

SpIM

Prediction of splice sites in Human DNA sequences.

The program developed by Salamov A and Solovyev V. It locates potential splice site positions based on 5 weight matrices for donor sites and a model including dinucleotide composition and weight matrix for acceptor splice site. Program includes prediction of potential GC -donor sites and non-standard splice sites as AT-AC

Program does not EXCLUDE splice sites close to sites predicted with higher scores or sites on different chains. User could make processing based on the reported scores. It designed to be useful to analyze ALTERNATIVE Splice variants and NON-CANONICAL splice sites. Program has much higher number of overpredicted sites comparing with Spl program.

For some description of this program see:

Solovyev V.V. (2001) Statistical approaches in Eukaryotic gene prediction. In Handbook of Statistical genetics (eds. Balding D. et al.), John Wiley & Sons, Ltd., p. 83-127.

Example of output:

Splm: Matrix-based prediction of splice sites in Human sequences

Parameters: -d 90 -a 90 -dGC 90 -nc 1 (non-st. consensus AT-AC)

Length of sequence 4500

Number of Donor sites: 22 Threshold: 90

Number Position Score Chain Type

1	167	33	-	GT
2	184	43	-	GC
3	460	25	-	GT
4	486	21	-	GC
5	710	97	+	GT
6	1077	48	+	GT
7	1081	18	+	GT
8	1181	75	-	GT
9	1920	24	+	GT
10	2179	36	-	GC
11	2691	45	+	GT
12	2745	43	-	GC
13	2906	18	+	GT
14	2937	83	+	GT
15	3006	14	-	GT
16	3023	90	-	GT
17	3041	29	-	GT
18	3107	11	-	GT
19	3174	46	+	GT
20	3290	12	-	GT
21	4156	51	-	GT
22	4308	22	+	GT

Number of Acceptor sites: 38 Threshold: 90

1	110	24	-	AG
2	498	12	+	AG
3	680	15	+	AG
4	702	18	-	AG
5	738	19	+	AG
6	780	27	-	AG
7	861	49	+	AG
8	912	34	-	AG
9	1033	24	+	AG
10	1384	8	-	AC
11	1399	16	+	AG
12	1780	11	-	AG
13	1809	14	-	AG
14	2072	13	+	AG
15	2120	29	-	AG

16	2212	61	+	AG
17	2238	24	-	AG
18	2258	18	-	AG
19	2453	8	-	AC
20	2474	12	-	AG
21	2508	9	-	AC
22	2576	94	+	AG
23	2691	9	-	AC
24	2755	33	+	AG
25	2841	41	-	AG
26	3045	8	+	AC
27	3108	27	-	AG
28	3185	14	-	AG
29	3241	39	+	AG
30	3267	23	-	AG
31	3776	25	+	AG
32	3825	13	-	AG
33	3885	8	+	AC
34	4200	12	+	AG
35	4252	29	+	AG
36	4290	18	-	AG
37	4334	9	+	AC
38	4388	13	+	AG