

5'SS	3'SS	Observed	Expected	Obs/Exp	P-value
CGCG	TTTT	661	440.2	1.5	7.2e-23
CGCG	TTTA	431	283.3	1.5	2.3e-16
CGCG	TTAA	341	212.5	1.6	3.3e-16
CGCG	AATT	336	210.1	1.6	8.6e-16
CGCG	GTTT	547	383.8	1.4	2.9e-15
CGCG	ATTT	502	350.2	1.4	1.5e-14
CGCG	TAAA	306	194.3	1.6	9.0e-14
CGCG	AAAT	352	232.8	1.5	2.2e-13
CGCG	ATTA	245	150.9	1.6	1.3e-12
GCGC	TTAA	655	496.4	1.3	6.5e-12
CGCG	AAAA	280	182.7	1.5	1.5e-11
CGCG	AATA	252	161.3	1.6	2.5e-11
CGCG	TATT	346	238.6	1.4	4.3e-11
CGCG	TTAT	335	229.6	1.5	4.4e-11
GGCG	TAAA	855	679.7	1.3	5.7e-11
CGCG	TTTC	785	617.8	1.3	6.0e-11
GCGC	TAAA	594	454.6	1.3	2.4e-10
GCGC	TTTT	1175	975.4	1.2	3.2e-10
CGCG	TTTG	583	446.3	1.3	3.6e-10
GCGG	TAAA	848	681.5	1.2	4.3e-10
CGCG	TAAT	286	194.9	1.5	6.3e-10
GGCG	AAAT	983	806.9	1.2	1.1e-09
GCGC	GTTT	995	819.7	1.2	1.7e-09
GCGG	GTTT	1508	1291.0	1.2	2.1e-09
GGCG	TTAA	908	742.5	1.2	2.4e-09
GGCG	TTTT	1715	1487.1	1.2	4.2e-09
GGCG	GTTT	1469	1260.7	1.2	5.8e-09
GCGG	TTAA	900	744.7	1.2	1.9e-08
CGCG	CTTT	750	610.7	1.2	2.8e-08
CGCG	AACA	320	232.9	1.4	3.7e-08
GCGC	AAAA	534	419.2	1.3	4.1e-08
CGCG	ATAA	224	152.7	1.5	4.1e-08
GCGC	AATT	611	489.2	1.2	6.3e-08
GCGC	ATTA	459	354.7	1.3	6.5e-08
GGCG	ATTT	1386	1200.2	1.2	8.7e-08
ACCC	CTCC	5368	4994.5	1.1	9.1e-08
GCGG	AAAA	769	634.7	1.2	1.3e-07
GCCG	TTAA	863	720.3	1.2	1.3e-07
GGCG	AAAA	764	630.3	1.2	1.3e-07
GCGG	TTTT	1716	1512.1	1.1	1.5e-07
GCGC	TTTA	790	654.4	1.2	1.5e-07
GCGC	AAAT	659	536.9	1.2	1.9e-07
GCGG	AAAT	960	812.0	1.2	2.3e-07
CGCG	ATAT	226	158.3	1.4	2.5e-07
GGCG	ATTA	648	529.1	1.2	3.2e-07
GGCG	AATT	872	733.1	1.2	3.4e-07

GCGC	TATT	676	555.0	1.2	3.6e-07
GCGC	ACGA	147	94.6	1.6	3.7e-07
GCGC	ATTT	936	792.8	1.2	4.1e-07
CGCG	TGTT	593	480.9	1.2	4.4e-07
CGCG	TTGT	512	408.9	1.3	5.1e-07
GCGC	ATAA	450	355.7	1.3	8.5e-07
CGCG	ACAA	236	170.3	1.4	1.1e-06
GCGC	AATA	476	380	1.3	1.1e-06
CGCG	TTAG	273	202.3	1.3	1.3e-06
CGGC	TTAA	718	599.2	1.2	1.3e-06
CGCG	TTCA	475	380.0	1.2	1.4e-06
CGCG	AAGT	278	207.0	1.3	1.5e-06
CCGC	GTTT	1192	1038.8	1.1	1.8e-06
ACCC	ACCC	3387	3128.4	1.1	2.6e-06
GCGC	AAAC	500	404.9	1.2	2.8e-06
CGCG	TATA	183	128.0	1.4	2.8e-06
ACCC	CCCA	4984	4671.5	1.1	3.1e-06
CGCC	TTAA	742	625.4	1.2	3.1e-06
GCCG	TAAA	779	659.8	1.2	3.4e-06
CCCC	ACCC	4064	3785.2	1.1	3.8e-06
CGCG	ATGT	397	314.1	1.3	3.9e-06
CGCC	TAAA	683	572.5	1.2	4.0e-06
CGCG	AAAC	252	187.3	1.3	4.0e-06
CCGC	TTTT	1377	1218.3	1.1	4.4e-06
GCGG	ATTT	1372	1213.9	1.1	4.6e-06
CGGC	TCGT	278	210.2	1.3	4.6e-06
GGCG	TTTA	1126	983.5	1.1	4.7e-06
CCGG	TTAA	856	734.1	1.2	6.2e-06
CCCC	CCCC	5332	5020.0	1.1	6.6e-06
GCGG	TTTC	2254	2054.2	1.1	7.4e-06
CGGC	ATTT	1108	969.7	1.1	7.4e-06
CGCG	TTGA	338	264.6	1.3	8.3e-06
CGGG	TTAA	1201	1057.9	1.1	8.7e-06
CGCT	GTCG	193	139.1	1.4	9.2e-06
CGCG	AGTT	275	209.8	1.3	9.7e-06
CGCG	TGAA	362	286.5	1.3	1.0e-05
CCGC	TAAA	655	552.0	1.2	1.1e-05
CGCG	GAAT	262	198.9	1.3	1.1e-05
GCGG	TTTG	1657	1491.3	1.1	1.3e-05
ACCC	CCTC	5338	5037.4	1.1	1.4e-05
GCGC	TAAT	544	452.0	1.2	1.4e-05

Table S1. Significant 4-mer Co-Occurrences in Constitutive Human Introns. Pairs

are listed with the 4-mer occurring in the first 80 nt of the intron (column 1), the 4-mer

occurring in the last 80 nt of the intron (column 2), the observed number of co-occurrences in the constitutive human dataset (column 3), the expected number calculated by coCOA (column 4), the ratio of observed to expected (column 5), and the Poisson P-value calculated by coCOA. Pairs are sorted by P-value.

5' SS	3' SS	Observed	Expected	Obs/Exp	P-value
CCGCG	TTTTT	143	70.3	2.0	2.1e-14
CGCGG	TTAAA	89	35.6	2.5	4.7e-14
CGGCG	ATTTT	116	55.6	2.1	1.0e-12
GGCGC	TAAAA	127	63.2	2.0	1.2e-12
GCGCG	ATTTT	123	61.3	2.0	2.8e-12
CGCGC	TTTTT	113	54.8	2.1	4.4e-12
CGGCG	TTTTA	101	47.4	2.1	9.4e-12
GGCGC	TTAAA	132	68.8	1.9	9.4e-12
CCGCG	ATTTT	119	60.2	2.0	1.5e-11
CCGCG	TTTTTC	177	102.9	1.7	2.3e-11
GCGCG	TTTTA	107	52.5	2.0	2.9e-11
CGCGC	ATTTT	99	47.5	2.1	4.9e-11
GGCGC	TTTAA	151	85.2	1.8	8.2e-11
CGCGC	TTTTTC	140	77.3	1.8	1.0e-10
GGCGC	AAAAT	134	73.0	1.8	1.0e-10
CCGCG	TTTGT	130	70.3	1.8	1.3e-10
CGCGG	TGTTT	180	108.4	1.7	2.2e-10
CGGCG	TATTT	93	45.1	2.1	3.0e-10
CGCGC	GTTTT	94	46.3	2.0	5.1e-10
CCGCG	TTTTG	116	62.1	1.9	7.1e-10
GCGCG	GTTTT	112	59.3	1.9	7.2e-10
GGCGC	TTTTA	194	121.2	1.6	7.5e-10
GGGCG	ATTTT	330	231.9	1.4	8.2e-10
CGCGG	TTTGT	146	84.5	1.7	8.3e-10
GGCGC	ATTTT	219	141.3	1.5	9.1e-10
CGCGG	TTTAA	90	44.2	2.0	1.0e-09
GCGCG	AATTT	74	33.7	2.2	1.5e-09
CGCGG	ATTTT	130	73.8	1.8	2.2e-09
CGCGG	TAAAA	72	32.8	2.2	2.4e-09
CGGCG	TTTTT	117	64.5	1.8	2.9e-09
CCGCG	TAAAA	62	26.5	2.3	2.9e-09
CGCCG	TTTAA	70	31.6	2.2	2.9e-09
CGCGG	GTTTT	129	73.4	1.8	3.0e-09
CGCGG	TAAAT	66	29.1	2.3	3.0e-09
CGCGG	AATTT	83	40.5	2.0	3.3e-09
CGGCG	TTAAA	62	26.8	2.3	4.9e-09
GCGCG	TTTTT	124	70.6	1.8	6.1e-09
CCGCG	TTTAA	75	35.6	2.1	6.4e-09
GCGCC	TTTAA	128	73.9	1.7	7.9e-09
CCGCG	CTTTT	154	93.8	1.6	8.2e-09
CGGCG	TGTAT	59	25.4	2.3	9.1e-09
GCGCG	TGTTT	144	86.3	1.7	9.1e-09
CCGCG	TCTTT	170	106.7	1.6	1.0e-08
GCGCG	TTTAT	84	42.4	2	1.2e-08
GCGCC	TTAAA	108	59.7	1.8	1.3e-08
CGCGC	CTTTT	122	70.3	1.7	1.4e-08

CCGCG	TTTCT	197	128.8	1.5	1.5e-08
CCGCG	TTAAT	60	26.5	2.3	1.7e-08
CGCGC	TTTTA	81	40.7	2.0	1.7e-08
CGCGC	TTTTG	90	47.0	1.9	1.7e-08
GCGGG	TTAAA	198	129.9	1.5	1.7e-08
CGCGC	TATTT	78	38.7	2.0	1.8e-08
CCCCA	ACCCT	978	815.6	1.2	1.9e-08
CGCCG	TATTT	84	42.9	2.0	1.9e-08
CGGCG	TAAAA	57	24.7	2.3	2.0e-08
CGGCG	TTCTT	144	87.5	1.6	2.0e-08
CCCCA	CACCC	1205	1024.4	1.2	2.1e-08
ACCCC	CCCCC	755	614.0	1.2	2.1e-08
CCCCA	CCCTC	1824	1600.0	1.1	2.2e-08
CGCCG	TTTTT	109	61.2	1.8	2.4e-08
CCGCG	GTTTT	108	60.9	1.8	3.5e-08
CCGCG	AAATA	56	24.5	2.3	3.9e-08
GCGCG	TTTTC	157	98.8	1.6	4.2e-08
GGCGG	ATTTT	312	226.5	1.4	4.4e-08
CCCCA	CCTCC	1813	1594.7	1.1	4.6e-08
CGCGG	TTTTT	140	85.7	1.6	4.8e-08
CGCGC	TTTCT	152	95.1	1.6	5.0e-08
CCGCG	TTAAA	62	28.7	2.2	5.0e-08
CGCGG	AATTA	52	22.4	2.3	6.6e-08
TCGCG	TTTAA	44	17.4	2.5	7.0e-08
GCGCC	ATTTT	186	122.9	1.5	7.2e-08
CGCGG	TTTTG	125	74.8	1.7	7.4e-08
CCGCG	ATATT	57	25.8	2.2	7.9e-08
GGCGG	ATTAA	123	73.4	1.7	8.0e-08
CGCTG	TTTAA	185	122.6	1.5	9.5e-08
GCGCC	TAAAA	98	54.9	1.8	1.0e-07
GCGCG	AAATT	57	26.0	2.2	1.0e-07
GCGCC	GTTTT	179	118.0	1.5	1.1e-07
GCGCG	TAAAA	59	27.4	2.2	1.1e-07
CGCGG	AACAA	50	21.5	2.3	1.1e-07
CGCCG	ATTTT	95	52.8	1.8	1.1e-07
CCCCA	GCCCC	1362	1179.8	1.2	1.1e-07
CGGCG	TTTAT	75	38.5	1.9	1.2e-07
GCGCC	TGTTT	243	171.0	1.4	1.2e-07
GCGCG	AAAAT	65	31.7	2.0	1.4e-07
CGCGG	TTTTA	108	62.9	1.7	1.6e-07
CGCGG	AAAAT	74	38.1	1.9	1.8e-07
CGCGG	AAATT	64	31.2	2.0	1.8e-07
TCGCG	AACAA	26	7.7	3.4	1.8e-07
CGCGG	TTGTT	118	70.9	1.7	2.1e-07
GCGCG	CTTTT	142	89.6	1.6	2.1e-07
ACCCC	ACCCT	609	492.3	1.2	2.1e-07
GGCGC	TGTTT	270	195.0	1.4	2.2e-07

GCGGG	TAAAA	179	119.8	1.5	2.8e-07
GCCGC	TTTTT	219	152.7	1.4	2.8e-07
CGCGG	TTTCT	220	153.7	1.4	2.9e-07
CCGCG	TTTTA	91	51.0	1.8	3.0e-07
ACCCC	CCCCA	1118	959.0	1.2	3.0e-07
TCGCG	TAAAA	35	12.9	2.7	3.0e-07
CGGCG	AAATA	51	22.9	2.2	3.1e-07
GGCGC	TTAAG	76	40.2	1.9	3.4e-07
CCGCG	TAATT	57	27.1	2.1	4.1e-07
CGCTG	ATTTT	274	200.1	1.4	4.2e-07
CGGCG	TTTGT	107	63.5	1.7	4.3e-07
CGGCG	TAAAT	49	21.9	2.2	4.4e-07
CCCCA	CCCCA	1784	1585.2	1.1	5.1e-07
GCGCG	TAATA	42	17.6	2.4	5.6e-07
GCGCG	ATAAT	44	18.9	2.3	5.8e-07
CCCCC	TCCCC	1058	907.3	1.2	5.8e-07
CGGCG	ATAAT	41	17.0	2.4	6.2e-07
CGGGC	ATTTT	282	207.9	1.4	6.2e-07
TCGCG	TTTTG	56	26.9	2.1	6.3e-07
CCGCG	TTGTA	57	27.6	2.1	6.5e-07
CCCCC	CTCCC	1161	1004.2	1.2	7.3e-07
CGCGG	GTTTA	53	25.0	2.1	7.4e-07
CGCGG	AAATA	61	30.4	2.0	7.4e-07
GCGCT	TATTT	173	116.9	1.5	7.5e-07
TCGCG	TTTTT	64	32.6	2.0	7.6e-07
GGCGG	TTTTT	346	263.9	1.3	7.9e-07
CCGCG	TATTT	86	48.6	1.8	8.5e-07
CGGCG	TTTTC	143	92.9	1.5	8.8e-07
CCGCG	TGTTT	140	90.5	1.5	8.9e-07
CCCCA	CTCCC	1920	1718.0	1.1	9.0e-07
CGCGC	TTTAA	58	28.6	2.0	9.4e-07

Table S2. Significant 5-mer Co-Occurrences in Constitutive Human Introns. Pairs are listed with the 5-mer occurring in the first 80 nt of the intron (column 1), the 5-mer occurring in the last 80 nt of the intron (column 2), the observed number of co-occurrences in the constitutive human dataset (column 3), the expected number calculated by coCOA (column 4), the ratio of observed to expected (column 5), and the Poisson P-value calculated by coCOA. Pairs are sorted by P-value.

5' SS	3' SS	Observed	Expected	Obs/Exp	P-value
GTATCC	CCTTAA	53	12.4	4.3	1.3e-17
TATCCT	CCTTAA	72	22.3	3.2	6.7e-17
ATCCTT	CCTTAA	78	26.5	2.9	4.2e-16
TATCCT	CTTAAC	61	20.2	3.0	2.2e-13
TATCCT	CCTTGA	56	19.9	2.8	2.6e-11
GTATCC	TCCTTA	53	18.9	2.8	9.9e-11
GGGCGC	TTAAAA	37	10.7	3.4	3.1e-10
GTATCC	CTTAAC	38	11.3	3.4	3.6e-10
ATCCTT	CTTAAC	58	24.0	2.4	3.0e-09
TATCCT	TCCTTA	72	33.0	2.2	3.3e-09
GGGCGC	TTTTAT	29	8.1	3.6	1.1e-08
CCCCTC	ACCCTC	151	93.3	1.6	2.6e-08
TATCCT	CTTGAC	46	18.4	2.5	5.0e-08
ATCCTT	TCCTTG	82	42.5	1.9	5.0e-08
CGGGCG	ATTTTT	40	14.9	2.7	5.9e-08

Table S3. Significant 6-mer Co-Occurrences in Constitutive Human Introns.

Pairs are listed with the 6-mer occurring in the first 80 nt of the intron (column 1), the 6-mer occurring in the last 80 nt of the intron (column 2), the observed number of co-occurrences in the constitutive human dataset (column 3), the expected number calculated by coCOA (column 4), the ratio of observed to expected (column 5), and the Poisson P-value calculated by coCOA. Pairs are sorted by P-value.

5' SS	3' SS	Observed	Expected	Obs/Exp	P-value
CGCG	TTAA	587	463.6	1.3	2.0e-08
CGCG	ATAT	428	336.2	1.3	8.7e-07
GCGG	ATTT	1311	1145.3	1.1	9.0e-07
GCGT	TTTA	610	500.6	1.2	1.2e-06
GCGT	AATT	464	369.5	1.3	1.2e-06
CGGC	ATAT	555	451.5	1.2	1.4e-06
CGCG	ATTT	793	669.3	1.2	1.8e-06
CGGC	ATTT	1047	905.0	1.2	2.1e-06
CGCG	TTTA	704	590.2	1.2	2.9e-06
GGCG	AATT	823	699.5	1.2	2.9e-06
CGCG	TATT	613	508.1	1.2	3.5e-06
CGCG	TTGT	750	633.9	1.2	3.9e-06
GCCG	TATT	867	741.7	1.2	4.0e-06
CGCG	AATA	433	346.3	1.2	4.0e-06
TCGC	ATTA	289	219.4	1.3	4.1e-06
GCGG	TTTA	1150	1005.3	1.1	4.2e-06
CGGC	TATT	804	684.3	1.2	4.5e-06
GCGG	ATAT	679	569.9	1.2	4.9e-06
CCGG	GTAT	555	457.2	1.2	5.2e-06
CGGC	TTTT	1244	1095.4	1.1	5.8e-06
GGCG	ATTT	1229	1082.2	1.1	6.6e-06
CGCG	TTTT	933	806.6	1.2	7.5e-06
CGGC	TTGT	994	863.8	1.2	8.1e-06
GCCG	TTGT	1065	930.6	1.1	8.7e-06
TCGC	TTGT	525	432.3	1.2	8.7e-06
GGCG	TTTT	1471	1312.6	1.1	9.3e-06
CCGC	ATTT	999	871.2	1.1	1.2e-05
GCGG	TTTT	1549	1388.9	1.1	1.2e-05
GCGG	TATT	991	864.4	1.1	1.3e-05
CGCT	AAAT	529	438.0	1.2	1.3e-05
GCGG	TTTG	1344	1196.4	1.1	1.5e-05

Table S4. Significant 4-mer Co-Occurrences Flanking Alternatively Spliced

Human Exons. Pairs are listed with the 4-mer occurring in the first 80 nt of the upstream intron (column 1), the 4-mer occurring in the last 80 nt of the downstream intron (column 2), the observed number of co-occurrences in the human alternative exon dataset (column 3), the expected number calculated by coCOA (column 4), the ratio of observed to expected (column 5), and the Poisson P-value calculated by coCOA. Pairs are sorted by P-value.

5' SS	3' SS	Observed	Expected	Obs/Exp	P-value
CGCTG	GTTAT	70	34.2	2.0	5.5e-08
GCGGC	TTTTG	290	210.1	1.4	1.0e-07
CGGCG	TTTTG	207	142.1	1.5	2.0e-07
GCGGC	TATGT	146	93.6	1.6	3.3e-07
CGGCG	TTGTT	190	129.8	1.5	4.5e-07
CCGGC	TTTGA	154	100.7	1.5	5.0e-07
CGGCG	TACAT	76	41.1	1.8	7.4e-07

Table S5. Significant 4-mer Co-Occurrences Flanking Alternatively Spliced

Human Exons. Pairs are listed with the 4-mer occurring in the first 80 nt of the upstream intron (column 1), the 4-mer occurring in the last 80 nt of the downstream intron (column 2), the observed number of co-occurrences in the human alternative exon dataset (column 3), the expected number calculated by coCOA (column 4), the ratio of observed to expected (column 5), and the Poisson P-value calculated by coCOA. Pairs are sorted by P-value.