTC
587	8mer	0.0419	CCATAGTG
588	8mer	0.04188	CGAAGGCG
589	8mer	0.04183	GCCGCCAG
590	8mer	0.04182	AGCTTTTG
591	8mer	0.04181	CCCCCCCC
592	8mer	0.04181	TCGACCTC
593	8mer	0.04173	CCCGGAGG
594	8mer	0.0417	CACCACTT
595	8mer	0.0417	GACCTCGG
596	6mer	0.04166	TCGCCC
597	Triplet	0.04159	Triplet
598	8mer	0.04158	TCCGGTTG
599	8mer	0.04153	CGGAAACG
600	6mer	0.04151	CTCACG
601	8mer	0.04145	ACAAATGAG
602	8mer	0.04145	CCCTCGTC
603	8mer	0.04144	AGGTTCAA
604	8mer	0.04144	GCCCCCGG
605	8mer	0.04142	CCGGGCCC
606	8mer	0.04142	GGGAGATC
607	8mer	0.04138	TACTCGGG
608	8mer	0.04134	TACGTAAA
609	8mer	0.0413	TCGCGGGC
610	8mer	0.04129	CCCGTATT
611	8mer	0.04124	GACCTGGC
612	6mer	0.04121	TTCCGC
613	6mer	0.04121	AGGCAC
614	8mer	0.04118	GCCCGAGG
615	8mer	0.04114	GCATTCCG
616	8mer	0.04113	CATTGAT
617	8mer	0.04111	GGGGCGGG
618	8mer	0.04108	AAGGGTCC
619	8mer	0.04108	TTGCGGAA
620	5mer	0.04104	GCGCG
621	8mer	0.04101	CTAATGCA
622	8mer	0.04095	CCCGTCTC
623	5mer	0.04091	GCCCC
624	6mer	0.04088	CGACGA
625	8mer	0.04083	CATCGCCG
626	8mer	0.04082	CCGCGCCG
627	8mer	0.0408	GGGACCCG
628	8mer	0.0408	TAACGGGA
629	8mer	0.0408	CGCGATGA
630	8mer	0.0408	ACGCGCTA
631	8mer	0.0408	GCGCTAGG
632	8mer	0.0408	GTAACGGG
633	8mer	0.0408	GTCGTAAT
634	8mer	0.04079	CTGGGCCT
635	8mer	0.04073	CCTGGCGT
636	6mer	0.04072	GAGGG
637	8mer	0.04071	TGGGACCT
638	8mer	0.04069	CGTAATAC
639	6mer	0.04065	CTCGCG
640	8mer	0.04064	GCGGTTCA
641	8mer	0.04057	TCACGCCA
642	6mer	0.04057	TCCGCC
643	8mer	0.04056	CGAACACC
644	8mer	0.04056	TGCTTACC
645	8mer	0.04048	AACGGGTT
646	8mer	0.04046	CACGCCTC
647	8mer	0.04046	GGGATCGC
648	8mer	0.0404	CACCCCCG
649	8mer	0.0404	CGAACGGG
650	8mer	0.0404	GCTCCAGC
651	8mer	0.04039	TGGCGTGC
652	8mer	0.04038	TGGGACTG
653	8mer	0.04037	CCCCTGAC
654	8mer	0.04037	GCCGGCTG
655	8mer	0.04033	TCCTACCT
656	8mer	0.04033	CGCTGGTA
657	8mer	0.04031	GGGCCTAG
658	8mer	0.04031	CTGACGGC
659	8mer	0.04029	GGATTATA
660	8mer	0.04026	CGACCAAA
661	8mer	0.04025	GAACGTGC
662	8mer	0.04025	TGGGCCAG

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	CACCTACG
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	TTCACCTG
950	8mer	0.03528	CCCTTCGA
951	6mer	0.035237	GGCCGG
952	8mer	0.03522	CTGCGCCT
953	6mer	0.035216	CCAGG
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	CTGTGCAC
960	8mer	0.035065	TAGTGGGA
961	8mer	0.035051	CTACAGCT
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	GTCGGGAA
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	TCCGCCAA
975	8mer	0.034779	TCCTAAGT
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	GGCGTCCG
984	8mer	0.034707	ACCAAGGG
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	CCGGGC
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

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	GGCGCCCA
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
339	8mer	0.048619	GGCGGGGG
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	CGCCCCGC
347	8mer	0.048409	GGCGCGGC
348	8mer	0.048382	ATCGTCGT
349	8mer	0.048319	GTTGTACG
350	6mer	0.048125	CCCGCA

663	8mer	0.04024	ACACGGGA
664	8mer	0.04022	AGCCGTTT
665	8mer	0.04021	TGCTCACG
666	8mer	0.04019	AGGAGGGG
667	8mer	0.04019	GGAGCGCT
668	8mer	0.04017	CCAATGCC
669	8mer	0.04017	ATCGCGCC
670	6mer	0.04012	GCTCAC
671	8mer	0.04004	GGGTGTTT
672	8mer	0.04003	GCCACTGC
673	8mer	0.03999	GGGAGGCT
674	8mer	0.03999	TATTGAAG
675	6mer	0.03993	GCGTTC
676	8mer	0.03991	CGAACTCG
677	8mer	0.0399	CGCTAAAT
678	8mer	0.03987	TAATCGAG
679	8mer	0.03987	AACCCGAA
680	8mer	0.03985	CGCGAACG
681	8mer	0.03983	CTGTGGGG
682	8mer	0.03981	CCTCGCGG
683	8mer	0.03981	CGAGATGG
684	6mer	0.03977	GCGAGA
685	8mer	0.03976	GCCGGTCT
686	8mer	0.03963	CATGTAAT
687	8mer	0.0396	TCGCGTCC
688	8mer	0.03955	TCGATGAG
689	8mer	0.03955	GGAGCACT
690	8mer	0.03951	AACGCCTG
691	8mer	0.03949	CGGCGCTG
692	8mer	0.03941	GCGCCGAG
693	8mer	0.0394	CTCGAGAT
694	8mer	0.03938	CGAAATTA
695	8mer	0.03936	CAATGCCC
696	8mer	0.03936	TCGACCCG
697	5mer	0.03936	GGTCG
698	8mer	0.03927	TATACTCG
699	8mer	0.03922	GGCGCGTG
700	8mer	0.03921	GGTAGCGT