

Motif	Consensus sequence*	Observed	Expected**	Enrichment	P-value
Human minisatellite conserved sequence/chi-like element	GCWGGWGG	4	1.1	3.54	0
DNA polymerase beta frameshift hotspot	ACCCWR	26	12.0	2.16	0
Deletion hotspot consensus sequence	TGRRKM	66	53.3	1.24	0.00001
Immunoglobulin heavy chain class switch repeat	GAGCT	15	9.8	1.53	0.00035
Translin target site	ATGCAG	6	3.3	1.84	0.00056
Immunoglobulin heavy chain class switch repeat	GGGCT	5	9.1	0.55	0.00275
Heptamer recombination signal	CACAGTG	0	1.3	0	0.0045
Vaccinia topo I consensus cleavage site	YCCTT	37	30.0	1.24	0.00501
Immunoglobulin heavy chain class switch repeat	TGAGC	17	13.1	1.29	0.02158
DNA polymerase alpha frameshift hotspot	CTGGCG	1	0.4	2.63	0.03029
Immunoglobulin heavy chain class switch repeat	GGGGT	11	8.5	1.3	0.0528
DNA polymerase alpha pause site core sequence	ACG	31	25.4	1.22	0.05872
DNA polymerase alpha/beta frameshift hotspot	TGGNGT	15	12.4	1.21	0.07057
DNA polymerase alpha pause site core sequence	GCS	149	137.6	1.08	0.10412
Vertebrate topo II consensus cleavage site	RNYNNCNGYNGKTNINY	0	0.1	0	0.1833
Murine parvovirus recombination hotspot	CTWTTY	28	25.2	1.11	0.19076
DNA polymerase arrest site	WGGAG	38	34.7	1.1	0.22032
Vertebrate/plant topo I consensus cleavage site	GTY	286	297.6	0.96	0.22833
Nonamer recombination signal	ACAAAAACC	0	0.2	0	0.27433
Chi element	GCTGGTGG	0	0.2	0	0.28396
Vertebrate/plant topo I consensus cleavage site	RAT	354	364.5	0.97	0.31916
DNA polymerase alpha/beta frameshift hotspot	ACCCCA	4	3.3	1.21	0.36606
DNA polymerase beta frameshift hotspot	TTTT	108	112.5	0.96	0.39429
Drosophila topo II consensus cleavage site	GTNWAYATTNATNNR	0	0.1	0	0.43717
Immunoglobulin heavy chain class switch repeat	TGGGG	14	13.1	1.07	0.57683
Translin target site	GCCCWSSW	3	2.7	1.12	0.58301
Human hypervariable minisatellite core sequence	GGGCAGGANG	0	0.1	0	0.66794
Murine MHC deletion hotspot*	CAGR	134	131.7	1.02	0.72595
Murine MHC recombination hotspot	CAGRCAGR	1	1.2	0.87	0.74703
Vertebrate/plant topo I consensus cleavage site	CAT	196	198.0	0.99	0.77397
DNA polymerase alpha frameshift hotspot	TCCCCC	2	2.2	0.91	0.78608
Vertebrate/plant topo I consensus cleavage site	CTY	428	424.4	1.01	0.8216
DNA polymerase alpha pause site core sequence	GAG	197	195.1	1.01	0.84906
Murine LTR recombination hotspot	TGGAAATCC	0	0.0	0	1
Human hypervariable minisatellite recombination sequence	AGAGGTGGGCAGGTGG	0	0.0	N/A	N/A
Human hypervariable minisatellite core sequence	GGAGGTGGGCAGGARG	0	0.0	N/A	N/A
Marine transposon-like element (3' end)	GAAAATGAAGCTATTTACCCAGGA	0	0.0	N/A	N/A

*From Abeyasinghe et al 2003, and references therein

**Based on occurrence in 100 matched control sequences

Supplementary Table 5. Analysis of motifs implicated in formation or repair of double stranded DNA breaks. Breakpoint and control sequences were searched for 37 consensus sequences representing 23 sequence motifs from Abeyasinghe *et al*¹³. Significant enrichment was determined by two-tailed Fisher Exact tests. The stringent Bonferroni correction for multiple testing (with $\alpha = 0.05/37 = 0.0014$) was applied.