

**Supplementary Material for “Genomic features that predict allelic imbalance in humans are negatively correlated with gene density, expression evenness, and selective constraint”**

Tung, J, Fédrigo, O, Haygood, R, Mukherjee, S, Wray, GA

**Table S1: Factor weights for the four factors produced from the NMF analysis. Features are ordered by decreasing weight for the AI factor.**

<b>Feature ID</b>	<b>AI factor</b>	<b>Factor 2</b>	<b>Factor 3</b>	<b>Factor 4</b>
proportion of C-T SNPs in 2 kb up relative to total num SNPs	0.4797	0	0.05023	0.05527
proportion of transitions in 2 kb up relative to total num of SNPs	0.43752	0.17686	0.19444	0.25054
number of C-T SNPs in 2kb flanking	0.39944	0.02578	0	0.16425
C-T SNPs per bp in 2kb flanking	0.39944	0.02578	0	0.16425
total snps per bp in 2 kb up	0.38989	0.01193	0.03526	0.04173
number of C-G SNPs in 2 kb up	0.38523	0	0	0
C-G SNPs per bp in 2 kb up	0.38523	0	0	0
C-G SNPs per bp in ccds plus 10kb flanking	0.37558	0.0774	0.0255	0.14676
number of A-G SNPs in 10 kb up	0.36722	0.08257	0.03047	0.1178
A-G SNPs per bp in 10 kb up	0.36722	0.08257	0.03047	0.1178
number of transversions in 10 kb up	0.3599	0.02407	0.0341	0.01629
transversions per bp 10 kb up	0.3599	0.02407	0.0341	0.01629
number of A-G SNPs in 10kb flanking	0.35711	0.13807	0.02004	0.18405
A-G SNPs per bp in 10kb flanking	0.35711	0.13807	0.02004	0.18405
transversions per bp 5 kb up	0.34326	0.0285	0.02865	0
number of total SNPS in 10 kb up	0.3432	0.0496	0.02818	0.0633
total snps per bp in 10 kb up	0.3432	0.0496	0.02818	0.0633
number of transitions in 10 kb up	0.33596	0.06312	0.02774	0.08668
transitions per bp in 10 kb up	0.33596	0.06312	0.02774	0.08668
number of C-G SNPs in 10kb flanking	0.32675	0.04723	0.03136	0.1032
C-G SNPs per bp in 10kb flanking	0.32675	0.04723	0.03136	0.1032
number of A-T SNPs in 10kb flanking	0.32046	0.05435	0	0
A-T SNPs per bp in 10kb flanking	0.32046	0.05435	0	0
number of A-G SNPs in 5kb flanking	0.30497	0.10462	0.01241	0.15311
A-G SNPs per bp in 5kb flanking	0.30497	0.10462	0.01241	0.15311
number of transversions in 5 kb down	0.28294	0.137	0.04818	0.08995
number of A-T SNPs in 10 kb up	0.26877	0	0	0
A-T SNPs per bp in 10 kb up	0.26877	0	0	0
proportion of divergent sites that (no gaps) are transversions in ccds + 10 kb flank	0.26002	0.48757	0.32599	0.39974
proportion of transitions in 2 kb down relative to total num of SNPs	0.25949	0.35319	0.14243	0.45487
number of C-G SNPs in 10 kb up	0.2586	0	0.02511	0
C-G SNPs per bp in 10 kb up	0.2586	0	0.02511	0
total snps per bp in 5 kb down	0.25821	0.12968	0.01153	0.17439
number of C-T SNPs in 2 kb up	0.25542	0	0	0.0098
C-T SNPs per bp in 2 kb up	0.25542	0	0	0.0098
proportion of divergent sites that (no gaps) are transversions in 10 kb flank	0.25423	0.46463	0.30301	0.38924
proportion of divergent sites that (no gaps) are transversions in 10 kb down	0.24007	0.42568	0.27452	0.35358
frequency of AAGGA in 5' flanking	0.2347	0.32837	0.13528	0.26699
transitions per bp in the CCDS	0.23439	0.07959	0.05169	0.14207
frequency of CAATC in 5' flanking	0.22377	0.31953	0.2924	0.10595

frequency of TTTC in 5' flanking	0.22363	0.42652	0.15209	0.22098
proportion of CT divergent sites relative to total num divergent sites (no gaps) in 5 kb flank	0.22103	0.4258	0.34228	0.40568
frequency of TCATT in 5' flanking	0.2196	0.44858	0.11839	0.03345
frequency of ATCCT in 5' flanking	0.21787	0.28418	0.26105	0.25126
AT divergent sites per bp in 10 kb flank	0.21753	0.25208	0.20607	0.12411
AT divergent sites per bp in 5 kb flank	0.21701	0.23544	0.22237	0.11264
frequency of CAACC in 5' flanking	0.21697	0.17839	0.2529	0.28675
frequency of AAGAG in 5' flanking	0.21673	0.29918	0.17583	0.22622
proportion of transitions in introns relative to total number of SNPs	0.21599	0.39823	0.2791	0.43817
frequency of ACATC in 5' flanking	0.21561	0.28283	0.11804	0.15084
number LINEs in 10 kb downstream	0.21555	0.14886	0.15893	0.07301
frequency of CAATG in 5' flanking	0.2154	0.28884	0.13303	0.14724
frequency of AAATG in 5' flanking	0.214	0.5679	0.14589	0.06679
frequency of CAGAT in 5' flanking	0.21393	0.39444	0.209	0.26759
proportion of CT divergent sites relative to total num divergent sites (no gaps) in 10 kb up	0.21305	0.42812	0.31515	0.40676
AC divergent sites per bp in 10 kb down	0.21279	0.17923	0.17411	0.2514
proportion of divergent sites that (no gaps) are transversions in 5 kb down	0.21243	0.39299	0.2408	0.31162
number of transitions in 2 kb down	0.21054	0.09345	0.00672	0.19206
transitions per bp in 2 kb down	0.21054	0.09345	0.00672	0.19206
proportion of divergent transversions per bp in 5 kb flank	0.2102	0.18633	0.20466	0.28708
frequency of ATTCC in 5' flanking	0.20875	0.31323	0.13368	0.16122
frequency of ATTGC in 5' flanking	0.20701	0.40654	0.2361	0.08142
frequency of ACTCA in 5' flanking	0.20674	0.34055	0.23947	0.2259
frequency of CATT A in 5' flanking	0.20554	0.44077	0.08044	0.00889
proportion of transitions in 5 kb down relative to total num of SNPs	0.20542	0.44051	0.23358	0.46361
proportion of transitions in the CCDS relative to total number of SNPs	0.20281	0.40762	0.2805	0.49304
frequency of CTATT in 5' flanking	0.20202	0.39281	0.09777	0.0008
frequency of CAACT in 5' flanking	0.20148	0.35428	0.14032	0.16113
frequency of ATGGT in 5' flanking	0.20074	0.36154	0.36094	0.1368
number of total SNPs in exons	0.20033	0.06614	0.01679	0.09354
frequency of ACTCT in 5' flanking	0.19894	0.35519	0.2082	0.2542
frequency of GCAAA in 5' flanking	0.19813	0.35402	0.13191	0.12823
frequency of TTCTC in 5' flanking	0.198	0.31551	0.1965	0.21531
frequency of CCAGA in 5' flanking	0.19752	0.15347	0.17942	0.43537
frequency of CCTAC in 5' flanking	0.1953	0.21523	0.17856	0.30115
frequency of TACCC in 5' flanking	0.19519	0.26548	0.14338	0.22891
frequency of CAAAT in 5' flanking	0.19457	0.47702	0.13454	0.05755
number of transversions in exons	0.19414	0.02519	0	0.00238
frequency of ATGAT in 5' flanking	0.19394	0.45193	0.15219	0.0524
number of A-G SNPs in 5 kb down	0.19378	0.11087	0	0.13174
A-G SNPs per bp in 5 kb down	0.19378	0.11087	0	0.13174
frequency of GAGTT in 5' flanking	0.19367	0.37888	0.31965	0.18281
frequency of CTTTG in 5' flanking	0.1931	0.39952	0.20143	0.20848
frequency of AAGCA in 5' flanking	0.19266	0.42989	0.19003	0.19953
frequency of CTAGA in 5' flanking	0.19202	0.28313	0.12486	0.14653
frequency of TCAAC in 5' flanking	0.19201	0.3015	0.12173	0.12347
A-T SNPs per bp in ccds plus 10kb flanking	0.19187	0.0776	0	0.01891

frequency of CAATA in 5' flanking	0.19117	0.39352	0.12945	0.01389
proportion of divergent sites that (no gaps) are transversions in ccds + 2kb flank	0.19108	0.38968	0.25964	0.31957
frequency of AGACT in 5' flanking	0.19037	0.29526	0.2387	0.24683
proportion of transitions in introns plus 5 kb flanking relative to total num of SNPs	0.19035	0.42522	0.31923	0.47136
frequency of AAGCT in 5' flanking	0.18954	0.36268	0.16976	0.22066
frequency of TATCT in 5' flanking	0.18912	0.41099	0.08929	0.04269
frequency of TGCTC in 5' flanking	0.18893	0.15506	0.16521	0.39228
frequency of TTTGC in 5' flanking	0.18853	0.43565	0.11153	0.16748
frequency of CTGTT in 5' flanking	0.18606	0.47336	0.21297	0.20718
proportion of transitions in ccds plus 5kb flanking relative to total num of SNPs	0.18567	0.42691	0.3223	0.47817
frequency of GATTC in 5' flanking	0.18506	0.38253	0.25288	0.16434
frequency of ACAAG in 5' flanking	0.1849	0.3625	0.18899	0.20155
frequency of TCTCC in 5' flanking	0.18388	0.14713	0.21218	0.39802
proportion of GT divergent sites relative to total num divergent sites (no gaps) in ccds + 10 kb flank	0.18264	0.42611	0.22959	0.30141
frequency of AGCTC in 5' flanking	0.18233	0.10802	0.18808	0.37578
frequency of ACATT in 5' flanking	0.18178	0.47505	0.10564	0.04506
frequency of AGAAT in 5' flanking	0.18143	0.47416	0.2005	0.07835
frequency of CTCTT in 5' flanking	0.18016	0.26647	0.13461	0.15942
frequency of GAGAG in 5' flanking	0.17887	0.13414	0.10163	0.31378
AT divergent sites per bp in 10 kb up	0.17869	0.21893	0.20785	0.12292
proportion of CT divergent sites relative to total num divergent sites (no gaps) in 5 kb up	0.17863	0.37844	0.28666	0.31929
frequency of TGGTA in 5' flanking	0.17668	0.35763	0.14071	0.07038
frequency of TTTTC in 5' flanking	0.17636	0.51201	0.14368	0.0761
AG divergent sites per bp in introns	0.1761	0.18147	0.15112	0.2811
frequency of AATGG in 5' flanking	0.17607	0.40113	0.1716	0.15005
frequency of ATTGA in 5' flanking	0.17504	0.49846	0.1016	0.02851
percentage LINEs in 10 kb downstream	0.1739	0.09815	0.05577	0.01557
frequency of TAGCA in 5' flanking	0.17298	0.3998	0.11421	0.08604
AT divergent sites per bp in 5 kb down	0.17279	0.18203	0.14414	0.06525
frequency of ATCTC in 5' flanking	0.17262	0.25617	0.3355	0.17549
frequency of ACTTG in 5' flanking	0.17243	0.33314	0.22228	0.17279
proportion of C-G SNPs in exons relative to total num SNPs	0.17154	0	0	0
frequency of TGATT in 5' flanking	0.17153	0.48771	0.14207	0.04225
frequency of CCTTC in 5' flanking	0.17146	0.17209	0.14078	0.37069
frequency of GGTAT in 5' flanking	0.16961	0.33316	0.13921	0.05453
frequency of TTGCT in 5' flanking	0.16941	0.39896	0.15834	0.15641
frequency of ATAGG in 5' flanking	0.16883	0.32897	0.20187	0.0871
frequency of GGCTT in 5' flanking	0.16759	0.2265	0.15101	0.36076
frequency of GGTA in 5' flanking	0.16637	0.41913	0.15359	0.11846
frequency of CAAGT in 5' flanking	0.16594	0.38665	0.27453	0.19997
frequency of GTTCC in 5' flanking	0.16575	0.18913	0.1385	0.34889
frequency of GAAGG in 5' flanking	0.16496	0.16374	0.12256	0.43041
frequency of TAATG in 5' flanking	0.16471	0.48239	0.09444	0.01624
number of G-T SNPs in 10 kb up	0.16461	0.02571	0.05762	0
G-T SNPs per bp in 10 kb up	0.16461	0.02571	0.05762	0
frequency of TTTCT in 5' flanking	0.16417	0.40462	0.141	0.07531
proportion of transitions in 10 kb up relative to total num of SNPs	0.16404	0.46327	0.26795	0.46645

frequency of AAGAC in 5' flanking	0.16397	0.30062	0.20202	0.19373
frequency of CACTG in 5' flanking	0.16364	0.21169	0.35423	0.40543
frequency of ACTGA in 5' flanking	0.16361	0.36228	0.12119	0.19359
frequency of CAATT in 5' flanking	0.16361	0.40527	0.13547	0.02415
proportion of C-G SNPs in 2 kb up relative to total num SNPs	0.16351	0	0	0
proportion of AT divergent sites relative to total num divergent sites (no gaps) in 5 kb flank	0.1626	0.35076	0.19642	0.08905
frequency of CTACC in 5' flanking	0.16183	0.23515	0.17438	0.20079
frequency of GATCC in 5' flanking	0.16116	0.18478	0.35412	0.30372
frequency of AGCAT in 5' flanking	0.16037	0.2978	0.11949	0.11579
percentage LINE1 in 5 kb downstream	0.15946	0.00302	0.04875	0
proportion of GT divergent sites relative to total num divergent sites (no gaps) in 10 kb up	0.15943	0.31447	0.16895	0.22803
frequency of TCAAG in 5' flanking	0.15941	0.33292	0.37135	0.19598
AT divergent sites per bp in 10 kb down	0.15939	0.18639	0.12879	0.085
frequency of TAGGG in 5' flanking	0.15848	0.22748	0.14394	0.29717
frequency of GATAC in 5' flanking	0.15802	0.35868	0.07785	0.0669
frequency of TGAAA in 5' flanking	0.158	0.46259	0.15218	0.05876
frequency of GAGCT in 5' flanking	0.15785	0.13736	0.17915	0.41239
frequency of AGAGC in 5' flanking	0.15752	0.15152	0.15109	0.40121
frequency of TCATA in 5' flanking	0.15698	0.44387	0.1215	0.00538
frequency of TCTCT in 5' flanking	0.15694	0.20899	0.18484	0.17692
frequency of CACTT in 5' flanking	0.15652	0.3331	0.27312	0.19447
proportion of CG divergent sites relative to total num divergent sites (no gaps) in 10 kb down	0.1552	0.23323	0.20782	0.37536
frequency of GTTGA in 5' flanking	0.15504	0.3309	0.13054	0.10398
percentage LINEs in 5 kb downstream	0.1549	0.01665	0.07093	0
frequency of CTTGA in 5' flanking	0.15474	0.34677	0.37235	0.16635
frequency of TAAGC in 5' flanking	0.15468	0.34755	0.15202	0.14525
number of synonymous snps	0.15466	0.05934	0.006	0.07503
proportion of G-T SNPs in introns plus 5 kb flanking relative to total num SNPs	0.1544	0.18558	0.13282	0.07462
number LINE1 in 10 kb downstream	0.15438	0.08457	0.08931	0.01202
frequency of TTAAA in 5' flanking	0.15399	0.53173	0.09392	0.01086
frequency of AATGA in 5' flanking	0.15376	0.38434	0.09424	0.06749
GC content in ccds	0.15346	0.30793	0.29313	0.46538
frequency of GAACA in 5' flanking	0.15322	0.32035	0.09363	0.15142
frequency of TGTCC in 5' flanking	0.15301	0.11913	0.1106	0.36095
frequency of ACTAG in 5' flanking	0.15239	0.32627	0.12122	0.09951
frequency of TTAGC in 5' flanking	0.15232	0.3712	0.31409	0.11636
frequency of TTGAC in 5' flanking	0.15216	0.35472	0.148	0.12699
frequency of CACTC in 5' flanking	0.15126	0.11663	0.24109	0.3462
proportion of AG divergent sites relative to total num divergent sites (no gaps) in 2 kb down	0.15123	0.31052	0.21385	0.31602
number of G-T SNPs in 5 kb up	0.14948	0.0138	0.01247	0
frequency of GCTGA in 5' flanking	0.1488	0.16052	0.31558	0.38175
frequency of AGATC in 5' flanking	0.1485	0.24337	0.25965	0.17054
frequency of AATTC in 5' flanking	0.14838	0.38712	0.12915	0.04679
frequency of ATCAT in 5' flanking	0.14799	0.29624	0.10389	0.03649
frequency of TAGGT in 5' flanking	0.14751	0.32645	0.1243	0.10062
proportion of AT divergent sites relative to total num divergent sites (no gaps) in 10 kb	0.14745	0.28069	0.19183	0.10838

up				
frequency of TAATA in 5' flanking	0.1466	0.46731	0.12087	0
frequency of TCTTT in 5' flanking	0.14646	0.36922	0.14253	0.0629
frequency of AGTGA in 5' flanking	0.14561	0.30019	0.31323	0.17944
frequency of TAGGC in 5' flanking	0.14487	0.24096	0.24133	0.19187
percentage LINE1 in 10 kb downstream	0.14373	0.07502	0.03214	0.00521
frequency of CTCTG in 5' flanking	0.14356	0.13426	0.21257	0.39804
frequency of TTATC in 5' flanking	0.14333	0.47505	0.11757	0.01849
frequency of TTTGG in 5' flanking	0.14307	0.43759	0.21919	0.17436
frequency of CTTGT in 5' flanking	0.14182	0.32323	0.13599	0.14933
frequency of GTACA in 5' flanking	0.14021	0.28078	0.11771	0.09952
frequency of TCACC in 5' flanking	0.13906	0.12977	0.28027	0.28998
frequency of CTTAC in 5' flanking	0.13877	0.46312	0.12736	0.11285
frequency of AACTG in 5' flanking	0.13871	0.36081	0.11754	0.16253
frequency of AAGTT in 5' flanking	0.13826	0.54164	0.12256	0.06489
transversions per bp in exons	0.13817	0.01133	0.01267	0
frequency of GTGCA in 5' flanking	0.1376	0.16743	0.26834	0.34006
frequency of TATTC in 5' flanking	0.13684	0.37699	0.08687	0.01732
proportion of A-C SNPs in ccds plus 10kb flanking relative to total num SNPs	0.1366	0.17682	0.14171	0.09695
frequency of ATAGC in 5' flanking	0.13659	0.37349	0.1591	0.05706
proportion of AT divergent sites relative to total num divergent sites (no gaps) in 5 kb up	0.13643	0.28078	0.18812	0.09786
proportion of divergent sites that (no gaps) are transversions in 2 kb down	0.13528	0.28997	0.19595	0.236
proportion of AG divergent sites relative to total num divergent sites (no gaps) in introns	0.13491	0.22326	0.15707	0.21053
frequency of GTACC in 5' flanking	0.1349	0.27101	0.1011	0.17268
frequency of CCTTG in 5' flanking	0.13458	0.19015	0.18106	0.36306
number of C-G SNPs in exons	0.13438	0	0	0
number of A-T SNPs in 2 kb up	0.13438	0	0	0
A-T SNPs per bp in 2 kb up	0.13438	0	0	0
proportion of divergent transversions per bp in 5 kb up	0.1341	0.14399	0.14318	0.20389
frequency of CTATA in 5' flanking	0.13387	0.37387	0.10871	0.01968
frequency of TTGGT in 5' flanking	0.13383	0.35727	0.14115	0.13314
frequency of TTACTION in 5' flanking	0.13245	0.43404	0.08252	0.02035
frequency of ACTAA in 5' flanking	0.13239	0.40697	0.22622	0.06005
proportion of AG divergent sites relative to total num divergent sites (no gaps) in exons	0.13229	0.2436	0.13405	0.18912
number of divergent transitions in 5 kb flank	0.13179	0.10184	0.14617	0.197
frequency of TCTTG in 5' flanking	0.13173	0.34382	0.23537	0.16084
frequency of ATAAC in 5' flanking	0.13133	0.35667	0.10423	0
frequency of GTTAC in 5' flanking	0.1312	0.38766	0.15762	0.06392
frequency of GTCTG in 5' flanking	0.13073	0.13145	0.1322	0.32892
proportion of A-T SNPs in 5 kb up relative to total num SNPs	0.13002	0	0	0
frequency of CTTGC in 5' flanking	0.12952	0.2539	0.1739	0.32864
CT divergent sites per bp in ccds	0.12946	0.10819	0.1333	0.21906
frequency of TGCTG in 5' flanking	0.12937	0.18553	0.20049	0.37373
frequency of GATCA in 5' flanking	0.1289	0.17943	0.25995	0.14898
frequency of AAAAA in 5' flanking	0.12767	0.35222	0.20322	0.06627
frequency of TCAGT in 5' flanking	0.12746	0.31341	0.11971	0.14739
frequency of TTGTT in 5' flanking	0.12736	0.38103	0.14762	0.02814

proportion of G-T SNPs in ccds plus 5kb flanking relative to total num SNPs	0.12708	0.2487	0.17705	0.09465
AT divergent sites per bp in 5 kb up	0.12708	0.14232	0.15874	0.08432
frequency of TTGAA in 5' flanking	0.12653	0.45706	0.18693	0.04897
frequency of TGCCT in 5' flanking	0.12638	0.16263	0.30211	0.33408
frequency of GCCTT in 5' flanking	0.12622	0.22627	0.19392	0.36108
frequency of TTGGG in 5' flanking	0.12547	0.24758	0.29887	0.37329
frequency of TTGGC in 5' flanking	0.12538	0.27018	0.28301	0.26324
proportion of A-G SNPs in introns plus 2 kb flanking relative to total num SNPs	0.12526	0.28223	0.20153	0.33992
frequency of GTAGA in 5' flanking	0.12353	0.39173	0.3133	0.18892
frequency of ACCAC in 5' flanking	0.1222	0.14265	0.30369	0.22604
frequency of GCAAC in 5' flanking	0.12213	0.23101	0.32594	0.19994
frequency of AGTGT in 5' flanking	0.12177	0.30713	0.09568	0.19158
C-G SNPs per bp in introns plus 5 kb flanking	0.12159	0.01692	0.005	0.03908
frequency of ATAAT in 5' flanking	0.12117	0.39871	0.10998	0
frequency of TATCA in 5' flanking	0.12086	0.34048	0.07482	0.00277
frequency of TAGTG in 5' flanking	0.1203	0.46006	0.20251	0.09452
frequency of CCAAG in 5' flanking	0.11979	0.11198	0.18685	0.30179
frequency of GTAAG in 5' flanking	0.11946	0.38691	0.15229	0.10798
proportion of AC divergent sites relative to total num divergent sites (no gaps) in ccds	0.11898	0.23815	0.15901	0.18489
frequency of GTATA in 5' flanking	0.11892	0.40926	0.09131	0
frequency of TGACC in 5' flanking	0.11808	0.12073	0.21313	0.28266
frequency of GAGTA in 5' flanking	0.11768	0.36045	0.28646	0.10328
frequency of ATTAT in 5' flanking	0.11755	0.45168	0.12488	0
frequency of GGTAG in 5' flanking	0.11723	0.24668	0.15384	0.22888
frequency of GGGTT in 5' flanking	0.11676	0.17989	0.32398	0.35173
frequency of GCTTG in 5' flanking	0.11516	0.23794	0.27135	0.35016
frequency of TTGTA in 5' flanking	0.11505	0.51979	0.22768	0.03069
frequency of GGATT in 5' flanking	0.115	0.40471	0.41344	0.14952
frequency of CTTAA in 5' flanking	0.11419	0.44229	0.09547	0.05104
transversions per bp introns plus 5 kb flanking	0.11402	0.028	0.00505	0.02356
frequency of TACAG in 5' flanking	0.11393	0.33658	0.37285	0.12951
frequency of TAACC in 5' flanking	0.11367	0.32321	0.13529	0.10671
frequency of CACAG in 5' flanking	0.113	0.09991	0.11743	0.30929
frequency of TACTC in 5' flanking	0.11234	0.3129	0.21327	0.08534
frequency of AGGCT in 5' flanking	0.11209	0.12689	0.41774	0.37146
number of A-C SNPs in 2 kb down	0.11136	0.03543	0	0.0087
A-C SNPs per bp in 2 kb down	0.11136	0.03543	0	0.0087
frequency of CCTGT in 5' flanking	0.11056	0.12212	0.29412	0.33205
frequency of GTAGG in 5' flanking	0.11054	0.26589	0.1605	0.25036
frequency of TAGCT in 5' flanking	0.10956	0.41251	0.38942	0.15427
frequency of CCATC in 5' flanking	0.10953	0.09017	0.15698	0.25978
frequency of CTGTG in 5' flanking	0.10893	0.12058	0.08727	0.28403
frequency of GAATG in 5' flanking	0.10773	0.24496	0.0732	0.10782
proportion of divergent sites per bp in 5 kb up	0.10753	0.09612	0.11467	0.15269
frequency of GTGAT in 5' flanking	0.10611	0.30864	0.30836	0.14893
frequency of GTCCC in 5' flanking	0.10365	0.01521	0.09635	0.38087
frequency of ACTAC in 5' flanking	0.10351	0.30585	0.24604	0.07147

frequency of TTTTG in 5' flanking	0.10332	0.45798	0.23932	0.06204
A-G SNPs per bp in introns plus 5 kb flanking	0.10296	0.02547	0.00489	0.04893
number MaLRs in 10 kb downstream	0.10279	0.05024	0.09316	0.00568
frequency of CAGTG in 5' flanking	0.10248	0.21251	0.30823	0.35101
frequency of ATGTA in 5' flanking	0.10225	0.43137	0.09823	0.03306
proportion of AT divergent sites relative to total num divergent sites (no gaps) in 2 kb flank	0.10175	0.23081	0.13294	0.05874
total % bases masked in 2 kb downstream	0.10142	0.00556	0.34395	0.00116
G-T SNPs per bp in introns plus 5 kb flanking	0.10103	0.02253	0.00546	0.01517
frequency of GCTCC in 5' flanking	0.10064	0.01907	0.08286	0.48974
frequency of AGTAG in 5' flanking	0.09989	0.30434	0.37513	0.1261
frequency of TTTGT in 5' flanking	0.0997	0.40476	0.18629	0.05767
percentage MER1 type in ceds plus 10 kb flanking	0.0978	0.14773	0.03176	0.02078
proportion of total divergent sites that are gaps, including gaps in 2 kb down	0.09739	0.32043	0.30494	0.25141
frequency of AGGTC in 5' flanking	0.09681	0.10813	0.16319	0.28209
frequency of CTCCA in 5' flanking	0.09593	0.04639	0.12698	0.23577
transitions per bp in introns plus 5 kb flanking	0.09586	0.02284	0.00547	0.04381
AT divergent sites per bp in 2 kb up	0.09565	0.06684	0.09817	0.06074
proportion of GT divergent sites relative to total num divergent sites (no gaps) in 5 kb up	0.09554	0.22303	0.12297	0.15606
proportion of AT divergent sites relative to total num divergent sites (no gaps) in 2 kb up	0.09509	0.14073	0.10514	0.06971
A-T SNPs per bp in introns plus 5 kb flanking	0.09508	0.0318	0	0.007
frequency of TTATT in 5' flanking	0.09464	0.37171	0.12713	0.00283
frequency of CCACC in 5' flanking	0.09383	0.0372	0.31635	0.43437
frequency of TCCCC in 5' flanking	0.09383	0.04484	0.05098	0.3466
proportion of A-C SNPs in 2 kb down relative to total num SNPs	0.09345	0.06614	0	0.02342
percentage MIRs in 10 kb downstream	0.09336	0.05089	0.08106	0.04886
proportion of CG divergent sites relative to total num divergent sites (no gaps) in exons	0.09211	0.05757	0.08382	0.1144
frequency of GAGAC in 5' flanking	0.09202	0.13495	0.39836	0.27748
number of divergent transitions in 5 kb up	0.09163	0.07466	0.09794	0.12582
frequency of ACGTT in 5' flanking	0.0912	0.19447	0.10355	0.18068
frequency of TCGTT in 5' flanking	0.09092	0.24197	0.07791	0.15182
A-C SNPs per bp in introns plus 2 kb flanking	0.08975	0.01608	0.00802	0.01291
frequency of CAACG in 5' flanking	0.08931	0.12177	0.07542	0.24406
proportion of C-T SNPs in 10 kb down relative to total num SNPs	0.08918	0.21869	0.12381	0.20918
proportion of GT divergent sites relative to total num divergent sites (no gaps) in ceds	0.08846	0.1995	0.10925	0.15981
frequency of ATGTG in 5' flanking	0.08676	0.17229	0.05106	0.09944
number MIRs in 5 kb downstream	0.08379	0.02944	0.12171	0.02769
frequency of CCTGG in 5' flanking	0.08285	0.01731	0.22998	0.41776
frequency of TAGCC in 5' flanking	0.08272	0.19929	0.26786	0.17023
proportion of CG divergent sites relative to total num divergent sites (no gaps) in ceds	0.0827	0.15367	0.15245	0.24329
percentage MIRs in 5 kb downstream	0.08256	0.02778	0.09857	0.01962
frequency of GCTGC in 5' flanking	0.0824	0.04441	0.06919	0.39742
number of C-T SNPs in exons	0.08199	0.02669	0.01833	0.06642
percentage LINEs in 2 kb downstream	0.07973	0	0.04148	0
proportion of C-T SNPs in exons relative to total num SNPs	0.07882	0.09232	0.12859	0.21953
frequency of TTATA in 5' flanking	0.07834	0.3846	0.07407	0
frequency of CTAGC in 5' flanking	0.07793	0.29884	0.17925	0.26547
number LINEs in 10 kb upstream	0.07772	0.22256	0.11958	0.11346

frequency of GACGT in 5' flanking	0.07668	0.06427	0.06582	0.34574
frequency of CAGGG in 5' flanking	0.07624	0.00472	0.08679	0.38215
number of nonsynonymous snps	0.07563	0.01003	0.01641	0.03392
frequency of GGTCC in 5' flanking	0.07493	0.01654	0.07185	0.37251
total % bases masked in 10 kb upstream	0.07303	0.31709	0.41593	0.1935
C-G SNPs per bp in introns plus 2 kb flanking	0.07288	0.00738	0.00087	0.01514
frequency of GCAGG in 5' flanking	0.07272	0.01945	0.14438	0.42104
frequency of ACGTG in 5' flanking	0.0722	0.04069	0.08926	0.29115
frequency of CTACA in 5' flanking	0.07106	0.11341	0.09132	0.06751
frequency of CAGCC in 5' flanking	0.07104	0.02285	0.2649	0.4044
proportion of C-T SNPs in 5 kb up relative to total num SNPs	0.07103	0.25252	0.12248	0.19947
frequency of CGAAA in 5' flanking	0.07097	0.13937	0.1116	0.14615
frequency of ACCCC in 5' flanking	0.06954	0.01747	0.10233	0.32202
number LINE1 in 10 kb upstream	0.06947	0.16055	0.09441	0.04386
CG divergent sites per bp in ceds	0.06924	0.06363	0.09873	0.17809
AT divergent sites per bp in 2 kb flank	0.06735	0.06017	0.0717	0.05076
transversions per bp introns plus 2 kb flanking	0.06681	0.0103	0.00174	0.01004
frequency of GGGTG in 5' flanking	0.06679	0.02393	0.09472	0.32226
frequency of GGACC in 5' flanking	0.06497	0.01967	0.06299	0.35694
A-G SNPs per bp in introns plus 2 kb flanking	0.06476	0.00774	0.00211	0.02696
frequency of GACCC in 5' flanking	0.0643	0.00723	0.05471	0.28108
proportion of A-C SNPs in 10 kb up relative to total num SNPs	0.06406	0.08035	0.01159	0.06079
number of A-T SNPs in exons	0.06239	0.0355	0	0.00291
proportion of C-G SNPs in introns relative to total num SNPs	0.06164	0.06312	0.04105	0.03761
frequency of GTAGT in 5' flanking	0.0615	0.34461	0.22576	0.06655
frequency of GGTGT in 5' flanking	0.06056	0.07738	0.10207	0.18919
transitions per bp in introns plus 2 kb flanking	0.05992	0.00648	0.00221	0.02178
frequency of ATCGT in 5' flanking	0.05963	0.16776	0.09735	0.20595
frequency of TACAC in 5' flanking	0.05786	0.09096	0.03545	0.05645
frequency of CGGTT in 5' flanking	0.05688	0.09535	0.03665	0.23357
percentage LINE1 in 2 kb downstream	0.05661	0	0.01148	0
frequency of TCGAT in 5' flanking	0.05624	0.16865	0.15185	0.10768
number MIRs in 2kb upstream	0.05518	0.0355	0.07064	0.05404
frequency of CGTTT in 5' flanking	0.05461	0.16283	0.05083	0.17946
A-T SNPs per bp in introns plus 2 kb flanking	0.05455	0.01159	0	0.00296
number of divergent transversions in exons	0.05432	0.01513	0.02872	0.04404
frequency of CTCGT in 5' flanking	0.05378	0.0779	0.09805	0.22858
proportion of A-G SNPs in 10 kb up relative to total num SNPs	0.05345	0.23628	0.14349	0.26245
frequency of GGCTC in 5' flanking	0.0516	0.05018	0.22459	0.39769
frequency of CGTTC in 5' flanking	0.05102	0.0898	0.05342	0.27492
frequency of ATGCG in 5' flanking	0.05035	0.08521	0.06121	0.24873
frequency of CGAAG in 5' flanking	0.05024	0.08825	0.00664	0.27529
G-T SNPs per bp in introns plus 2 kb flanking	0.05023	0.00706	0.00154	0.0087
percentage LINE2 in 5 kb downstream	0.04973	0.01662	0.06638	0.00095
percentage LTR in 5 kb downstream	0.0489	0.00641	0.07651	0
percentage MIRs in 2 kb downstream	0.04829	0	0.0998	0
frequency of ACCCG in 5' flanking	0.04771	0	0.18077	0.37794



frequency of GGATG in 5' flanking	0.04769	0.06145	0.04803	0.15899
proportion of G-T SNPs in 10 kb down relative to total num SNPs	0.04752	0.10864	0.11718	0.03685
frequency of GGAGG in 5' flanking	0.04745	0.04075	0.19493	0.34938
frequency of AAGCG in 5' flanking	0.04727	0.0788	0.15822	0.26548
CG divergent sites per bp in introns	0.04644	0.04787	0.05588	0.14157
frequency of TATGG in 5' flanking	0.04615	0.07655	0.03499	0.0237
CG divergent sites per bp in exons	0.04589	0	0.04082	0.05963
number of G-T SNPs in exons	0.04476	0.01495	0.01634	0
proportion of G-T SNPs in introns relative to total num SNPs	0.04445	0.05303	0.03942	0.0164
frequency of GCGAC in 5' flanking	0.04384	0	0.10419	0.35243
frequency of GCCCC in 5' flanking	0.04373	0	0.02841	0.40195
frequency of CGCAC in 5' flanking	0.04368	0.01944	0.109	0.30984
number of exons in the gene	0.04311	0.16458	0.06191	0.13084
proportion of A-C SNPs in 5 kb down relative to total num SNPs	0.04238	0.02911	0.0369	0.04808
frequency of CGTGT in 5' flanking	0.04214	0.0528	0.04625	0.22645
frequency of ACACG in 5' flanking	0.04157	0.02449	0.04206	0.20316
frequency of AGTCG in 5' flanking	0.04139	0.0772	0.04636	0.30174
frequency of ACCGG in 5' flanking	0.03958	0.01247	0.03493	0.23007
percentage LTR in 10 kb downstream	0.03938	0.04172	0.08506	0
frequency of ATCCG in 5' flanking	0.03918	0.08542	0.17834	0.27633
percentage unclassified in 2kb upstream	0.03887	0	0.02785	0.02982
frequency of CGTCA in 5' flanking	0.03852	0.04769	0.02744	0.16135
percentage LINE2 in 5 kb upstream	0.03796	0.04319	0.04495	0.05462
proportion of G-T SNPs in 2 kb down relative to total num SNPs	0.03586	0.04518	0	0.0414
frequency of TCGAG in 5' flanking	0.03556	0.0714	0.16824	0.18139
proportion of G-T SNPs in introns plus 2 kb flanking relative to total num SNPs	0.03261	0.05288	0.04789	0.02076
frequency of CGAGA in 5' flanking	0.03209	0.05953	0.24954	0.19968
total % bases masked in 2kb upstream	0.03191	0.13663	0.37021	0.1015
frequency of GCGTA in 5' flanking	0.02929	0.08296	0.04362	0.13776
frequency of TCGGA in 5' flanking	0.02908	0.07741	0.04787	0.34219
number L3/CR1 in 5 kb downstream	0.02803	0	0	0.04523
frequency of GGCCC in 5' flanking	0.02701	0	0.03868	0.38452
frequency of GGTTCG in 5' flanking	0.02628	0.01486	0.07626	0.36403
frequency of CGGAT in 5' flanking	0.02613	0.05611	0.14507	0.24306
frequency of CCGAC in 5' flanking	0.02487	0.00563	0.03679	0.27188
frequency of GGACG in 5' flanking	0.02438	0	0.03114	0.30157
frequency of ACTCG in 5' flanking	0.02419	0.08191	0.15095	0.29109
frequency of GTCCG in 5' flanking	0.02387	0.01527	0.03691	0.29858
frequency of TAGCG in 5' flanking	0.02377	0.12674	0.09714	0.17394
frequency of CGGAG in 5' flanking	0.02336	0.02019	0.09872	0.3723
frequency of CTCGA in 5' flanking	0.02243	0.07375	0.22825	0.23989
frequency of CGCTG in 5' flanking	0.02225	0	0.02715	0.4069
frequency of AGCGT in 5' flanking	0.02167	0.09358	0.03991	0.21071
frequency of CGGTG in 5' flanking	0.02164	0.03782	0.12163	0.27857
number simple repeats in 2 kb downstream	0.02156	0.03977	0.01881	0.02932
percentage ERV classII in 2kb upstream	0.02064	0	0.02892	0.03451
percentage unclassified in 5 kb downstream	0.02008	0.01162	0.03547	0

proportion of C-G SNPs in 5 kb down relative to total num SNPs	0.01993	0.03974	0.08074	0.06388
proportion of C-G SNPs in 2 kb down relative to total num SNPs	0.01947	0.02355	0.01881	0.065
frequency of GTGTG in 5' flanking	0.01917	0.04631	0.02385	0.06993
frequency of ACGAC in 5' flanking	0.01915	0.10566	0.10811	0.23684
frequency of CCGAG in 5' flanking	0.01751	0.00358	0.21412	0.41633
frequency of CCGTA in 5' flanking	0.01721	0.12069	0.06881	0.18652
percentage unclassified in ccds	0.01702	0	0.01059	0.01542
frequency of TCGTC in 5' flanking	0.01639	0.09834	0.04524	0.30837
percentage unclassified in ccds plus 2kb flanking	0.01619	0	0.01416	0.01807
frequency of CGTCC in 5' flanking	0.01583	0.01388	0.02158	0.39855
frequency of GTCGG in 5' flanking	0.01491	0.00954	0.04157	0.26739
frequency of CGGGT in 5' flanking	0.01477	0.00701	0.14198	0.34003
frequency of CACGG in 5' flanking	0.01415	0.01826	0.07054	0.28779
frequency of CTGCG in 5' flanking	0.01395	0.0283	0.03618	0.36808
number MaLRs in ccds plus 10 kb flanking	0.0135	0.10186	0.00402	0.02129
percentage low complexity in 5 kb downstream	0.01324	0.10694	0.02245	0.03963
number of transversions in introns	0.01317	0.10387	0	0.01098
percentage unclassified in ccds plus 5 kb flanking	0.01271	0	0.04401	0.01633
frequency of CTACG in 5' flanking	0.01203	0.12961	0.11042	0.25733
frequency of CGAGG in 5' flanking	0.01186	0.01337	0.10624	0.38545
frequency of TTCGC in 5' flanking	0.01081	0.07591	0.11264	0.27502
percentage simple repeats in 5 kb upstream	0.01041	0.01564	0.00829	0.06686
frequency of GTGCG in 5' flanking	0.00951	0.02926	0.07567	0.31567
frequency of ACGGG in 5' flanking	0.00829	0.00925	0.0734	0.21662
percentage low complexity in 2 kb downstream	0.00806	0.07745	0	0.00949
percentage ERVL in 2kb upstream	0.00795	0.0164	0.00274	0
frequency of CCCGC in 5' flanking	0.00751	0.0203	0.0767	0.4499
percentage ALUs in 5 kb downstream	0.00711	0.02315	0.35254	0.03151
frequency of CCGCT in 5' flanking	0.00697	0.01398	0.03711	0.46551
proportion of G-T SNPs in 10 kb up relative to total num SNPs	0.00637	0.0755	0.08807	0.01859
frequency of TGGCG in 5' flanking	0.00637	0.02493	0.02731	0.32488
number SINEs in 10 kb upstream	0.00606	0.12324	0.42013	0.17238
frequency of CGAGC in 5' flanking	0.00564	0.05072	0.034	0.37464
proportion unalignable (gaps in human or other species) per bp in ccds + 2kb flank	0.0054	0	0.04127	0.04197
proportion unalignable (gaps in human or other species) per bp in introns	0.00523	0.00111	0.04315	0.03857
frequency of TACGA in 5' flanking	0.00479	0.18895	0.08199	0.11796
proportion unalignable (gaps in other species) per bp in 2 kb down	0.00371	0.00071	0.03753	0.01598
number ERV classI in ccds plus 2kb flanking	0.00306	0.04895	0.00119	0.00732
number ERV classI in ccds	0.00294	0.04605	0	0.00422
frequency of CCGGC in 5' flanking	0.00291	0.00572	0.05887	0.27873
percentage unclassified in ccds plus 10 kb flanking	0.00246	0	0.05516	0.00845
percentage LTR in 5 kb upstream	0.00222	0.05156	0.05414	0.04206
percentage simple repeats in 2kb upstream	0.00134	0.00734	0.00285	0.04722
number LINE1 in ccds plus 5 kb flanking	0.00085	0.11567	0	0.00995
frequency of GCGTC in 5' flanking	0.00073	0.04343	0.02028	0.3595
number LINEs in ccds plus 5 kb flanking	0.00001	0.10169	0.00102	0.01323
frequency of ACGCC in 5' flanking	0.00001	0.02942	0.26939	0.31071

total repeats (number SINEs + LINEs + LTR + DNA elements) in ccds	0	0.11274	0.00657	0.01867
percentage ALUs in 2kb upstream	0	0.01987	0.31486	0.02592
percentage L3/CR1 in 2kb upstream	0	0.00921	0	0.00334
percentage MER1 type in 2kb upstream	0	0	0.01716	0
percentage low complexity in 2kb upstream	0	0.05246	0	0.10204
number SINEs in 2 kb downstream	0	0	0.3296	0
percentage SINEs in 2 kb downstream	0	0	0.33995	0
number ALUs in 2 kb downstream	0	0	0.3225	0
percentage LTR in 2 kb downstream	0	0	0.08369	0
number MaLRs in 2 kb downstream	0	0	0.05324	0
percentage MaLRs in 2 kb downstream	0	0	0.07514	0
percentage ERV classII in 2 kb downstream	0	0	0.01066	0
percentage L3/CR1 in 5 kb upstream	0	0.03022	0	0
number MaLRs in 5 kb upstream	0	0.10089	0.03402	0.01733
percentage unclassified in 5 kb upstream	0	0	0.01819	0
number low complexity in 5 kb upstream	0	0.16661	0.01662	0.1244
percentage SINEs in 10 kb upstream	0	0.11048	0.43388	0.15884
number ALUs in 10 kb upstream	0	0.08827	0.43125	0.13562
percentage unclassified in 10 kb upstream	0	0	0.03804	0
percentage low complexity in 10 kb upstream	0	0.14137	0.04097	0.19922
percentage unclassified in 10 kb downstream	0	0.01757	0.01065	0.02368
percentage satellites in 10 kb downstream	0	0.02422	0	0.01323
number LINEs in ccds plus 2kb flanking	0	0.09702	0	0.01025
number LINE1 in ccds plus 2kb flanking	0	0.11099	0	0.00664
number L3/CR1 in ccds plus 5 kb flanking	0	0.05523	0	0.00425
number L3/CR1 in ccds plus 10 kb flanking	0	0.05878	0	0.00829
total repeats (number SINEs + LINEs + LTR + DNA elements) in ccds plus 10 kb flanking	0	0.12554	0.03174	0.03551
frequency of CGCCT in 5' flanking	0	0.06604	0.34595	0.3973
frequency of CGGCG in 5' flanking	0	0.02128	0	0.19489
frequency of CGGGC in 5' flanking	0	0.00606	0.0722	0.3526
frequency of GACGG in 5' flanking	0	0.01319	0.11014	0.18598
frequency of GCGAT in 5' flanking	0	0.07248	0.26616	0.15793
frequency of GCGCC in 5' flanking	0	0.01863	0.08467	0.2732
frequency of GCGGG in 5' flanking	0	0.01747	0.03713	0.31672
frequency of GCGGT in 5' flanking	0	0.06198	0.09682	0.33797
frequency of GGCGC in 5' flanking	0	0.00992	0.06939	0.25885
frequency of TCGCC in 5' flanking	0	0.05465	0.16983	0.34933
frequency of TCGGC in 5' flanking	0	0.03052	0.22286	0.278
proportion unalignable (gaps in other species) per bp in 10 kb flank	0	0.01313	0.09386	0.10386

Table S2. Complete list of 2,269 features used in full model SVM. Features are color-coded by feature type (gray indicates polymorphism features, lavender indicates sequence features, and blue indicates divergence features).

No.	Feature Description	No.	Feature Description	No.	Feature Description
1	number of A-C SNPs in the CCDS	758	number satellites in ccds plus 2kb flanking	1515	frequency of GGCGA in 5 kb flanking
2	number of A-G SNPs in the CCDS	759	percentage satellites in ccds plus 2kb flanking	1516	frequency of GGCGC in 5 kb flanking
3	number of A-T SNPs in the CCDS	760	number simple repeats in ccds plus 2kb flanking	1517	frequency of GGCGG in 5 kb flanking
4	number of C-G SNPs in the CCDS	761	percentage simple repeats in ccds plus 2kb flanking	1518	frequency of GGCGT in 5 kb flanking
5	number of C-T SNPs in the CCDS	762	number low complexity in ccds plus 2kb flanking	1519	frequency of GGCTA in 5 kb flanking
6	number of G-T SNPs in the CCDS	763	percentage low complexity in ccds plus 2kb flanking	1520	frequency of GGCTC in 5 kb flanking
7	number of total snps in the CCDS	764	total repeats (number SINEs + LINEs + LTR + DNA elements) in ccds plus 2kb flanking	1521	frequency of GGCTG in 5 kb flanking
8	number of transitions in the CCDS	765	GC content in ccds plus 5 kb flanking	1522	frequency of GGCTT in 5 kb flanking
9	number of transversions in the CCDS	766	total % bases masked in ccds plus 5 kb flanking	1523	frequency of GGGAA in 5 kb flanking
10	proportion of transitions in the CCDS relative to total number of SNPs	767	number SINEs in ccds plus 5 kb flanking	1524	frequency of GGGAC in 5 kb flanking
11	proportion of A-C SNPs in the CCDS relative to total num SNPs	768	percentage SINEs in ccds plus 5 kb flanking	1525	frequency of GGGAG in 5 kb flanking
12	proportion of A-G SNPs in the CCDS relative to total num SNPs	769	number ALUs in ccds plus 5 kb flanking	1526	frequency of GGGAT in 5 kb flanking
13	proportion of A-T SNPs in the CCDS relative to total num SNPs	770	percentage ALUs in ccds plus 5 kb flanking	1527	frequency of GGGCA in 5 kb flanking
14	proportion of C-G SNPs in the CCDS relative to total num SNPs	771	number MIRs in ccds plus 5 kb flanking	1528	frequency of GGGCC in 5 kb flanking
15	proportion of C-T SNPs in the CCDS relative to total num SNPs	772	percentage MIRs in ccds plus 5 kb flanking	1529	frequency of GGGCG in 5 kb flanking
16	proportion of G-T SNPs in the CCDS relative to total num SNPs	773	number LINEs in ccds plus 5 kb flanking	1530	frequency of GGGCT in 5 kb flanking
17	A-C SNPs per bp in the CCDS	774	percentage LINEs in ccds plus 5 kb flanking	1531	frequency of GGGGA in 5 kb flanking
18	A-G SNPs per bp in the CCDS	775	number LINE1 in ccds plus 5 kb flanking	1532	frequency of GGGGC in 5 kb flanking
19	A-T SNPs per bp in the CCDS	776	percentage LINE1 in ccds plus 5 kb flanking	1533	frequency of GGGGG in 5 kb flanking
20	C-G SNPs per bp in the CCDS	777	number LINE2 in ccds plus 5 kb flanking	1534	frequency of GGGGT in 5 kb flanking
21	C-T SNPs per bp in the CCDS	778	percentage LINE2 in ccds plus 5 kb flanking	1535	frequency of GGGTA in 5 kb flanking
22	G-T SNPs per bp in the CCDS	779	number L3/CR1 in ccds plus 5 kb flanking	1536	frequency of GGGTC in 5 kb flanking
23	transitions per bp in the CCDS	780	percentage L3/CR1 in ccds plus 5 kb flanking	1537	frequency of GGGTG in 5 kb flanking
24	transversions per bp in the CCDS	781	number LTR in ccds plus 5 kb flanking	1538	frequency of GGGTT in 5 kb flanking
25	total snps per bp in the CCDS	782	percentage LTR in ccds plus 5 kb flanking	1539	frequency of GGTA in 5 kb flanking
26	number of A-C SNPs in exons	783	number MaLRs in ccds plus 5 kb flanking	1540	frequency of GGTAC in 5 kb flanking
27	number of A-G SNPs in exons	784	percentage MaLRs in ccds plus 5 kb flanking	1541	frequency of GGTAG in 5 kb flanking
28	number of A-T SNPs in exons	785	number ERVL in ccds plus 5 kb flanking	1542	frequency of GGTAT in 5 kb flanking

29	number of C-G SNPs in exons	786	percentage ERVL in ccds plus 5 kb flanking	1543	frequency of GGTCA in 5 kb flanking
30	number of C-T SNPs in exons	787	number ERV_classI in ccds plus 5 kb flanking	1544	frequency of GGTC in 5 kb flanking
31	number of G-T SNPs in exons	788	percentage ERV_classI in ccds plus 5 kb flanking	1545	frequency of GGTCG in 5 kb flanking
32	number of total SNPs in exons	789	number ERV_classII in ccds plus 5 kb flanking	1546	frequency of GGTCT in 5 kb flanking
33	number of transitions in exons	790	percentage ERV_classII in ccds plus 5 kb flanking	1547	frequency of GGTGA in 5 kb flanking
34	number of transversions in exons	791	number DNA elements in ccds plus 5 kb flanking	1548	frequency of GGTGC in 5 kb flanking
35	proportion of transitions in exons relative to total number of SNPs	792	percentage DNA elements in ccds plus 5 kb flanking	1549	frequency of GGTGG in 5 kb flanking
36	proportion of A-C SNPs in exons relative to total num SNPs	793	number MER1_type in ccds plus 5 kb flanking	1550	frequency of GGTGT in 5 kb flanking
37	proportion of A-G SNPs in exons relative to total num SNPs	794	percentage MER1_type in ccds plus 5 kb flanking	1551	frequency of GGTTA in 5 kb flanking
38	proportion of A-T SNPs in exons relative to total num SNPs	795	number MER2_type in ccds plus 5 kb flanking	1552	frequency of GGTTTC in 5 kb flanking
39	proportion of C-G SNPs in exons relative to total num SNPs	796	percentage MER2_type in ccds plus 5 kb flanking	1553	frequency of GGTTG in 5 kb flanking
40	proportion of C-T SNPs in exons relative to total num SNPs	797	number unclassified in ccds plus 5 kb flanking	1554	frequency of GGTTT in 5 kb flanking
41	proportion of G-T SNPs in exons relative to total num SNPs	798	percentage unclassified in ccds plus 5 kb flanking	1555	frequency of GTAAA in 5 kb flanking
42	A-C SNPs per bp in exons	799	number small RNA in ccds plus 5 kb flanking	1556	frequency of GTAAC in 5 kb flanking
43	A-G SNPs per bp in exons	800	percentage small RNA in ccds plus 5 kb flanking	1557	frequency of GTAAG in 5 kb flanking
44	A-T SNPs per bp in exons	801	number satellites in ccds plus 5 kb flanking	1558	frequency of GTAAT in 5 kb flanking
45	C-G SNPs per bp in exons	802	percentage satellites in ccds plus 5 kb flanking	1559	frequency of GTACA in 5 kb flanking
46	C-T SNPs per bp in exons	803	number simple repeats in ccds plus 5 kb flanking	1560	frequency of GTACC in 5 kb flanking
47	G-T SNPs per bp in exons	804	percentage simple repeats in ccds plus 5 kb flanking	1561	frequency of GTACG in 5 kb flanking
48	transitions per bp in exons	805	number low complexity in ccds plus 5 kb flanking	1562	frequency of GTACT in 5 kb flanking
49	transversions per bp in exons	806	percentage low complexity in ccds plus 5 kb flanking	1563	frequency of GTAGA in 5 kb flanking
50	total snps per bp in exons	807	total repeats (number SINEs + LINEs + LTR + DNA elements) in ccds plus 5 kb flanking	1564	frequency of GTAGC in 5 kb flanking
51	number of A-C SNPs in introns	808	GC content in ccds plus 10 kb flanking	1565	frequency of GTAGG in 5 kb flanking
52	number of A-G SNPs in introns	809	total % bases masked in ccds plus 10 kb flanking	1566	frequency of GTAGT in 5 kb flanking
53	number of A-T SNPs in introns	810	number SINEs in ccds plus 10 kb flanking	1567	frequency of GTATA in 5 kb flanking
54	number of C-G SNPs in introns	811	percentage SINEs in ccds plus 10 kb flanking	1568	frequency of GTATC in 5 kb flanking
55	number of C-T SNPs in introns	812	number ALUs in ccds plus 10 kb flanking	1569	frequency of GTATG in 5 kb flanking
56	number of G-T SNPs in introns	813	percentage ALUs in ccds plus 10 kb flanking	1570	frequency of GTATT in 5 kb flanking
57	number of total SNPs in introns	814	number MIRs in ccds plus 10 kb flanking	1571	frequency of GTCAA in 5 kb flanking
58	number of transitions in introns	815	percentage MIRs in ccds plus 10 kb flanking	1572	frequency of GTCAC in 5 kb flanking
59	number of transversions in introns	816	number LINEs in ccds plus 10 kb flanking	1573	frequency of GTCAG in 5 kb flanking

60	proportion of transitions in introns relative to total number of SNPs	817	percentage LINEs in ccds plus 10 kb flanking	1574	frequency of GTCAT in 5 kb flanking
61	proportion of A-C SNPs in introns relative to total num SNPs	818	number LINE1 in ccds plus 10 kb flanking	1575	frequency of GTCCA in 5 kb flanking
62	proportion of A-G SNPs in introns relative to total num SNPs	819	percentage LINE1 in ccds plus 10 kb flanking	1576	frequency of GTCCC in 5 kb flanking
63	proportion of A-T SNPs in introns relative to total num SNPs	820	number LINE2 in ccds plus 10 kb flanking	1577	frequency of GTCCG in 5 kb flanking
64	proportion of C-G SNPs in introns relative to total num SNPs	821	percentage LINE2 in ccds plus 10 kb flanking	1578	frequency of GTCCT in 5 kb flanking
65	proportion of C-T SNPs in introns relative to total num SNPs	822	number L3/CR1 in ccds plus 10 kb flanking	1579	frequency of GTCGA in 5 kb flanking
66	proportion of G-T SNPs in introns relative to total num SNPs	823	percentage L3/CR1 in ccds plus 10 kb flanking	1580	frequency of GTCGC in 5 kb flanking
67	A-C SNPs per bp in introns	824	number LTR in ccds plus 10 kb flanking	1581	frequency of GTCGG in 5 kb flanking
68	A-G SNPs per bp in introns	825	percentage LTR in ccds plus 10 kb flanking	1582	frequency of GTCGT in 5 kb flanking
69	A-T SNPs per bp in introns	826	number MaLRs in ccds plus 10 kb flanking	1583	frequency of GTCTA in 5 kb flanking
70	C-G SNPs per bp in introns	827	percentage MaLRs in ccds plus 10 kb flanking	1584	frequency of GTCTC in 5 kb flanking
71	C-T SNPs per bp in introns	828	number ERVL in ccds plus 10 kb flanking	1585	frequency of GTCTG in 5 kb flanking
72	G-T SNPs per bp in introns	829	percentage ERVL in ccds plus 10 kb flanking	1586	frequency of GTCTT in 5 kb flanking
73	transitions per bp in introns	830	number ERV_classI in ccds plus 10 kb flanking	1587	frequency of GTGAA in 5 kb flanking
74	transversions per bp in introns	831	percentage ERV_classI in ccds plus 10 kb flanking	1588	frequency of GTGAC in 5 kb flanking
75	total snps per bp in introns	832	number ERV_classII in ccds plus 10 kb flanking	1589	frequency of GTGAG in 5 kb flanking
76	number of A-C SNPs in 2 kb up	833	percentage ERV_classII in ccds plus 10 kb flanking	1590	frequency of GTGAT in 5 kb flanking
77	number of A-G SNPs in 2 kb up	834	number DNA elements in ccds plus 10 kb flanking	1591	frequency of GTGCA in 5 kb flanking
78	number of A-T SNPs in 2 kb up	835	percentage DNA elements in ccds plus 10 kb flanking	1592	frequency of GTGCC in 5 kb flanking
79	number of C-G SNPs in 2 kb up	836	number MER1_type in ccds plus 10 kb flanking	1593	frequency of GTGCG in 5 kb flanking
80	number of C-T SNPs in 2 kb up	837	percentage MER1_type in ccds plus 10 kb flanking	1594	frequency of GTGCT in 5 kb flanking
81	number of G-T SNPs in 2 kb up	838	number MER2_type in ccds plus 10 kb flanking	1595	frequency of GTGGA in 5 kb flanking
82	number of total SNPs in 2 kb up	839	percentage MER2_type in ccds plus 10 kb flanking	1596	frequency of GTGGC in 5 kb flanking
83	number of transitions in 2 kb up	840	number unclassified in ccds plus 10 kb flanking	1597	frequency of GTGGG in 5 kb flanking
84	number of transversions in 2 kb up	841	percentage unclassified in ccds plus 10 kb flanking	1598	frequency of GTGGT in 5 kb flanking
85	proportion of transitions in 2 kb up relative to total num of SNPs	842	number small RNA in ccds plus 10 kb flanking	1599	frequency of GTGTA in 5 kb flanking
86	proportion of A-C SNPs in 2 kb up relative to total num SNPs	843	percentage small RNA in ccds plus 10 kb flanking	1600	frequency of GTGTC in 5 kb flanking
87	proportion of A-G SNPs in 2 kb up relative to total num SNPs	844	number satellites in ccds plus 10 kb flanking	1601	frequency of GTGTG in 5 kb flanking
88	proportion of A-T SNPs in 2 kb up relative to total num SNPs	845	percentage satellites in ccds plus 10 kb flanking	1602	frequency of GTGTT in 5 kb flanking
89	proportion of C-G SNPs in 2 kb up relative to total num SNPs	846	number simple repeats in ccds plus 10 kb flanking	1603	frequency of GTTAA in 5 kb flanking
90	proportion of C-T SNPs in 2 kb up relative to total num SNPs	847	percentage simple repeats in ccds plus 10 kb flanking	1604	frequency of GTTAC in 5 kb flanking

91	proportion of G-T SNPs in 2 kb up relative to total num SNPs	848	number low complexity in ccds plus 10 kb flanking	1605	frequency of GTTAG in 5 kb flanking
92	A-C SNPs per bp in 2 kb up	849	percentage low complexity in ccds plus 10 kb flanking	1606	frequency of GTTAT in 5 kb flanking
93	A-G SNPs per bp in 2 kb up	850	total repeats (number SINEs + LINEs + LTR + DNA elements) in ccds plus 10 kb flanking	1607	frequency of GTTCA in 5 kb flanking
94	A-T SNPs per bp in 2 kb up	851	frequency of AAAAA in 5 kb flanking	1608	frequency of GTTCC in 5 kb flanking
95	C-G SNPs per bp in 2 kb up	852	frequency of AAAAC in 5 kb flanking	1609	frequency of GTTCG in 5 kb flanking
96	C-T SNPs per bp in 2 kb up	853	frequency of AAAAG in 5 kb flanking	1610	frequency of GTTCT in 5 kb flanking
97	G-T SNPs per bp in 2 kb up	854	frequency of AAAAT in 5 kb flanking	1611	frequency of GTTGA in 5 kb flanking
98	transitions per bp in 2 kb up	855	frequency of AAACA in 5 kb flanking	1612	frequency of GTTGC in 5 kb flanking
99	transversions per bp 2 kb up	856	frequency of AAACC in 5 kb flanking	1613	frequency of GTTGG in 5 kb flanking
100	total snps per bp in 2 kb up	857	frequency of AAACG in 5 kb flanking	1614	frequency of GTTGT in 5 kb flanking
101	number of A-C SNPs in 2 kb down	858	frequency of AAACT in 5 kb flanking	1615	frequency of GTTTA in 5 kb flanking
102	number of A-G SNPs in 2 kb down	859	frequency of AAAGA in 5 kb flanking	1616	frequency of GTTTC in 5 kb flanking
103	number of A-T SNPs in 2 kb down	860	frequency of AAAGC in 5 kb flanking	1617	frequency of GTTTG in 5 kb flanking
104	number of C-G SNPs in 2 kb down	861	frequency of AAAGG in 5 kb flanking	1618	frequency of GTTTT in 5 kb flanking
105	number of C-T SNPs in 2 kb down	862	frequency of AAAGT in 5 kb flanking	1619	frequency of TAAAA in 5 kb flanking
106	number of G-T SNPs in 2 kb down	863	frequency of AAATA in 5 kb flanking	1620	frequency of TAAAC in 5 kb flanking
107	number of total SNPS in 2 kb down	864	frequency of AAATC in 5 kb flanking	1621	frequency of TAAAG in 5 kb flanking
108	number of transitions in 2 kb down	865	frequency of AAATG in 5 kb flanking	1622	frequency of TAAAT in 5 kb flanking
109	number of transversions in 2 kb down	866	frequency of AAATT in 5 kb flanking	1623	frequency of TAACA in 5 kb flanking
110	proportion of transitions in 2 kb down relative to total num of SNPs	867	frequency of AACAA in 5 kb flanking	1624	frequency of TAACC in 5 kb flanking
111	proportion of A-C SNPs in 2 kb down relative to total num SNPs	868	frequency of AACAC in 5 kb flanking	1625	frequency of TAACG in 5 kb flanking
112	proportion of A-G SNPs in 2 kb down relative to total num SNPs	869	frequency of AACAG in 5 kb flanking	1626	frequency of TAACT in 5 kb flanking
113	proportion of A-T SNPs in 2 kb down relative to total num SNPs	870	frequency of AACAT in 5 kb flanking	1627	frequency of TAAGA in 5 kb flanking
114	proportion of C-G SNPs in 2 kb down relative to total num SNPs	871	frequency of AACCA in 5 kb flanking	1628	frequency of TAAGC in 5 kb flanking
115	proportion of C-T SNPs in 2 kb down relative to total num SNPs	872	frequency of AACCC in 5 kb flanking	1629	frequency of TAAGG in 5 kb flanking
116	proportion of G-T SNPs in 2 kb down relative to total num SNPs	873	frequency of AACCG in 5 kb flanking	1630	frequency of TAAGT in 5 kb flanking
117	A-C SNPs per bp in 2 kb down	874	frequency of AACCT in 5 kb flanking	1631	frequency of TAATA in 5 kb flanking
118	A-G SNPs per bp in 2 kb down	875	frequency of AACGA in 5 kb flanking	1632	frequency of TAATC in 5 kb flanking
119	A-T SNPs per bp in 2 kb down	876	frequency of AACGC in 5 kb flanking	1633	frequency of TAATG in 5 kb flanking
120	C-G SNPs per bp in 2 kb down	877	frequency of AACGG in 5 kb flanking	1634	frequency of TAATT in 5 kb flanking
121	C-T SNPs per bp in 2 kb down	878	frequency of AACGT in 5 kb flanking	1635	frequency of TACAA in 5 kb flanking
122	G-T SNPs per bp in 2 kb down	879	frequency of AACTA in 5 kb flanking	1636	frequency of TACAC in 5 kb flanking
123	transitions per bp in 2 kb down	880	frequency of AACTC in 5 kb	1637	frequency of TACAG in 5 kb

			flanking		flanking
124	transversions per bp 2 kb down	881	frequency of AACTG in 5 kb flanking	1638	frequency of TACAT in 5 kb flanking
125	total snps per bp in 2 kb down	882	frequency of AACCT in 5 kb flanking	1639	frequency of TACCA in 5 kb flanking
126	number of A-C SNPs in 5 kb up	883	frequency of AAGAA in 5 kb flanking	1640	frequency of TACCC in 5 kb flanking
127	number of A-G SNPs in 5 kb up	884	frequency of AAGAC in 5 kb flanking	1641	frequency of TACCG in 5 kb flanking
128	number of A-T SNPs in 5 kb up	885	frequency of AAGAG in 5 kb flanking	1642	frequency of TACCT in 5 kb flanking
129	number of C-G SNPs in 5 kb up	886	frequency of AAGAT in 5 kb flanking	1643	frequency of TACGA in 5 kb flanking
130	number of C-T SNPs in 5 kb up	887	frequency of AAGCA in 5 kb flanking	1644	frequency of TACGC in 5 kb flanking
131	number of G-T SNPs in 5 kb up	888	frequency of AAGCC in 5 kb flanking	1645	frequency of TACGG in 5 kb flanking
132	number of total SNPs in 5 kb up	889	frequency of AAGCG in 5 kb flanking	1646	frequency of TACGT in 5 kb flanking
133	number of transitions in 5 kb up	890	frequency of AAGCT in 5 kb flanking	1647	frequency of TACTA in 5 kb flanking
134	number of transversions in 5 kb up	891	frequency of AAGGA in 5 kb flanking	1648	frequency of TACTC in 5 kb flanking
135	proportion of transitions in 5 kb up relative to total num of SNPs	892	frequency of AAGGC in 5 kb flanking	1649	frequency of TACTG in 5 kb flanking
136	proportion of A-C SNPs in 5 kb up relative to total num SNPs	893	frequency of AAGGG in 5 kb flanking	1650	frequency of TACTT in 5 kb flanking
137	proportion of A-G SNPs in 5 kb up relative to total num SNPs	894	frequency of AAGGT in 5 kb flanking	1651	frequency of TAGAA in 5 kb flanking
138	proportion of A-T SNPs in 5 kb up relative to total num SNPs	895	frequency of AAGTA in 5 kb flanking	1652	frequency of TAGAC in 5 kb flanking
139	proportion of C-G SNPs in 5 kb up relative to total num SNPs	896	frequency of AAGTC in 5 kb flanking	1653	frequency of TAGAG in 5 kb flanking
140	proportion of C-T SNPs in 5 kb up relative to total num SNPs	897	frequency of AAGTG in 5 kb flanking	1654	frequency of TAGAT in 5 kb flanking
141	proportion of G-T SNPs in 5 kb up relative to total num SNPs	898	frequency of AAGTT in 5 kb flanking	1655	frequency of TAGCA in 5 kb flanking
142	A-C SNPs per bp in 5 kb up	899	frequency of AATAA in 5 kb flanking	1656	frequency of TAGCC in 5 kb flanking
143	A-G SNPs per bp in 5 kb up	900	frequency of AATAC in 5 kb flanking	1657	frequency of TAGCG in 5 kb flanking
144	A-T SNPs per bp in 5 kb up	901	frequency of AATAG in 5 kb flanking	1658	frequency of TAGCT in 5 kb flanking
145	C-G SNPs per bp in 5 kb up	902	frequency of AATAT in 5 kb flanking	1659	frequency of TAGGA in 5 kb flanking
146	C-T SNPs per bp in 5 kb up	903	frequency of AATCA in 5 kb flanking	1660	frequency of TAGGC in 5 kb flanking
147	G-T SNPs per bp in 5 kb up	904	frequency of AATCC in 5 kb flanking	1661	frequency of TAGGG in 5 kb flanking
148	transitions per bp in 5 kb up	905	frequency of AATCG in 5 kb flanking	1662	frequency of TAGGT in 5 kb flanking
149	transversions per bp 5 kb up	906	frequency of AATCT in 5 kb flanking	1663	frequency of TAGTA in 5 kb flanking
150	total snps per bp in 5 kb up	907	frequency of AATGA in 5 kb flanking	1664	frequency of TAGTC in 5 kb flanking
151	number of A-C SNPs in 5 kb down	908	frequency of AATGC in 5 kb flanking	1665	frequency of TAGTG in 5 kb flanking
152	number of A-G SNPs in 5 kb down	909	frequency of AATGG in 5 kb flanking	1666	frequency of TAGTT in 5 kb flanking
153	number of A-T SNPs in 5 kb down	910	frequency of AATGT in 5 kb flanking	1667	frequency of TATAA in 5 kb flanking
154	number of C-G SNPs in 5 kb down	911	frequency of AATTA in 5 kb flanking	1668	frequency of TATAC in 5 kb flanking
155	number of C-T SNPs in 5 kb down	912	frequency of AATTC in 5 kb flanking	1669	frequency of TATAG in 5 kb flanking
156	number of G-T SNPs in 5 kb down	913	frequency of AATTG in 5 kb flanking	1670	frequency of TATAT in 5 kb flanking



157	number of total SNPS in 5 kb down	914	frequency of AATTT in 5 kb flanking	1671	frequency of TATCA in 5 kb flanking
158	number of transitions in 5 kb down	915	frequency of ACAA in 5 kb flanking	1672	frequency of TATCC in 5 kb flanking
159	number of transversions in 5 kb down	916	frequency of ACAAC in 5 kb flanking	1673	frequency of TATCG in 5 kb flanking
160	proportion of transitions in 5 kb down relative to total num of SNPs	917	frequency of ACAAG in 5 kb flanking	1674	frequency of TATCT in 5 kb flanking
161	proportion of A-C SNPs in 5 kb down relative to total num SNPs	918	frequency of ACAAT in 5 kb flanking	1675	frequency of TATGA in 5 kb flanking
162	proportion of A-G SNPs in 5 kb down relative to total num SNPs	919	frequency of ACACA in 5 kb flanking	1676	frequency of TATGC in 5 kb flanking
163	proportion of A-T SNPs in 5 kb down relative to total num SNPs	920	frequency of ACACC in 5 kb flanking	1677	frequency of TATGG in 5 kb flanking
164	proportion of C-G SNPs in 5 kb down relative to total num SNPs	921	frequency of ACACG in 5 kb flanking	1678	frequency of TATGT in 5 kb flanking
165	proportion of C-T SNPs in 5 kb down relative to total num SNPs	922	frequency of ACACT in 5 kb flanking	1679	frequency of TATTA in 5 kb flanking
166	proportion of G-T SNPs in 5 kb down relative to total num SNPs	923	frequency of ACAGA in 5 kb flanking	1680	frequency of TATTC in 5 kb flanking
167	A-C SNPs per bp in 5 kb down	924	frequency of ACAGC in 5 kb flanking	1681	frequency of TATTG in 5 kb flanking
168	A-G SNPs per bp in 5 kb down	925	frequency of ACAGG in 5 kb flanking	1682	frequency of TATTT in 5 kb flanking
169	A-T SNPs per bp in 5 kb down	926	frequency of ACAGT in 5 kb flanking	1683	frequency of TCAAA in 5 kb flanking
170	C-G SNPs per bp in 5 kb down	927	frequency of ACATA in 5 kb flanking	1684	frequency of TCAAC in 5 kb flanking
171	C-T SNPs per bp in 5 kb down	928	frequency of ACATC in 5 kb flanking	1685	frequency of TCAAG in 5 kb flanking
172	G-T SNPs per bp in 5 kb down	929	frequency of ACATG in 5 kb flanking	1686	frequency of TCAAT in 5 kb flanking
173	transitions per bp in 5 kb down	930	frequency of ACATT in 5 kb flanking	1687	frequency of TCACA in 5 kb flanking
174	transversions per bp 5 kb down	931	frequency of ACCAA in 5 kb flanking	1688	frequency of TCACC in 5 kb flanking
175	total snps per bp in 5 kb down	932	frequency of ACCAC in 5 kb flanking	1689	frequency of TCACG in 5 kb flanking
176	number of A-C SNPs in 10 kb up	933	frequency of ACCAG in 5 kb flanking	1690	frequency of TCACT in 5 kb flanking
177	number of A-G SNPs in 10 kb up	934	frequency of ACCAT in 5 kb flanking	1691	frequency of TCAGA in 5 kb flanking
178	number of A-T SNPs in 10 kb up	935	frequency of ACCCA in 5 kb flanking	1692	frequency of TCAGC in 5 kb flanking
179	number of C-G SNPs in 10 kb up	936	frequency of ACCCC in 5 kb flanking	1693	frequency of TCAGG in 5 kb flanking
180	number of C-T SNPs in 10 kb up	937	frequency of ACCCG in 5 kb flanking	1694	frequency of TCAGT in 5 kb flanking
181	number of G-T SNPs in 10 kb up	938	frequency of ACCCT in 5 kb flanking	1695	frequency of TCATA in 5 kb flanking
182	number of total SNPS in 10 kb up	939	frequency of ACCGA in 5 kb flanking	1696	frequency of TCATC in 5 kb flanking
183	number of transitions in 10 kb up	940	frequency of ACCGC in 5 kb flanking	1697	frequency of TCATG in 5 kb flanking
184	number of transversions in 10 kb up	941	frequency of ACCGG in 5 kb flanking	1698	frequency of TCATT in 5 kb flanking
185	proportion of transitions in 10 kb up relative to total num of SNPs	942	frequency of ACCGT in 5 kb flanking	1699	frequency of TCCAA in 5 kb flanking
186	proportion of A-C SNPs in 10 kb up relative to total num SNPs	943	frequency of ACCTA in 5 kb flanking	1700	frequency of TCCAC in 5 kb flanking
187	proportion of A-G SNPs in 10 kb up relative to total num SNPs	944	frequency of ACCTC in 5 kb flanking	1701	frequency of TCCAG in 5 kb flanking
188	proportion of A-T SNPs in 10 kb up relative to total num SNPs	945	frequency of ACCTG in 5 kb flanking	1702	frequency of TCCAT in 5 kb flanking
189	proportion of C-G SNPs in 10 kb up relative to total num SNPs	946	frequency of ACCTT in 5 kb flanking	1703	frequency of TCCCA in 5 kb flanking

190	proportion of C-T SNPs in 10 kb up relative to total num SNPs	947	frequency of ACGAA in 5 kb flanking	1704	frequency of TCCCC in 5 kb flanking
191	proportion of G-T SNPs in 10 kb up relative to total num SNPs	948	frequency of ACGAC in 5 kb flanking	1705	frequency of TCCCG in 5 kb flanking
192	A-C SNPs per bp in 10 kb up	949	frequency of ACGAG in 5 kb flanking	1706	frequency of TCCCT in 5 kb flanking
193	A-G SNPs per bp in 10 kb up	950	frequency of ACGAT in 5 kb flanking	1707	frequency of TCCGA in 5 kb flanking
194	A-T SNPs per bp in 10 kb up	951	frequency of ACGCA in 5 kb flanking	1708	frequency of TCCGC in 5 kb flanking
195	C-G SNPs per bp in 10 kb up	952	frequency of ACGCC in 5 kb flanking	1709	frequency of TCCGG in 5 kb flanking
196	C-T SNPs per bp in 10 kb up	953	frequency of ACGCG in 5 kb flanking	1710	frequency of TCCGT in 5 kb flanking
197	G-T SNPs per bp in 10 kb up	954	frequency of ACGCT in 5 kb flanking	1711	frequency of TCCTA in 5 kb flanking
198	transitions per bp in 10 kb up	955	frequency of ACGGA in 5 kb flanking	1712	frequency of TCCTC in 5 kb flanking
199	transversions per bp 10 kb up	956	frequency of ACGGC in 5 kb flanking	1713	frequency of TCCTG in 5 kb flanking
200	total snps per bp in 10 kb up	957	frequency of ACGGG in 5 kb flanking	1714	frequency of TCCTT in 5 kb flanking
201	number of A-C SNPs in 10 kb down	958	frequency of ACGGT in 5 kb flanking	1715	frequency of TCGAA in 5 kb flanking
202	number of A-G SNPs in 10 kb down	959	frequency of ACGTA in 5 kb flanking	1716	frequency of TCGAC in 5 kb flanking
203	number of A-T SNPs in 10 kb down	960	frequency of ACGTC in 5 kb flanking	1717	frequency of TCGAG in 5 kb flanking
204	number of C-G SNPs in 10 kb down	961	frequency of ACGTG in 5 kb flanking	1718	frequency of TCGAT in 5 kb flanking
205	number of C-T SNPs in 10 kb down	962	frequency of ACGTT in 5 kb flanking	1719	frequency of TCGCA in 5 kb flanking
206	number of G-T SNPs in 10 kb down	963	frequency of ACTAA in 5 kb flanking	1720	frequency of TCGCC in 5 kb flanking
207	number of total SNPS in 10 kb down	964	frequency of ACTAC in 5 kb flanking	1721	frequency of TCGCG in 5 kb flanking
208	number of transitions in 10 kb down	965	frequency of ACTAG in 5 kb flanking	1722	frequency of TCGCT in 5 kb flanking
209	number of transversions in 10 kb down	966	frequency of ACTAT in 5 kb flanking	1723	frequency of TCGGA in 5 kb flanking
210	proportion of transitions in 10 kb down relative to total num of SNPs	967	frequency of ACTCA in 5 kb flanking	1724	frequency of TCGGC in 5 kb flanking
211	proportion of A-C SNPs in 10 kb down relative to total num SNPs	968	frequency of ACTCC in 5 kb flanking	1725	frequency of TCGGG in 5 kb flanking
212	proportion of A-G SNPs in 10 kb down relative to total num SNPs	969	frequency of ACTCG in 5 kb flanking	1726	frequency of TCGGT in 5 kb flanking
213	proportion of A-T SNPs in 10 kb down relative to total num SNPs	970	frequency of ACTCT in 5 kb flanking	1727	frequency of TCGTA in 5 kb flanking
214	proportion of C-G SNPs in 10 kb down relative to total num SNPs	971	frequency of ACTGA in 5 kb flanking	1728	frequency of TCGTC in 5 kb flanking
215	proportion of C-T SNPs in 10 kb down relative to total num SNPs	972	frequency of ACTGC in 5 kb flanking	1729	frequency of TCGTG in 5 kb flanking
216	proportion of G-T SNPs in 10 kb down relative to total num SNPs	973	frequency of ACTGG in 5 kb flanking	1730	frequency of TCGTT in 5 kb flanking
217	A-C SNPs per bp in 10 kb down	974	frequency of ACTGT in 5 kb flanking	1731	frequency of TCTAA in 5 kb flanking
218	A-G SNPs per bp in 10 kb down	975	frequency of ACTTA in 5 kb flanking	1732	frequency of TCTAC in 5 kb flanking
219	A-T SNPs per bp in 10 kb down	976	frequency of ACTTC in 5 kb flanking	1733	frequency of TCTAG in 5 kb flanking
220	C-G SNPs per bp in 10 kb down	977	frequency of ACTTG in 5 kb flanking	1734	frequency of TCTAT in 5 kb flanking
221	C-T SNPs per bp in 10 kb down	978	frequency of ACTTT in 5 kb flanking	1735	frequency of TCTCA in 5 kb flanking
222	G-T SNPs per bp in 10 kb down	979	frequency of AGAAA in 5 kb flanking	1736	frequency of TCTCC in 5 kb flanking

223	transitions per bp in 10 kb down	980	frequency of AGAAC in 5 kb flanking	1737	frequency of TCTCG in 5 kb flanking
224	transversions per bp 10 kb down	981	frequency of AGAAG in 5 kb flanking	1738	frequency of TCTCT in 5 kb flanking
225	total snps per bp in 10 kb down	982	frequency of AGAAT in 5 kb flanking	1739	frequency of TCTGA in 5 kb flanking
226	number of A-C SNPs in ccds plus 2kb flanking	983	frequency of AGACA in 5 kb flanking	1740	frequency of TCTGC in 5 kb flanking
227	number of A-G SNPs in ccds plus 2kb flanking	984	frequency of AGACC in 5 kb flanking	1741	frequency of TCTGG in 5 kb flanking
228	number of A-T SNPs in ccds plus 2kb flanking	985	frequency of AGACG in 5 kb flanking	1742	frequency of TCTGT in 5 kb flanking
229	number of C-G SNPs in ccds plus 2kb flanking	986	frequency of AGACT in 5 kb flanking	1743	frequency of TCTTA in 5 kb flanking
230	number of C-T SNPs in ccds plus 2kb flanking	987	frequency of AGAGA in 5 kb flanking	1744	frequency of TCTTC in 5 kb flanking
231	number of G-T SNPs in ccds plus 2kb flanking	988	frequency of AGAGC in 5 kb flanking	1745	frequency of TCTTG in 5 kb flanking
232	number of total SNPs in ccds plus 2kb flanking	989	frequency of AGAGG in 5 kb flanking	1746	frequency of TCTTT in 5 kb flanking
233	number of transitions in ccds plus 2kb flanking	990	frequency of AGAGT in 5 kb flanking	1747	frequency of TGAAA in 5 kb flanking
234	number of transversions in ccds plus 2kb flanking	991	frequency of AGATA in 5 kb flanking	1748	frequency of TGAAC in 5 kb flanking
235	proportion of transitions in ccds plus 2kb flanking relative to total num of SNPs	992	frequency of AGATC in 5 kb flanking	1749	frequency of TGAAG in 5 kb flanking
236	proportion of A-C SNPs in ccds plus 2kb flanking relative to total num SNPs	993	frequency of AGATG in 5 kb flanking	1750	frequency of TGAAT in 5 kb flanking
237	proportion of A-G SNPs in ccds plus 2kb flanking relative to total num SNPs	994	frequency of AGATT in 5 kb flanking	1751	frequency of TGACA in 5 kb flanking
238	proportion of A-T SNPs in ccds plus 2kb flanking relative to total num SNPs	995	frequency of AGCAA in 5 kb flanking	1752	frequency of TGACC in 5 kb flanking
239	proportion of C-G SNPs in ccds plus 2kb flanking relative to total num SNPs	996	frequency of AGCAC in 5 kb flanking	1753	frequency of TGACG in 5 kb flanking
240	proportion of C-T SNPs in ccds plus 2kb flanking relative to total num SNPs	997	frequency of AGCAG in 5 kb flanking	1754	frequency of TGACT in 5 kb flanking
241	proportion of G-T SNPs in ccds plus 2kb flanking relative to total num SNPs	998	frequency of AGCAT in 5 kb flanking	1755	frequency of TGAGA in 5 kb flanking
242	A-C SNPs per bp in ccds plus 2kb flanking	999	frequency of AGCCA in 5 kb flanking	1756	frequency of TGAGC in 5 kb flanking
243	A-G SNPs per bp in ccds plus 2kb flanking	1000	frequency of AGCCC in 5 kb flanking	1757	frequency of TGAGG in 5 kb flanking
244	A-T SNPs per bp in ccds plus 2kb flanking	1001	frequency of AGCCG in 5 kb flanking	1758	frequency of TGAGT in 5 kb flanking
245	C-G SNPs per bp in ccds plus 2kb flanking	1002	frequency of AGCCT in 5 kb flanking	1759	frequency of TGATA in 5 kb flanking
246	C-T SNPs per bp in ccds plus 2kb flanking	1003	frequency of AGCGA in 5 kb flanking	1760	frequency of TGATC in 5 kb flanking
247	G-T SNPs per bp in ccds plus 2kb flanking	1004	frequency of AGCGC in 5 kb flanking	1761	frequency of TGATG in 5 kb flanking
248	transitions per bp in ccds plus 2kb flanking	1005	frequency of AGCGG in 5 kb flanking	1762	frequency of TGATT in 5 kb flanking
249	transversions per bp ccds plus 2kb flanking	1006	frequency of AGCGT in 5 kb flanking	1763	frequency of TGCAA in 5 kb flanking
250	total snps per bp in ccds plus 2kb flanking	1007	frequency of AGCTA in 5 kb flanking	1764	frequency of TGCAC in 5 kb flanking
251	number of A-C SNPs in ccds plus 5kb flanking	1008	frequency of AGCTC in 5 kb flanking	1765	frequency of TGCAG in 5 kb flanking

252	number of A-G SNPs in ccds plus 5kb flanking	1009	frequency of AGCTG in 5 kb flanking	1766	frequency of TGCAT in 5 kb flanking
253	number of A-T SNPs in ccds plus 5kb flanking	1010	frequency of AGCTT in 5 kb flanking	1767	frequency of TGCCA in 5 kb flanking
254	number of C-G SNPs in ccds plus 5kb flanking	1011	frequency of AGGAA in 5 kb flanking	1768	frequency of TGCCC in 5 kb flanking
255	number of C-T SNPs in ccds plus 5kb flanking	1012	frequency of AGGAC in 5 kb flanking	1769	frequency of TGCCG in 5 kb flanking
256	number of G-T SNPs in ccds plus 5kb flanking	1013	frequency of AGGAG in 5 kb flanking	1770	frequency of TGCCT in 5 kb flanking
257	number of total SNPs in ccds plus 5kb flanking	1014	frequency of AGGAT in 5 kb flanking	1771	frequency of TGCGA in 5 kb flanking
258	number of transitions in ccds plus 5kb flanking	1015	frequency of AGGCA in 5 kb flanking	1772	frequency of TGCGC in 5 kb flanking
259	number of transversions in ccds plus 5kb flanking	1016	frequency of AGGCC in 5 kb flanking	1773	frequency of TGCGG in 5 kb flanking
260	proportion of transitions in ccds plus 5kb flanking relative to total num of SNPs	1017	frequency of AGGCG in 5 kb flanking	1774	frequency of TGCGT in 5 kb flanking
261	proportion of A-C SNPs in ccds plus 5kb flanking relative to total num SNPs	1018	frequency of AGGCT in 5 kb flanking	1775	frequency of TGCTA in 5 kb flanking
262	proportion of A-G SNPs in ccds plus 5kb flanking relative to total num SNPs	1019	frequency of AGGGA in 5 kb flanking	1776	frequency of TGCTC in 5 kb flanking
263	proportion of A-T SNPs in ccds plus 5kb flanking relative to total num SNPs	1020	frequency of AGGGC in 5 kb flanking	1777	frequency of TGCTG in 5 kb flanking
264	proportion of C-G SNPs in ccds plus 5kb flanking relative to total num SNPs	1021	frequency of AGGGG in 5 kb flanking	1778	frequency of TGCTT in 5 kb flanking
265	proportion of C-T SNPs in ccds plus 5kb flanking relative to total num SNPs	1022	frequency of AGGGT in 5 kb flanking	1779	frequency of TGGAA in 5 kb flanking
266	proportion of G-T SNPs in ccds plus 5kb flanking relative to total num SNPs	1023	frequency of AGGTA in 5 kb flanking	1780	frequency of TGGAC in 5 kb flanking
267	A-C SNPs per bp in ccds plus 5kb flanking	1024	frequency of AGGTC in 5 kb flanking	1781	frequency of TGGAG in 5 kb flanking
268	A-G SNPs per bp in ccds plus 5kb flanking	1025	frequency of AGGTG in 5 kb flanking	1782	frequency of TGGAT in 5 kb flanking
269	A-T SNPs per bp in ccds plus 5kb flanking	1026	frequency of AGGTT in 5 kb flanking	1783	frequency of TGGCA in 5 kb flanking
270	C-G SNPs per bp in ccds plus 5kb flanking	1027	frequency of AGTAA in 5 kb flanking	1784	frequency of TGGCC in 5 kb flanking
271	C-T SNPs per bp in ccds plus 5kb flanking	1028	frequency of AGTAC in 5 kb flanking	1785	frequency of TGGCG in 5 kb flanking
272	G-T SNPs per bp in ccds plus 5kb flanking	1029	frequency of AGTAG in 5 kb flanking	1786	frequency of TGGCT in 5 kb flanking
273	transitions per bp in ccds plus 5kb flanking	1030	frequency of AGTAT in 5 kb flanking	1787	frequency of TGGGA in 5 kb flanking
274	transversions per bp ccds plus 5kb flanking	1031	frequency of AGTCA in 5 kb flanking	1788	frequency of TGGGC in 5 kb flanking
275	total snps per bp in ccds plus 5kb flanking	1032	frequency of AGTCC in 5 kb flanking	1789	frequency of TGGGG in 5 kb flanking
276	number of A-C SNPs in ccds plus 10kb flanking	1033	frequency of AGTCG in 5 kb flanking	1790	frequency of TGGGT in 5 kb flanking
277	number of A-G SNPs in ccds plus 10kb flanking	1034	frequency of AGTCT in 5 kb flanking	1791	frequency of TGGTA in 5 kb flanking
278	number of A-T SNPs in ccds plus 10kb flanking	1035	frequency of AGTGA in 5 kb flanking	1792	frequency of TGGTC in 5 kb flanking
279	number of C-G SNPs in ccds plus 10kb flanking	1036	frequency of AGTGC in 5 kb flanking	1793	frequency of TGGTG in 5 kb flanking

280	number of C-T SNPs in ccds plus 10kb flanking	1037	frequency of AGTGG in 5 kb flanking	1794	frequency of TGGTT in 5 kb flanking
281	number of G-T SNPs in ccds plus 10kb flanking	1038	frequency of AGTGT in 5 kb flanking	1795	frequency of TGTAA in 5 kb flanking
282	number of total SNPS in ccds plus 10kb flanking	1039	frequency of AGTTA in 5 kb flanking	1796	frequency of TGTAC in 5 kb flanking
283	number of transitions in ccds plus 10kb flanking	1040	frequency of AGTTC in 5 kb flanking	1797	frequency of TGTAG in 5 kb flanking
284	number of transversions in ccds plus 10kb flanking	1041	frequency of AGTTG in 5 kb flanking	1798	frequency of TGTAT in 5 kb flanking
285	proportion of transitions in ccds plus 10kb flanking relative to total num of SNPs	1042	frequency of AGTTT in 5 kb flanking	1799	frequency of TGTCA in 5 kb flanking
286	proportion of A-C SNPs in ccds plus 10kb flanking relative to total num SNPs	1043	frequency of ATAAA in 5 kb flanking	1800	frequency of TGTCC in 5 kb flanking
287	proportion of A-G SNPs in ccds plus 10kb flanking relative to total num SNPs	1044	frequency of ATAAC in 5 kb flanking	1801	frequency of TGTCCG in 5 kb flanking
288	proportion of A-T SNPs in ccds plus 10kb flanking relative to total num SNPs	1045	frequency of ATAAG in 5 kb flanking	1802	frequency of TGTCT in 5 kb flanking
289	proportion of C-G SNPs in ccds plus 10kb flanking relative to total num SNPs	1046	frequency of ATAAT in 5 kb flanking	1803	frequency of TGTGA in 5 kb flanking
290	proportion of C-T SNPs in ccds plus 10kb flanking relative to total num SNPs	1047	frequency of ATACA in 5 kb flanking	1804	frequency of TGTGC in 5 kb flanking
291	proportion of G-T SNPs in ccds plus 10kb flanking relative to total num SNPs	1048	frequency of ATACC in 5 kb flanking	1805	frequency of TGTGG in 5 kb flanking
292	A-C SNPs per bp in ccds plus 10kb flanking	1049	frequency of ATACG in 5 kb flanking	1806	frequency of TGTGT in 5 kb flanking
293	A-G SNPs per bp in ccds plus 10kb flanking	1050	frequency of ATACT in 5 kb flanking	1807	frequency of TGTTA in 5 kb flanking
294	A-T SNPs per bp in ccds plus 10kb flanking	1051	frequency of ATAGA in 5 kb flanking	1808	frequency of TGTTC in 5 kb flanking
295	C-G SNPs per bp in ccds plus 10kb flanking	1052	frequency of ATAGC in 5 kb flanking	1809	frequency of TGTTG in 5 kb flanking
296	C-T SNPs per bp in ccds plus 10kb flanking	1053	frequency of ATAGG in 5 kb flanking	1810	frequency of TGTTT in 5 kb flanking
297	G-T SNPs per bp in ccds plus 10kb flanking	1054	frequency of ATAGT in 5 kb flanking	1811	frequency of TTAAC in 5 kb flanking
298	transitions per bp in ccds plus 10kb flanking	1055	frequency of ATATA in 5 kb flanking	1812	frequency of TTAAC in 5 kb flanking
299	transversions per bp ccds plus 10kb flanking	1056	frequency of ATATC in 5 kb flanking	1813	frequency of TTAAG in 5 kb flanking
300	total snps per bp in ccds plus 10kb flanking	1057	frequency of ATATG in 5 kb flanking	1814	frequency of TTAAT in 5 kb flanking
301	number of A-C SNPs in 2kb flanking	1058	frequency of ATATT in 5 kb flanking	1815	frequency of TTACA in 5 kb flanking
302	number of A-G SNPs in 2kb flanking	1059	frequency of ATCAA in 5 kb flanking	1816	frequency of TTACC in 5 kb flanking
303	number of A-T SNPs in 2kb flanking	1060	frequency of ATCAC in 5 kb flanking	1817	frequency of TTACG in 5 kb flanking
304	number of C-G SNPs in 2kb flanking	1061	frequency of ATCAG in 5 kb flanking	1818	frequency of TTAAC in 5 kb flanking
305	number of C-T SNPs in 2kb flanking	1062	frequency of ATCAT in 5 kb flanking	1819	frequency of TTAGA in 5 kb flanking
306	number of G-T SNPs in 2kb flanking	1063	frequency of ATCCA in 5 kb flanking	1820	frequency of TTAGC in 5 kb flanking
307	number of total SNPS in 2kb flanking	1064	frequency of ATCCC in 5 kb flanking	1821	frequency of TTAGG in 5 kb flanking
308	number of transitions in 2kb flanking	1065	frequency of ATCCG in 5 kb flanking	1822	frequency of TTAGT in 5 kb flanking

			flanking		flanking
309	number of transversions in 2kb flanking	1066	frequency of ATCCT in 5 kb flanking	1823	frequency of TTATA in 5 kb flanking
310	proportion of transitions in 2kb flanking relative to total num of SNPs	1067	frequency of ATCGA in 5 kb flanking	1824	frequency of TTATC in 5 kb flanking
311	proportion of A-C SNPs in 2kb flanking relative to total num SNPs	1068	frequency of ATCGC in 5 kb flanking	1825	frequency of TTATG in 5 kb flanking
312	proportion of A-G SNPs in 2kb flanking relative to total num SNPs	1069	frequency of ATCGG in 5 kb flanking	1826	frequency of TTATT in 5 kb flanking
313	proportion of A-T SNPs in 2kb flanking relative to total num SNPs	1070	frequency of ATCGT in 5 kb flanking	1827	frequency of TTCAA in 5 kb flanking
314	proportion of C-G SNPs in 2kb flanking relative to total num SNPs	1071	frequency of ATCTA in 5 kb flanking	1828	frequency of TTCAC in 5 kb flanking
315	proportion of C-T SNPs in 2kb flanking relative to total num SNPs	1072	frequency of ATCTC in 5 kb flanking	1829	frequency of TTCAG in 5 kb flanking
316	proportion of G-T SNPs in 2kb flanking relative to total num SNPs	1073	frequency of ATCTG in 5 kb flanking	1830	frequency of TTCAT in 5 kb flanking
317	A-C SNPs per bp in 2kb flanking	1074	frequency of ATCTT in 5 kb flanking	1831	frequency of TTCCA in 5 kb flanking
318	A-G SNPs per bp in 2kb flanking	1075	frequency of ATGAA in 5 kb flanking	1832	frequency of TTCCC in 5 kb flanking
319	A-T SNPs per bp in 2kb flanking	1076	frequency of ATGAC in 5 kb flanking	1833	frequency of TTCCG in 5 kb flanking
320	C-G SNPs per bp in 2kb flanking	1077	frequency of ATGAG in 5 kb flanking	1834	frequency of TTCCT in 5 kb flanking
321	C-T SNPs per bp in 2kb flanking	1078	frequency of ATGAT in 5 kb flanking	1835	frequency of TTCGA in 5 kb flanking
322	G-T SNPs per bp in 2kb flanking	1079	frequency of ATGCA in 5 kb flanking	1836	frequency of TTCGC in 5 kb flanking
323	transitions per bp in 2kb flanking	1080	frequency of ATGCC in 5 kb flanking	1837	frequency of TTCGG in 5 kb flanking
324	transversions per bp 2kb flanking	1081	frequency of ATGCG in 5 kb flanking	1838	frequency of TTCGT in 5 kb flanking
325	total snps per bp in 2kb flanking	1082	frequency of ATGCT in 5 kb flanking	1839	frequency of TTCTA in 5 kb flanking
326	number of A-C SNPs in 5kb flanking	1083	frequency of ATGGA in 5 kb flanking	1840	frequency of TTCTC in 5 kb flanking
327	number of A-G SNPs in 5kb flanking	1084	frequency of ATGGC in 5 kb flanking	1841	frequency of TTCTG in 5 kb flanking
328	number of A-T SNPs in 5kb flanking	1085	frequency of ATGGG in 5 kb flanking	1842	frequency of TTCTT in 5 kb flanking
329	number of C-G SNPs in 5kb flanking	1086	frequency of ATGGT in 5 kb flanking	1843	frequency of TTGAA in 5 kb flanking
330	number of C-T SNPs in 5kb flanking	1087	frequency of ATGTA in 5 kb flanking	1844	frequency of TTGAC in 5 kb flanking
331	number of G-T SNPs in 5kb flanking	1088	frequency of ATGTC in 5 kb flanking	1845	frequency of TTGAG in 5 kb flanking
332	number of total SNPs in 5kb flanking	1089	frequency of ATGTG in 5 kb flanking	1846	frequency of TTGAT in 5 kb flanking
333	number of transitions in 5kb flanking	1090	frequency of ATGTT in 5 kb flanking	1847	frequency of TTGCA in 5 kb flanking
334	number of transversions in 5kb flanking	1091	frequency of ATTAA in 5 kb flanking	1848	frequency of TTGCC in 5 kb flanking
335	proportion of transitions in 5kb flanking relative to total num of SNPs	1092	frequency of ATTAC in 5 kb flanking	1849	frequency of TTGCG in 5 kb flanking
336	proportion of A-C SNPs in 5kb flanking relative to total num SNPs	1093	frequency of ATTAG in 5 kb flanking	1850	frequency of TTGCT in 5 kb flanking
337	proportion of A-G SNPs in 5kb flanking relative to total num SNPs	1094	frequency of ATTAT in 5 kb flanking	1851	frequency of TTGGA in 5 kb flanking
338	proportion of A-T SNPs in 5kb flanking relative to total num SNPs	1095	frequency of ATTCA in 5 kb flanking	1852	frequency of TTGGC in 5 kb flanking
339	proportion of C-G SNPs in 5kb flanking relative to total num SNPs	1096	frequency of ATTCC in 5 kb flanking	1853	frequency of TTGGG in 5 kb flanking

340	proportion of C-T SNPs in 5kb flanking relative to total num SNPs	1097	frequency of ATTCG in 5 kb flanking	1854	frequency of TTGGT in 5 kb flanking
341	proportion of G-T SNPs in 5kb flanking relative to total num SNPs	1098	frequency of ATTCT in 5 kb flanking	1855	frequency of TTGTA in 5 kb flanking
342	A-C SNPs per bp in 5kb flanking	1099	frequency of ATTGA in 5 kb flanking	1856	frequency of TTGTC in 5 kb flanking
343	A-G SNPs per bp in 5kb flanking	1100	frequency of ATTGC in 5 kb flanking	1857	frequency of TTGTG in 5 kb flanking
344	A-T SNPs per bp in 5kb flanking	1101	frequency of ATTGG in 5 kb flanking	1858	frequency of TTGTT in 5 kb flanking
345	C-G SNPs per bp in 5kb flanking	1102	frequency of ATTGT in 5 kb flanking	1859	frequency of TTTAA in 5 kb flanking
346	C-T SNPs per bp in 5kb flanking	1103	frequency of ATTTA in 5 kb flanking	1860	frequency of TTTAC in 5 kb flanking
347	G-T SNPs per bp in 5kb flanking	1104	frequency of ATTTC in 5 kb flanking	1861	frequency of TTTAG in 5 kb flanking
348	transitions per bp in 5kb flanking	1105	frequency of ATTTG in 5 kb flanking	1862	frequency of TTTAT in 5 kb flanking
349	transversions per bp 5kb flanking	1106	frequency of ATTTT in 5 kb flanking	1863	frequency of TTTCA in 5 kb flanking
350	total snps per bp in 5kb flanking	1107	frequency of CAAAA in 5 kb flanking	1864	frequency of TTTCC in 5 kb flanking
351	number of A-C SNPs in 10kb flanking	1108	frequency of CAAAC in 5 kb flanking	1865	frequency of TTTCG in 5 kb flanking
352	number of A-G SNPs in 10kb flanking	1109	frequency of CAAAG in 5 kb flanking	1866	frequency of TTTCT in 5 kb flanking
353	number of A-T SNPs in 10kb flanking	1110	frequency of CAAAT in 5 kb flanking	1867	frequency of TTTGA in 5 kb flanking
354	number of C-G SNPs in 10kb flanking	1111	frequency of CAACA in 5 kb flanking	1868	frequency of TTTGC in 5 kb flanking
355	number of C-T SNPs in 10kb flanking	1112	frequency of CAACC in 5 kb flanking	1869	frequency of TTTGG in 5 kb flanking
356	number of G-T SNPs in 10kb flanking	1113	frequency of CAACG in 5 kb flanking	1870	frequency of TTTGT in 5 kb flanking
357	number of total SNPs in 10kb flanking	1114	frequency of CAACT in 5 kb flanking	1871	frequency of TTTTA in 5 kb flanking
358	number of transitions in 10kb flanking	1115	frequency of CAAGA in 5 kb flanking	1872	frequency of TTTTC in 5 kb flanking
359	number of transversions in 10kb flanking	1116	frequency of CAAGC in 5 kb flanking	1873	frequency of TTTTG in 5 kb flanking
360	proportion of transitions in 10kb flanking relative to total num of SNPs	1117	frequency of CAAGG in 5 kb flanking	1874	frequency of TTTTT in 5 kb flanking
361	proportion of A-C SNPs in 10kb flanking relative to total num SNPs	1118	frequency of CAAGT in 5 kb flanking	1875	# of cpg islands
362	proportion of A-G SNPs in 10kb flanking relative to total num SNPs	1119	frequency of CAATA in 5 kb flanking	1876	# of cpg islands/bp of sequence
363	proportion of A-T SNPs in 10kb flanking relative to total num SNPs	1120	frequency of CAATC in 5 kb flanking	1877	% of the sequence composed of cpg islands
364	proportion of C-G SNPs in 10kb flanking relative to total num SNPs	1121	frequency of CAATG in 5 kb flanking	1878	number of exons in the gene
365	proportion of C-T SNPs in 10kb flanking relative to total num SNPs	1122	frequency of CAATT in 5 kb flanking	1879	proportion (by bp) of exon content in the ccds
366	proportion of G-T SNPs in 10kb flanking relative to total num SNPs	1123	frequency of CACAA in 5 kb flanking	1880	proportion of divergent sites per bp in ccds
367	A-C SNPs per bp in 10kb flanking	1124	frequency of CACAC in 5 kb flanking	1881	number of divergent sites in total in ccds
368	A-G SNPs per bp in 10kb flanking	1125	frequency of CACAG in 5 kb flanking	1882	proportion of divergent sites per bp, including gaps in ccds
369	A-T SNPs per bp in 10kb flanking	1126	frequency of CACAT in 5 kb flanking	1883	number of divergent sites in total, including gaps in ccds
370	C-G SNPs per bp in 10kb flanking	1127	frequency of CACCA in 5 kb flanking	1884	proportion of divergent transitions per bp in ccds
371	C-T SNPs per bp in 10kb flanking	1128	frequency of CACCC in 5 kb flanking	1885	number of divergent transitions in ccds

372	G-T SNPs per bp in 10kb flanking	1129	frequency of CACCG in 5 kb flanking	1886	proportion of divergent transversions per bp in ccds
373	transitions per bp in 10kb flanking	1130	frequency of CACCT in 5 kb flanking	1887	number of divergent transversions in ccds
374	transversions per bp 10kb flanking	1131	frequency of CACGA in 5 kb flanking	1888	proportion of divergent sites (no gaps) that are transitions in ccds
375	total snps per bp in 10kb flanking	1132	frequency of CACGC in 5 kb flanking	1889	proportion of divergent sites that (no gaps) are transversions in ccds
376	number of A-C SNPs in introns plus 2 kb flanking	1133	frequency of CACGG in 5 kb flanking	1890	proportion of total divergent sites that are gaps, including gaps in ccds
377	number of A-G SNPs in introns plus 2 kb flanking	1134	frequency of CACGT in 5 kb flanking	1891	proportion of AC divergent sites relative to total num divergent sites (no gaps) in ccds
378	number of A-T SNPs in introns plus 2 kb flanking	1135	frequency of CACTA in 5 kb flanking	1892	proportion of AG divergent sites relative to total num divergent sites (no gaps) in ccds
379	number of C-G SNPs in introns plus 2 kb flanking	1136	frequency of CACTC in 5 kb flanking	1893	proportion of AT divergent sites relative to total num divergent sites (no gaps) in ccds
380	number of C-T SNPs in introns plus 2 kb flanking	1137	frequency of CACTG in 5 kb flanking	1894	proportion of CG divergent sites relative to total num divergent sites (no gaps) in ccds
381	number of G-T SNPs in introns plus 2 kb flanking	1138	frequency of CACTT in 5 kb flanking	1895	proportion of CT divergent sites relative to total num divergent sites (no gaps) in ccds
382	number of total SNPS in introns plus 2 kb flanking	1139	frequency of CAGAA in 5 kb flanking	1896	proportion of GT divergent sites relative to total num divergent sites (no gaps) in ccds
383	number of transitions in introns plus 2 kb flanking	1140	frequency of CAGAC in 5 kb flanking	1897	AC divergent sites per bp in ccds
384	number of transversions in introns plus 2 kb flanking	1141	frequency of CAGAG in 5 kb flanking	1898	AG divergent sites per bp in ccds
385	proportion of transitions in introns plus 2 kb flanking relative to total num of SNPs	1142	frequency of CAGAT in 5 kb flanking	1899	AT divergent sites per bp in ccds
386	proportion of A-C SNPs in introns plus 2 kb flanking relative to total num SNPs	1143	frequency of CAGCA in 5 kb flanking	1900	CG divergent sites per bp in ccds
387	proportion of A-G SNPs in introns plus 2 kb flanking relative to total num SNPs	1144	frequency of CAGCC in 5 kb flanking	1901	CT divergent sites per bp in ccds
388	proportion of A-T SNPs in introns plus 2 kb flanking relative to total num SNPs	1145	frequency of CAGCG in 5 kb flanking	1902	GT divergent sites per bp in ccds
389	proportion of C-G SNPs in introns plus 2 kb flanking relative to total num SNPs	1146	frequency of CAGCT in 5 kb flanking	1903	proportion unalignable (gaps in human or other species) per bp in ccds
390	proportion of C-T SNPs in introns plus 2 kb flanking relative to total num SNPs	1147	frequency of CAGGA in 5 kb flanking	1904	proportion unalignable (gaps in human) per bp in ccds
391	proportion of G-T SNPs in introns plus 2 kb flanking relative to total num SNPs	1148	frequency of CAGGC in 5 kb flanking	1905	proportion unalignable (gaps in other species) per bp in ccds
392	A-C SNPs per bp in introns plus 2 kb flanking	1149	frequency of CAGGG in 5 kb flanking	1906	proportion of divergent sites per bp in exons
393	A-G SNPs per bp in introns plus 2 kb flanking	1150	frequency of CAGGT in 5 kb flanking	1907	number of divergent sites in total in exons
394	A-T SNPs per bp in introns plus 2 kb flanking	1151	frequency of CAGTA in 5 kb flanking	1908	proportion of divergent sites per bp, including gaps in exons
395	C-G SNPs per bp in introns plus 2 kb flanking	1152	frequency of CAGTC in 5 kb flanking	1909	number of divergent sites in total, including gaps in exons
396	C-T SNPs per bp in introns plus 2 kb flanking	1153	frequency of CAGTG in 5 kb flanking	1910	proportion of divergent transitions per bp in exons



397	G-T SNPs per bp in introns plus 2 kb flanking	1154	frequency of CAGTT in 5 kb flanking	1911	number of divergent transitions in exons
398	transitions per bp in introns plus 2 kb flanking	1155	frequency of CATAA in 5 kb flanking	1912	proportion of divergent transversions per bp in exons
399	transversions per bp introns plus 2 kb flanking	1156	frequency of CATAAC in 5 kb flanking	1913	number of divergent transversions in exons
400	total snps per bp in introns plus 2 kb flanking	1157	frequency of CATAG in 5 kb flanking	1914	proportion of divergent sites (no gaps) that are transitions in exons
401	number of A-C SNPs in introns plus 5 kb flanking	1158	frequency of CATAT in 5 kb flanking	1915	proportion of divergent sites that (no gaps) are transversions in exons
402	number of A-G SNPs in introns plus 5 kb flanking	1159	frequency of CATCA in 5 kb flanking	1916	proportion of total divergent sites that are gaps, including gaps in exons
403	number of A-T SNPs in introns plus 5 kb flanking	1160	frequency of CATCC in 5 kb flanking	1917	proportion of AC divergent sites relative to total num divergent sites (no gaps) in exons
404	number of C-G SNPs in introns plus 5 kb flanking	1161	frequency of CATCG in 5 kb flanking	1918	proportion of AG divergent sites relative to total num divergent sites (no gaps) in exons
405	number of C-T SNPs in introns plus 5 kb flanking	1162	frequency of CATCT in 5 kb flanking	1919	proportion of AT divergent sites relative to total num divergent sites (no gaps) in exons
406	number of G-T SNPs in introns plus 5 kb flanking	1163	frequency of CATGA in 5 kb flanking	1920	proportion of CG divergent sites relative to total num divergent sites (no gaps) in exons
407	number of total SNPS in introns plus 5 kb flanking	1164	frequency of CATGC in 5 kb flanking	1921	proportion of CT divergent sites relative to total num divergent sites (no gaps) in exons
408	number of transitions in introns plus 5 kb flanking	1165	frequency of CATGG in 5 kb flanking	1922	proportion of GT divergent sites relative to total num divergent sites (no gaps) in exons
409	number of transversions in introns plus 5 kb flanking	1166	frequency of CATGT in 5 kb flanking	1923	AC divergent sites per bp in exons
410	proportion of transitions in introns plus 5 kb flanking relative to total num of SNPs	1167	frequency of CATAA in 5 kb flanking	1924	AG divergent sites per bp in exons
411	proportion of A-C SNPs in introns plus 5 kb flanking relative to total num SNPs	1168	frequency of CATTC in 5 kb flanking	1925	AT divergent sites per bp in exons
412	proportion of A-G SNPs in introns plus 5 kb flanking relative to total num SNPs	1169	frequency of CATTG in 5 kb flanking	1926	CG divergent sites per bp in exons
413	proportion of A-T SNPs in introns plus 5 kb flanking relative to total num SNPs	1170	frequency of CATTT in 5 kb flanking	1927	CT divergent sites per bp in exons
414	proportion of C-G SNPs in introns plus 5 kb flanking relative to total num SNPs	1171	frequency of CCAAA in 5 kb flanking	1928	GT divergent sites per bp in exons
415	proportion of C-T SNPs in introns plus 5 kb flanking relative to total num SNPs	1172	frequency of CCAAC in 5 kb flanking	1929	proportion unalignable (gaps in human or other species) per bp in exons
416	proportion of G-T SNPs in introns plus 5 kb flanking relative to total num SNPs	1173	frequency of CCAAG in 5 kb flanking	1930	proportion unalignable (gaps in human) per bp in exons
417	A-C SNPs per bp in introns plus 5 kb flanking	1174	frequency of CCAAT in 5 kb flanking	1931	proportion unalignable (gaps in other species) per bp in exons
418	A-G SNPs per bp in introns plus 5 kb flanking	1175	frequency of CCACA in 5 kb flanking	1932	proportion of divergent sites per bp in introns
419	A-T SNPs per bp in introns plus 5 kb flanking	1176	frequency of CCACC in 5 kb flanking	1933	number of divergent sites in total in introns
420	C-G SNPs per bp in introns plus 5 kb flanking	1177	frequency of CCACG in 5 kb flanking	1934	proportion of divergent sites per bp, including gaps in introns
421	C-T SNPs per bp in introns plus 5 kb flanking	1178	frequency of CCACT in 5 kb flanking	1935	number of divergent sites in total, including gaps in introns

422	G-T SNPs per bp in introns plus 5 kb flanking	1179	frequency of CCAGA in 5 kb flanking	1936	proportion of divergent transitions per bp in introns
423	transitions per bp in introns plus 5 kb flanking	1180	frequency of CCAGC in 5 kb flanking	1937	number of divergent transitions in introns
424	transversions per bp introns plus 5 kb flanking	1181	frequency of CCAGG in 5 kb flanking	1938	proportion of divergent transversions per bp in introns
425	total snps per bp in introns plus 5 kb flanking	1182	frequency of CCAGT in 5 kb flanking	1939	number of divergent transversions in introns
426	number of synonymous snps	1183	frequency of CCATA in 5 kb flanking	1940	proportion of divergent sites (no gaps) that are transitions in introns
427	number of nonsynonymous snps	1184	frequency of CCATC in 5 kb flanking	1941	proportion of divergent sites that (no gaps) are transversions in introns
428	proportion of coding snps that are nonsynonymous ('dn/ds' like)	1185	frequency of CCATG in 5 kb flanking	1942	proportion of total divergent sites that are gaps, including gaps in introns
429	proportion of coding bp that are synonymous snps	1186	frequency of CCATT in 5 kb flanking	1943	proportion of AC divergent sites relative to total num divergent sites (no gaps) in introns
430	proportion of coding bp that are nonsynonymous snps	1187	frequency of CCCAA in 5 kb flanking	1944	proportion of AG divergent sites relative to total num divergent sites (no gaps) in introns
431	GC content in ccds	1188	frequency of CCCAC in 5 kb flanking	1945	proportion of AT divergent sites relative to total num divergent sites (no gaps) in introns
432	total % bases masked in ccds	1189	frequency of CCCAG in 5 kb flanking	1946	proportion of CG divergent sites relative to total num divergent sites (no gaps) in introns
433	number SINEs in ccds	1190	frequency of CCCAT in 5 kb flanking	1947	proportion of CT divergent sites relative to total num divergent sites (no gaps) in introns
434	percentage SINEs in ccds	1191	frequency of CCCCA in 5 kb flanking	1948	proportion of GT divergent sites relative to total num divergent sites (no gaps) in introns
435	number ALUs in ccds	1192	frequency of CCCCC in 5 kb flanking	1949	AC divergent sites per bp in introns
436	percentage ALUs in ccds	1193	frequency of CCCCG in 5 kb flanking	1950	AG divergent sites per bp in introns
437	number MIRs in ccds	1194	frequency of CCCCT in 5 kb flanking	1951	AT divergent sites per bp in introns
438	percentage MIRs in ccds	1195	frequency of CCCGA in 5 kb flanking	1952	CG divergent sites per bp in introns
439	number LINES in ccds	1196	frequency of CCCGC in 5 kb flanking	1953	CT divergent sites per bp in introns
440	percentage LINES in ccds	1197	frequency of CCCGG in 5 kb flanking	1954	GT divergent sites per bp in introns
441	number LINE1 in ccds	1198	frequency of CCCGT in 5 kb flanking	1955	proportion unalignable (gaps in human or other species) per bp in introns
442	percentage LINE1 in ccds	1199	frequency of CCCTA in 5 kb flanking	1956	proportion unalignable (gaps in human) per bp in introns
443	number LINE2 in ccds	1200	frequency of CCCTC in 5 kb flanking	1957	proportion unalignable (gaps in other species) per bp in introns
444	percentage LINE2 in ccds	1201	frequency of CCCTG in 5 kb flanking	1958	proportion of divergent sites per bp in 2 kb up
445	number L3/CR1 in ccds	1202	frequency of CCCTT in 5 kb flanking	1959	number of divergent sites in total in 2 kb up
446	percentage L3/CR1 in ccds	1203	frequency of CCGAA in 5 kb flanking	1960	proportion of divergent sites per bp, including gaps in 2 kb up
447	number LTR in ccds	1204	frequency of CCGAC in 5 kb flanking	1961	number of divergent sites in total, including gaps in 2 kb up
448	percentage LTR in ccds	1205	frequency of CCGAG in 5 kb flanking	1962	proportion of divergent transitions per bp in 2 kb up
449	number MaLRs in ccds	1206	frequency of CCGAT in 5 kb flanking	1963	number of divergent transitions in 2 kb up

450	percentage MaLRs in ccds	1207	frequency of CCGCA in 5 kb flanking	1964	proportion of divergent transversions per bp in 2 kb up
451	number ERVL in ccds	1208	frequency of CCGCC in 5 kb flanking	1965	number of divergent transversions in 2 kb up
452	percentage ERVL in ccds	1209	frequency of CCGCG in 5 kb flanking	1966	proportion of divergent sites (no gaps) that are transitions in 2 kb up
453	number ERV_classI in ccds	1210	frequency of CCGCT in 5 kb flanking	1967	proportion of divergent sites that (no gaps) are transversions in 2 kb up
454	percentage ERV_classI in ccds	1211	frequency of CCGGA in 5 kb flanking	1968	proportion of total divergent sites that are gaps, including gaps in 2 kb up
455	number ERV_classII in ccds	1212	frequency of CCGGC in 5 kb flanking	1969	proportion of AC divergent sites relative to total num divergent sites (no gaps) in 2 kb up
456	percentage ERV_classII in ccds	1213	frequency of CCGGG in 5 kb flanking	1970	proportion of AG divergent sites relative to total num divergent sites (no gaps) in 2 kb up
457	number DNA elements in ccds	1214	frequency of CCGGT in 5 kb flanking	1971	proportion of AT divergent sites relative to total num divergent sites (no gaps) in 2 kb up
458	percentage DNA elements in ccds	1215	frequency of CCGTA in 5 kb flanking	1972	proportion of CG divergent sites relative to total num divergent sites (no gaps) in 2 kb up
459	number MER1_type in ccds	1216	frequency of CCGTC in 5 kb flanking	1973	proportion of CT divergent sites relative to total num divergent sites (no gaps) in 2 kb up
460	percentage MER1_type in ccds	1217	frequency of CCGTG in 5 kb flanking	1974	proportion of GT divergent sites relative to total num divergent sites (no gaps) in 2 kb up
461	number MER2_type in ccds	1218	frequency of CCGTT in 5 kb flanking	1975	AC divergent sites per bp in 2 kb up
462	percentage MER2_type in ccds	1219	frequency of CCTAA in 5 kb flanking	1976	AG divergent sites per bp in 2 kb up
463	number unclassified in ccds	1220	frequency of CCTAC in 5 kb flanking	1977	AT divergent sites per bp in 2 kb up
464	percentage unclassified in ccds	1221	frequency of CCTAG in 5 kb flanking	1978	CG divergent sites per bp in 2 kb up
465	number small RNA in ccds	1222	frequency of CCTAT in 5 kb flanking	1979	CT divergent sites per bp in 2 kb up
466	percentage small RNA in ccds	1223	frequency of CCTCA in 5 kb flanking	1980	GT divergent sites per bp in 2 kb up
467	number satellites in ccds	1224	frequency of CCTCC in 5 kb flanking	1981	proportion unalignable (gaps in human or other species) per bp in 2 kb up
468	percentage satellites in ccds	1225	frequency of CCTCG in 5 kb flanking	1982	proportion unalignable (gaps in human) per bp in 2 kb up
469	number simple repeats in ccds	1226	frequency of CCTCT in 5 kb flanking	1983	proportion unalignable (gaps in other species) per bp in 2 kb up
470	percentage simple repeats in ccds	1227	frequency of CCTGA in 5 kb flanking	1984	proportion of divergent sites per bp in 2 kb down
471	number low complexity in ccds	1228	frequency of CCTGC in 5 kb flanking	1985	number of divergent sites in total in 2 kb down
472	percentage low complexity in ccds	1229	frequency of CCTGG in 5 kb flanking	1986	proportion of divergent sites per bp, including gaps in 2 kb down
473	total repeats (number SINES + LINES + LTR + DNA elements) in ccds	1230	frequency of CCTGT in 5 kb flanking	1987	number of divergent sites in total, including gaps in 2 kb down
474	GC content in 2kb upstream	1231	frequency of CCTTA in 5 kb flanking	1988	proportion of divergent transitions per bp in 2 kb down
475	total % bases masked in 2kb upstream	1232	frequency of CCTTC in 5 kb flanking	1989	number of divergent transitions in 2 kb down
476	number SINES in 2kb upstream	1233	frequency of CCTTG in 5 kb flanking	1990	proportion of divergent transversions per bp in 2 kb down
477	percentage SINES in 2kb upstream	1234	frequency of CCTTT in 5 kb flanking	1991	number of divergent transversions in 2 kb down

478	number ALUs in 2kb upstream	1235	frequency of CGAAA in 5 kb flanking	1992	proportion of divergent sites (no gaps) that are transitions in 2 kb down
479	percentage ALUs in 2kb upstream	1236	frequency of CGAAC in 5 kb flanking	1993	proportion of divergent sites that (no gaps) are transversions in 2 kb down
480	number MIRs in 2kb upstream	1237	frequency of CGAAG in 5 kb flanking	1994	proportion of total divergent sites that are gaps, including gaps in 2 kb down
481	percentage MIRs in 2kb upstream	1238	frequency of CGAAT in 5 kb flanking	1995	proportion of AC divergent sites relative to total num divergent sites (no gaps) in 2 kb down
482	number LINEs in 2kb upstream	1239	frequency of CGACA in 5 kb flanking	1996	proportion of AG divergent sites relative to total num divergent sites (no gaps) in 2 kb down
483	percentage LINEs in 2kb upstream	1240	frequency of CGACC in 5 kb flanking	1997	proportion of AT divergent sites relative to total num divergent sites (no gaps) in 2 kb down
484	number LINE1 in 2kb upstream	1241	frequency of CGACG in 5 kb flanking	1998	proportion of CG divergent sites relative to total num divergent sites (no gaps) in 2 kb down
485	percentage LINE1 in 2kb upstream	1242	frequency of CGACT in 5 kb flanking	1999	proportion of CT divergent sites relative to total num divergent sites (no gaps) in 2 kb down
486	number LINE2 in 2kb upstream	1243	frequency of CGAGA in 5 kb flanking	2000	proportion of GT divergent sites relative to total num divergent sites (no gaps) in 2 kb down
487	percentage LINE2 in 2kb upstream	1244	frequency of CGAGC in 5 kb flanking	2001	AC divergent sites per bp in 2 kb down
488	number L3/CR1 in 2kb upstream	1245	frequency of CGAGG in 5 kb flanking	2002	AG divergent sites per bp in 2 kb down
489	percentage L3/CR1 in 2kb upstream	1246	frequency of CGAGT in 5 kb flanking	2003	AT divergent sites per bp in 2 kb down
490	number LTR in 2kb upstream	1247	frequency of CGATA in 5 kb flanking	2004	CG divergent sites per bp in 2 kb down
491	percentage LTR in 2kb upstream	1248	frequency of CGATC in 5 kb flanking	2005	CT divergent sites per bp in 2 kb down
492	number MaLRs in 2kb upstream	1249	frequency of CGATG in 5 kb flanking	2006	GT divergent sites per bp in 2 kb down
493	percentage MaLRs in 2kb upstream	1250	frequency of CGATT in 5 kb flanking	2007	proportion unalignable (gaps in human or other species) per bp in 2 kb down
494	number ERVL in 2kb upstream	1251	frequency of CGCAA in 5 kb flanking	2008	proportion unalignable (gaps in human) per bp in 2 kb down
495	percentage ERVL in 2kb upstream	1252	frequency of CGCAC in 5 kb flanking	2009	proportion unalignable (gaps in other species) per bp in 2 kb down
496	number ERV_classII in 2kb upstream	1253	frequency of CGCAG in 5 kb flanking	2010	proportion of divergent sites per bp in 5 kb up
497	percentage ERV_classII in 2kb upstream	1254	frequency of CGCAT in 5 kb flanking	2011	number of divergent sites in total in 5 kb up
498	number MER1_type in 2kb upstream	1255	frequency of CGCCA in 5 kb flanking	2012	proportion of divergent sites per bp, including gaps in 5 kb up
499	percentage MER1_type in 2kb upstream	1256	frequency of CGCCC in 5 kb flanking	2013	number of divergent sites in total, including gaps in 5 kb up
500	number MER2_type in 2kb upstream	1257	frequency of CGCCG in 5 kb flanking	2014	proportion of divergent transitions per bp in 5 kb up
501	percentage MER2_type in 2kb upstream	1258	frequency of CGCCT in 5 kb flanking	2015	number of divergent transitions in 5 kb up
502	number unclassified in 2kb upstream	1259	frequency of CGCGA in 5 kb flanking	2016	proportion of divergent transversions per bp in 5 kb up
503	percentage unclassified in 2kb upstream	1260	frequency of CGCGC in 5 kb flanking	2017	number of divergent transversions in 5 kb up
504	number small RNA in 2kb upstream	1261	frequency of CGCGG in 5 kb flanking	2018	proportion of divergent sites (no gaps) that are transitions in 5 kb up
505	percentage small RNA in 2kb upstream	1262	frequency of CGCGT in 5 kb flanking	2019	proportion of divergent sites that (no gaps) are transversions in 5 kb up

506	number satellites in 2kb upstream	1263	frequency of CGCTA in 5 kb flanking	2020	proportion of total divergent sites that are gaps, including gaps in 5 kb up
507	percentage satellites in 2kb upstream	1264	frequency of CGCTC in 5 kb flanking	2021	proportion of AC divergent sites relative to total num divergent sites (no gaps) in 5 kb up
508	number simple repeats in 2kb upstream	1265	frequency of CGCTG in 5 kb flanking	2022	proportion of AG divergent sites relative to total num divergent sites (no gaps) in 5 kb up
509	percentage simple repeats in 2kb upstream	1266	frequency of CGCTT in 5 kb flanking	2023	proportion of AT divergent sites relative to total num divergent sites (no gaps) in 5 kb up
510	number low complexity in 2kb upstream	1267	frequency of CGGAA in 5 kb flanking	2024	proportion of CG divergent sites relative to total num divergent sites (no gaps) in 5 kb up
511	percentage low complexity in 2kb upstream	1268	frequency of CGGAC in 5 kb flanking	2025	proportion of CT divergent sites relative to total num divergent sites (no gaps) in 5 kb up
512	total repeats (number SINEs + LINEs + LTR + DNA elements) in 2kb upstream	1269	frequency of CGGAG in 5 kb flanking	2026	proportion of GT divergent sites relative to total num divergent sites (no gaps) in 5 kb up
513	GC content in 2 kb downstream	1270	frequency of CGGAT in 5 kb flanking	2027	AC divergent sites per bp in 5 kb up
514	total % bases masked in 2 kb downstream	1271	frequency of CGGCA in 5 kb flanking	2028	AG divergent sites per bp in 5 kb up
515	number SINEs in 2 kb downstream	1272	frequency of CGGCC in 5 kb flanking	2029	AT divergent sites per bp in 5 kb up
516	percentage SINEs in 2 kb downstream	1273	frequency of CGGCG in 5 kb flanking	2030	CG divergent sites per bp in 5 kb up
517	number ALUs in 2 kb downstream	1274	frequency of CGGCT in 5 kb flanking	2031	CT divergent sites per bp in 5 kb up
518	percentage ALUs in 2 kb downstream	1275	frequency of CGGGA in 5 kb flanking	2032	GT divergent sites per bp in 5 kb up
519	number MIRs in 2 kb downstream	1276	frequency of CGGGC in 5 kb flanking	2033	proportion unalignable (gaps in human or other species) per bp in 5 kb up
520	percentage MIRs in 2 kb downstream	1277	frequency of CGGGG in 5 kb flanking	2034	proportion unalignable (gaps in human) per bp in 5 kb up
521	number LINEs in 2 kb downstream	1278	frequency of CGGGT in 5 kb flanking	2035	proportion unalignable (gaps in other species) per bp in 5 kb up
522	percentage LINEs in 2 kb downstream	1279	frequency of CGGTA in 5 kb flanking	2036	proportion of divergent sites per bp in 5 kb down
523	number LINE1 in 2 kb downstream	1280	frequency of CGGTC in 5 kb flanking	2037	number of divergent sites in total in 5 kb down
524	percentage LINE1 in 2 kb downstream	1281	frequency of CGGTG in 5 kb flanking	2038	proportion of divergent sites per bp, including gaps in 5 kb down
525	number LINE2 in 2 kb downstream	1282	frequency of CGGTT in 5 kb flanking	2039	number of divergent sites in total, including gaps in 5 kb down
526	percentage LINE2 in 2 kb downstream	1283	frequency of CGTAA in 5 kb flanking	2040	proportion of divergent transitions per bp in 5 kb down
527	number L3/CR1 in 2 kb downstream	1284	frequency of CGTAC in 5 kb flanking	2041	number of divergent transitions in 5 kb down
528	percentage L3/CR1 in 2 kb downstream	1285	frequency of CGTAG in 5 kb flanking	2042	proportion of divergent transversions per bp in 5 kb down
529	number LTR in 2 kb downstream	1286	frequency of CGTAT in 5 kb flanking	2043	number of divergent transversions in 5 kb down
530	percentage LTR in 2 kb downstream	1287	frequency of CGTCA in 5 kb flanking	2044	proportion of divergent sites (no gaps) that are transitions in 5 kb down
531	number MaLRs in 2 kb downstream	1288	frequency of CGTCC in 5 kb flanking	2045	proportion of divergent sites that (no gaps) are transversions in 5 kb down
532	percentage MaLRs in 2 kb downstream	1289	frequency of CGTCG in 5 kb flanking	2046	proportion of total divergent sites that are gaps, including gaps in 5 kb down

533	number ERVL in 2 kb downstream	1290	frequency of CGTCT in 5 kb flanking	2047	proportion of AC divergent sites relative to total num divergent sites (no gaps) in 5 kb down
534	percentage ERVL in 2 kb downstream	1291	frequency of CGTGA in 5 kb flanking	2048	proportion of AG divergent sites relative to total num divergent sites (no gaps) in 5 kb down
535	number ERV_classI in 2 kb downstream	1292	frequency of CGTGC in 5 kb flanking	2049	proportion of AT divergent sites relative to total num divergent sites (no gaps) in 5 kb down
536	percentage ERV_classI in 2 kb downstream	1293	frequency of CGTGG in 5 kb flanking	2050	proportion of CG divergent sites relative to total num divergent sites (no gaps) in 5 kb down
537	number ERV_classII in 2 kb downstream	1294	frequency of CGTGT in 5 kb flanking	2051	proportion of CT divergent sites relative to total num divergent sites (no gaps) in 5 kb down
538	percentage ERV_classII in 2 kb downstream	1295	frequency of CGTTA in 5 kb flanking	2052	proportion of GT divergent sites relative to total num divergent sites (no gaps) in 5 kb down
539	number MER1_type in 2 kb downstream	1296	frequency of CGTTC in 5 kb flanking	2053	AC divergent sites per bp in 5 kb down
540	percentage MER1_type in 2 kb downstream	1297	frequency of CGTTG in 5 kb flanking	2054	AG divergent sites per bp in 5 kb down
541	number MER2_type in 2 kb downstream	1298	frequency of CGTTT in 5 kb flanking	2055	AT divergent sites per bp in 5 kb down
542	percentage MER2_type in 2 kb downstream	1299	frequency of CTAAC in 5 kb flanking	2056	CG divergent sites per bp in 5 kb down
543	number unclassified in 2 kb downstream	1300	frequency of CTAAC in 5 kb flanking	2057	CT divergent sites per bp in 5 kb down
544	percentage unclassified in 2 kb downstream	1301	frequency of CTAAG in 5 kb flanking	2058	GT divergent sites per bp in 5 kb down
545	number small RNA in 2 kb downstream	1302	frequency of CTAAT in 5 kb flanking	2059	proportion unalignable (gaps in human or other species) per bp in 5 kb down
546	percentage small RNA in 2 kb downstream	1303	frequency of CTACA in 5 kb flanking	2060	proportion unalignable (gaps in human) per bp in 5 kb down
547	number satellites in 2 kb downstream	1304	frequency of CTACC in 5 kb flanking	2061	proportion unalignable (gaps in other species) per bp in 5 kb down
548	percentage satellites in 2 kb downstream	1305	frequency of CTACG in 5 kb flanking	2062	proportion of divergent sites per bp in 10 kb up
549	number simple repeats in 2 kb downstream	1306	frequency of CTAAT in 5 kb flanking	2063	number of divergent sites in total in 10 kb up
550	percentage simple repeats in 2 kb downstream	1307	frequency of CTAGA in 5 kb flanking	2064	proportion of divergent sites per bp, including gaps in 10 kb up
551	number low complexity in 2 kb downstream	1308	frequency of CTAGC in 5 kb flanking	2065	number of divergent sites in total, including gaps in 10 kb up
552	percentage low complexity in 2 kb downstream	1309	frequency of CTAGG in 5 kb flanking	2066	proportion of divergent transitions per bp in 10 kb up
553	total repeats (number SINES + LINES + LTR + DNA elements) in 2 kb downstream	1310	frequency of CTAGT in 5 kb flanking	2067	number of divergent transitions in 10 kb up
554	GC content in 5 kb upstream	1311	frequency of CTATA in 5 kb flanking	2068	proportion of divergent transversions per bp in 10 kb up
555	total % bases masked in 5 kb upstream	1312	frequency of CTATC in 5 kb flanking	2069	number of divergent transversions in 10 kb up
556	number SINES in 5 kb upstream	1313	frequency of CTATG in 5 kb flanking	2070	proportion of divergent sites (no gaps) that are transitions in 10 kb up
557	percentage SINES in 5 kb upstream	1314	frequency of CTATT in 5 kb flanking	2071	proportion of divergent sites that (no gaps) are transversions in 10 kb up
558	number ALUs in 5 kb upstream	1315	frequency of CTCAA in 5 kb flanking	2072	proportion of total divergent sites that are gaps, including gaps in 10 kb up

559	percentage ALUs in 5 kb upstream	1316	frequency of CTCAC in 5 kb flanking	2073	proportion of AC divergent sites relative to total num divergent sites (no gaps) in 10 kb up
560	number MIRs in 5 kb upstream	1317	frequency of CTCAG in 5 kb flanking	2074	proportion of AG divergent sites relative to total num divergent sites (no gaps) in 10 kb up
561	percentage MIRs in 5 kb upstream	1318	frequency of CTCAT in 5 kb flanking	2075	proportion of AT divergent sites relative to total num divergent sites (no gaps) in 10 kb up
562	number LINES in 5 kb upstream	1319	frequency of CTCCA in 5 kb flanking	2076	proportion of CG divergent sites relative to total num divergent sites (no gaps) in 10 kb up
563	percentage LINES in 5 kb upstream	1320	frequency of CTCCC in 5 kb flanking	2077	proportion of CT divergent sites relative to total num divergent sites (no gaps) in 10 kb up
564	number LINE1 in 5 kb upstream	1321	frequency of CTCCG in 5 kb flanking	2078	proportion of GT divergent sites relative to total num divergent sites (no gaps) in 10 kb up
565	percentage LINE1 in 5 kb upstream	1322	frequency of CTCTT in 5 kb flanking	2079	AC divergent sites per bp in 10 kb up
566	number LINE2 in 5 kb upstream	1323	frequency of CTCGA in 5 kb flanking	2080	AG divergent sites per bp in 10 kb up
567	percentage LINE2 in 5 kb upstream	1324	frequency of CTCGC in 5 kb flanking	2081	AT divergent sites per bp in 10 kb up
568	number L3/CR1 in 5 kb upstream	1325	frequency of CTCGG in 5 kb flanking	2082	CG divergent sites per bp in 10 kb up
569	percentage L3/CR1 in 5 kb upstream	1326	frequency of CTCGT in 5 kb flanking	2083	CT divergent sites per bp in 10 kb up
570	number LTR in 5 kb upstream	1327	frequency of CTCTA in 5 kb flanking	2084	GT divergent sites per bp in 10 kb up
571	percentage LTR in 5 kb upstream	1328	frequency of CTCTC in 5 kb flanking	2085	proportion unalignable (gaps in human or other species) per bp in 10 kb up
572	number MaLRs in 5 kb upstream	1329	frequency of CTCTG in 5 kb flanking	2086	proportion unalignable (gaps in human) per bp in 10 kb up
573	percentage MaLRs in 5 kb upstream	1330	frequency of CTCTT in 5 kb flanking	2087	proportion unalignable (gaps in other species) per bp in 10 kb up
574	number ERVL in 5 kb upstream	1331	frequency of CTGAA in 5 kb flanking	2088	proportion of divergent sites per bp in 10 kb down
575	percentage ERVL in 5 kb upstream	1332	frequency of CTGAC in 5 kb flanking	2089	number of divergent sites in total in 10 kb down
576	number ERV_classI in 5 kb upstream	1333	frequency of CTGAG in 5 kb flanking	2090	proportion of divergent sites per bp, including gaps in 10 kb down
577	percentage ERV_classI in 5 kb upstream	1334	frequency of CTGAT in 5 kb flanking	2091	number of divergent sites in total, including gaps in 10 kb down
578	number ERV_classII in 5 kb upstream	1335	frequency of CTGCA in 5 kb flanking	2092	proportion of divergent transitions per bp in 10 kb down
579	percentage ERV_classII in 5 kb upstream	1336	frequency of CTGCC in 5 kb flanking	2093	number of divergent transitions in 10 kb down
580	number DNA elements in 5 kb upstream	1337	frequency of CTGCG in 5 kb flanking	2094	proportion of divergent transversions per bp in 10 kb down
581	percentage DNA elements in 5 kb upstream	1338	frequency of CTGCT in 5 kb flanking	2095	number of divergent transversions in 10 kb down
582	number MER1_type in 5 kb upstream	1339	frequency of CTGGA in 5 kb flanking	2096	proportion of divergent sites (no gaps) that are transitions in 10 kb down
583	percentage MER1_type in 5 kb upstream	1340	frequency of CTGGC in 5 kb flanking	2097	proportion of divergent sites that (no gaps) are transversions in 10 kb down
584	number MER2_type in 5 kb upstream	1341	frequency of CTGGG in 5 kb flanking	2098	proportion of total divergent sites that are gaps, including gaps in 10 kb down
585	percentage MER2_type in 5 kb upstream	1342	frequency of CTGGT in 5 kb flanking	2099	proportion of AC divergent sites relative to total num divergent sites (no gaps) in 10 kb down

586	number unclassified in 5 kb upstream	1343	frequency of CTGTA in 5 kb flanking	2100	proportion of AG divergent sites relative to total num divergent sites (no gaps) in 10 kb down
587	percentage unclassified in 5 kb upstream	1344	frequency of CTGTC in 5 kb flanking	2101	proportion of AT divergent sites relative to total num divergent sites (no gaps) in 10 kb down
588	number small RNA in 5 kb upstream	1345	frequency of CTGTG in 5 kb flanking	2102	proportion of CG divergent sites relative to total num divergent sites (no gaps) in 10 kb down
589	percentage small RNA in 5 kb upstream	1346	frequency of CTGTT in 5 kb flanking	2103	proportion of CT divergent sites relative to total num divergent sites (no gaps) in 10 kb down
590	number satellites in 5 kb upstream	1347	frequency of CTTAA in 5 kb flanking	2104	proportion of GT divergent sites relative to total num divergent sites (no gaps) in 10 kb down
591	percentage satellites in 5 kb upstream	1348	frequency of CTTAC in 5 kb flanking	2105	AC divergent sites per bp in 10 kb down
592	number simple repeats in 5 kb upstream	1349	frequency of CTTAG in 5 kb flanking	2106	AG divergent sites per bp in 10 kb down
593	percentage simple repeats in 5 kb upstream	1350	frequency of CTTAT in 5 kb flanking	2107	AT divergent sites per bp in 10 kb down
594	number low complexity in 5 kb upstream	1351	frequency of CTTCA in 5 kb flanking	2108	CG divergent sites per bp in 10 kb down
595	percentage low complexity in 5 kb upstream	1352	frequency of CTTCC in 5 kb flanking	2109	CT divergent sites per bp in 10 kb down
596	total repeats (number SINES + LINES + LTR + DNA elements) in 5 kb upstream	1353	frequency of CTTCG in 5 kb flanking	2110	GT divergent sites per bp in 10 kb down
597	GC content in 5 kb downstream	1354	frequency of CTTCT in 5 kb flanking	2111	proportion unalignable (gaps in human or other species) per bp in 10 kb down
598	total % bases masked in 5 kb downstream	1355	frequency of CTTGA in 5 kb flanking	2112	proportion unalignable (gaps in human) per bp in 10 kb down
599	number SINES in 5 kb downstream	1356	frequency of CTTGC in 5 kb flanking	2113	proportion unalignable (gaps in other species) per bp in 10 kb down
600	percentage SINES in 5 kb downstream	1357	frequency of CTTGG in 5 kb flanking	2114	proportion of divergent sites per bp in ccds + 2kb flank
601	number ALUs in 5 kb downstream	1358	frequency of CTTGT in 5 kb flanking	2115	number of divergent sites in total in ccds + 2kb flank
602	percentage ALUs in 5 kb downstream	1359	frequency of CTTTA in 5 kb flanking	2116	proportion of divergent sites per bp, including gaps in ccds + 2kb flank
603	number MIRs in 5 kb downstream	1360	frequency of CTTTC in 5 kb flanking	2117	number of divergent sites in total, including gaps in ccds + 2kb flank
604	percentage MIRs in 5 kb downstream	1361	frequency of CTTTG in 5 kb flanking	2118	proportion of divergent transitions per bp in ccds + 2kb flank
605	number LINES in 5 kb downstream	1362	frequency of CTTTT in 5 kb flanking	2119	number of divergent transitions in ccds + 2kb flank
606	percentage LINES in 5 kb downstream	1363	frequency of GAAAA in 5 kb flanking	2120	proportion of divergent transversions per bp in ccds + 2kb flank
607	number LINE1 in 5 kb downstream	1364	frequency of GAAAC in 5 kb flanking	2121	number of divergent transversions in ccds + 2kb flank
608	percentage LINE1 in 5 kb downstream	1365	frequency of GAAAG in 5 kb flanking	2122	proportion of divergent sites (no gaps) that are transitions in ccds + 2kb flank
609	number LINE2 in 5 kb downstream	1366	frequency of GAAAT in 5 kb flanking	2123	proportion of divergent sites that (no gaps) are transversions in ccds + 2kb flank
610	percentage LINE2 in 5 kb downstream	1367	frequency of GAACA in 5 kb flanking	2124	proportion of total divergent sites that are gaps, including gaps in ccds + 2kb flank
611	number L3/CR1 in 5 kb downstream	1368	frequency of GAACC in 5 kb flanking	2125	proportion of AC divergent sites relative to total num divergent sites (no gaps) in ccds + 2kb flank



612	percentage L3/CR1 in 5 kb downstream	1369	frequency of GAACG in 5 kb flanking	2126	proportion of AG divergent sites relative to total num divergent sites (no gaps) in ccds + 2kb flank
613	number LTR in 5 kb downstream	1370	frequency of GAACT in 5 kb flanking	2127	proportion of AT divergent sites relative to total num divergent sites (no gaps) in ccds + 2kb flank
614	percentage LTR in 5 kb downstream	1371	frequency of GAAGA in 5 kb flanking	2128	proportion of CG divergent sites relative to total num divergent sites (no gaps) in ccds + 2kb flank
615	number MaLRs in 5 kb downstream	1372	frequency of GAAGC in 5 kb flanking	2129	proportion of CT divergent sites relative to total num divergent sites (no gaps) in ccds + 2kb flank
616	percentage MaLRs in 5 kb downstream	1373	frequency of GAAGG in 5 kb flanking	2130	proportion of GT divergent sites relative to total num divergent sites (no gaps) in ccds + 2kb flank
617	number ERVL in 5 kb downstream	1374	frequency of GAAGT in 5 kb flanking	2131	AC divergent sites per bp in ccds + 2kb flank
618	percentage ERVL in 5 kb downstream	1375	frequency of GAATA in 5 kb flanking	2132	AG divergent sites per bp in ccds + 2kb flank
619	number ERV_classI in 5 kb downstream	1376	frequency of GAATC in 5 kb flanking	2133	AT divergent sites per bp in ccds + 2kb flank
620	percentage ERV_classI in 5 kb downstream	1377	frequency of GAATG in 5 kb flanking	2134	CG divergent sites per bp in ccds + 2kb flank
621	number ERV_classII in 5 kb downstream	1378	frequency of GAATT in 5 kb flanking	2135	CT divergent sites per bp in ccds + 2kb flank
622	percentage ERV_classII in 5 kb downstream	1379	frequency of GACAA in 5 kb flanking	2136	GT divergent sites per bp in ccds + 2kb flank
623	number MER1_type in 5 kb downstream	1380	frequency of GACAC in 5 kb flanking	2137	proportion unalignable (gaps in human or other species) per bp in ccds + 2kb flank
624	percentage MER1_type in 5 kb downstream	1381	frequency of GACAG in 5 kb flanking	2138	proportion unalignable (gaps in human) per bp in ccds + 2kb flank
625	number MER2_type in 5 kb downstream	1382	frequency of GACAT in 5 kb flanking	2139	proportion unalignable (gaps in other species) per bp in ccds + 2kb flank
626	percentage MER2_type in 5 kb downstream	1383	frequency of GACCA in 5 kb flanking	2140	proportion of divergent sites per bp in ccds + 5 kb flank
627	number unclassified in 5 kb downstream	1384	frequency of GACCC in 5 kb flanking	2141	number of divergent sites in total in ccds + 5 kb flank
628	percentage unclassified in 5 kb downstream	1385	frequency of GACCG in 5 kb flanking	2142	proportion of divergent sites per bp, including gaps in ccds + 5 kb flank
629	number small RNA in 5 kb downstream	1386	frequency of GACCT in 5 kb flanking	2143	number of divergent sites in total, including gaps in ccds + 5 kb flank
630	percentage small RNA in 5 kb downstream	1387	frequency of GACGA in 5 kb flanking	2144	proportion of divergent transitions per bp in ccds + 5 kb flank
631	number satellites in 5 kb downstream	1388	frequency of GACGC in 5 kb flanking	2145	number of divergent transitions in ccds + 5 kb flank
632	percentage satellites in 5 kb downstream	1389	frequency of GACGG in 5 kb flanking	2146	proportion of divergent transversions per bp in ccds + 5 kb flank
633	number simple repeats in 5 kb downstream	1390	frequency of GACGT in 5 kb flanking	2147	number of divergent transversions in ccds + 5 kb flank
634	percentage simple repeats in 5 kb downstream	1391	frequency of GACTA in 5 kb flanking	2148	proportion of divergent sites (no gaps) that are transitions in ccds + 5 kb flank
635	number low complexity in 5 kb downstream	1392	frequency of GACTC in 5 kb flanking	2149	proportion of divergent sites that (no gaps) are transversions in ccds + 5 kb flank
636	percentage low complexity in 5 kb downstream	1393	frequency of GACTG in 5 kb flanking	2150	proportion of total divergent sites that are gaps, including gaps in ccds + 5 kb flank
637	total repeats (number SINEs + LINEs + LTR + DNA elements) in 5 kb downstream	1394	frequency of GACTT in 5 kb flanking	2151	proportion of AC divergent sites relative to total num divergent sites (no gaps) in ccds + 5 kb flank

638	GC content in 10 kb upstream	1395	frequency of GAGAA in 5 kb flanking	2152	proportion of AG divergent sites relative to total num divergent sites (no gaps) in ccds + 5 kb flank
639	total % bases masked in 10 kb upstream	1396	frequency of GAGAC in 5 kb flanking	2153	proportion of AT divergent sites relative to total num divergent sites (no gaps) in ccds + 5 kb flank
640	number SINEs in 10 kb upstream	1397	frequency of GAGAG in 5 kb flanking	2154	proportion of CG divergent sites relative to total num divergent sites (no gaps) in ccds + 5 kb flank
641	percentage SINEs in 10 kb upstream	1398	frequency of GAGAT in 5 kb flanking	2155	proportion of CT divergent sites relative to total num divergent sites (no gaps) in ccds + 5 kb flank
642	number ALUs in 10 kb upstream	1399	frequency of GAGCA in 5 kb flanking	2156	proportion of GT divergent sites relative to total num divergent sites (no gaps) in ccds + 5 kb flank
643	percentage ALUs in 10 kb upstream	1400	frequency of GAGCC in 5 kb flanking	2157	AC divergent sites per bp in ccds + 5 kb flank
644	number MIRs in 10 kb upstream	1401	frequency of GAGCG in 5 kb flanking	2158	AG divergent sites per bp in ccds + 5 kb flank
645	percentage MIRs in 10 kb upstream	1402	frequency of GAGCT in 5 kb flanking	2159	AT divergent sites per bp in ccds + 5 kb flank
646	number LINEs in 10 kb upstream	1403	frequency of GAGGA in 5 kb flanking	2160	CG divergent sites per bp in ccds + 5 kb flank
647	percentage LINEs in 10 kb upstream	1404	frequency of GAGGC in 5 kb flanking	2161	CT divergent sites per bp in ccds + 5 kb flank
648	number LINE1 in 10 kb upstream	1405	frequency of GAGGG in 5 kb flanking	2162	GT divergent sites per bp in ccds + 5 kb flank
649	percentage LINE1 in 10 kb upstream	1406	frequency of GAGGT in 5 kb flanking	2163	proportion unalignable (gaps in human or other species) per bp in ccds + 5 kb flank
650	number LINE2 in 10 kb upstream	1407	frequency of GAGTA in 5 kb flanking	2164	proportion unalignable (gaps in human) per bp in ccds + 5 kb flank
651	percentage LINE2 in 10 kb upstream	1408	frequency of GAGTC in 5 kb flanking	2165	proportion unalignable (gaps in other species) per bp in ccds + 5 kb flank
652	number L3/CR1 in 10 kb upstream	1409	frequency of GAGTG in 5 kb flanking	2166	proportion of divergent sites per bp in ccds + 10 kb flank
653	percentage L3/CR1 in 10 kb upstream	1410	frequency of GAGTT in 5 kb flanking	2167	number of divergent sites in total in ccds + 10 kb flank
654	number LTR in 10 kb upstream	1411	frequency of GATAA in 5 kb flanking	2168	proportion of divergent sites per bp, including gaps in ccds + 10 kb flank
655	percentage LTR in 10 kb upstream	1412	frequency of GATAC in 5 kb flanking	2169	number of divergent sites in total, including gaps in ccds + 10 kb flank
656	number MaLRs in 10 kb upstream	1413	frequency of GATAG in 5 kb flanking	2170	proportion of divergent transitions per bp in ccds + 10 kb flank
657	percentage MaLRs in 10 kb upstream	1414	frequency of GATAT in 5 kb flanking	2171	number of divergent transitions in ccds + 10 kb flank
658	number ERVL in 10 kb upstream	1415	frequency of GATCA in 5 kb flanking	2172	proportion of divergent transversions per bp in ccds + 10 kb flank
659	percentage ERVL in 10 kb upstream	1416	frequency of GATCC in 5 kb flanking	2173	number of divergent transversions in ccds + 10 kb flank
660	number ERV_classI in 10 kb upstream	1417	frequency of GATCG in 5 kb flanking	2174	proportion of divergent sites (no gaps) that are transitions in ccds + 10 kb flank
661	percentage ERV_classI in 10 kb upstream	1418	frequency of GATCT in 5 kb flanking	2175	proportion of divergent sites that (no gaps) are transversions in ccds + 10 kb flank
662	number ERV_classII in 10 kb upstream	1419	frequency of GATGA in 5 kb flanking	2176	proportion of total divergent sites that are gaps, including gaps in ccds + 10 kb flank
663	percentage ERV_classII in 10 kb upstream	1420	frequency of GATGC in 5 kb flanking	2177	proportion of AC divergent sites relative to total num divergent sites (no gaps) in ccds + 10 kb flank

664	number DNA elements in 10 kb upstream	1421	frequency of GATGG in 5 kb flanking	2178	proportion of AG divergent sites relative to total num divergent sites (no gaps) in ccds + 10 kb flank
665	percentage DNA elements in 10 kb upstream	1422	frequency of GATGT in 5 kb flanking	2179	proportion of AT divergent sites relative to total num divergent sites (no gaps) in ccds + 10 kb flank
666	number MER1_type in 10 kb upstream	1423	frequency of GATTA in 5 kb flanking	2180	proportion of CG divergent sites relative to total num divergent sites (no gaps) in ccds + 10 kb flank
667	percentage MER1_type in 10 kb upstream	1424	frequency of GATTC in 5 kb flanking	2181	proportion of CT divergent sites relative to total num divergent sites (no gaps) in ccds + 10 kb flank
668	number MER2_type in 10 kb upstream	1425	frequency of GATTG in 5 kb flanking	2182	proportion of GT divergent sites relative to total num divergent sites (no gaps) in ccds + 10 kb flank
669	percentage MER2_type in 10 kb upstream	1426	frequency of GATTT in 5 kb flanking	2183	AC divergent sites per bp in ccds + 10 kb flank
670	number unclassified in 10 kb upstream	1427	frequency of GCAAA in 5 kb flanking	2184	AG divergent sites per bp in ccds + 10 kb flank
671	percentage unclassified in 10 kb upstream	1428	frequency of GCAAC in 5 kb flanking	2185	AT divergent sites per bp in ccds + 10 kb flank
672	number small RNA in 10 kb upstream	1429	frequency of GCAAG in 5 kb flanking	2186	CG divergent sites per bp in ccds + 10 kb flank
673	percentage small RNA in 10 kb upstream	1430	frequency of GCAAT in 5 kb flanking	2187	CT divergent sites per bp in ccds + 10 kb flank
674	number satellites in 10 kb upstream	1431	frequency of GCACA in 5 kb flanking	2188	GT divergent sites per bp in ccds + 10 kb flank
675	percentage satellites in 10 kb upstream	1432	frequency of GCACC in 5 kb flanking	2189	proportion unalignable (gaps in human or other species) per bp in ccds + 10 kb flank
676	number simple repeats in 10 kb upstream	1433	frequency of GCACG in 5 kb flanking	2190	proportion unalignable (gaps in human) per bp in ccds + 10 kb flank
677	percentage simple repeats in 10 kb upstream	1434	frequency of GCACT in 5 kb flanking	2191	proportion unalignable (gaps in other species) per bp in ccds + 10 kb flank
678	number low complexity in 10 kb upstream	1435	frequency of GCAGA in 5 kb flanking	2192	proportion of divergent sites per bp in 2 kb flank
679	percentage low complexity in 10 kb upstream	1436	frequency of GCAGC in 5 kb flanking	2193	number of divergent sites in total in 2 kb flank
680	total repeats (number SINES + LINES + LTR + DNA elements) in 10 kb upstream	1437	frequency of GCAGG in 5 kb flanking	2194	proportion of divergent sites per bp, including gaps in 2 kb flank
681	GC content in 10 kb downstream	1438	frequency of GCAGT in 5 kb flanking	2195	number of divergent sites in total, including gaps in 2 kb flank
682	total % bases masked in 10 kb downstream	1439	frequency of GCATA in 5 kb flanking	2196	proportion of divergent transitions per bp in 2 kb flank
683	number SINES in 10 kb downstream	1440	frequency of GCATC in 5 kb flanking	2197	number of divergent transitions in 2 kb flank
684	percentage SINES in 10 kb downstream	1441	frequency of GCATG in 5 kb flanking	2198	proportion of divergent transversions per bp in 2 kb flank
685	number ALUs in 10 kb downstream	1442	frequency of GCATT in 5 kb flanking	2199	number of divergent transversions in 2 kb flank
686	percentage ALUs in 10 kb downstream	1443	frequency of GCCAA in 5 kb flanking	2200	proportion of divergent sites (no gaps) that are transitions in 2 kb flank
687	number MIRs in 10 kb downstream	1444	frequency of GCCAC in 5 kb flanking	2201	proportion of divergent sites that (no gaps) are transversions in 2 kb flank
688	percentage MIRs in 10 kb downstream	1445	frequency of GCCAG in 5 kb flanking	2202	proportion of total divergent sites that are gaps, including gaps in 2 kb flank
689	number LINES in 10 kb downstream	1446	frequency of GCCAT in 5 kb flanking	2203	proportion of AC divergent sites relative to total num divergent sites (no gaps) in 2 kb flank

690	percentage LINEs in 10 kb downstream	1447	frequency of GCCCA in 5 kb flanking	2204	proportion of AG divergent sites relative to total num divergent sites (no gaps) in 2 kb flank
691	number LINE1 in 10 kb downstream	1448	frequency of GCCCC in 5 kb flanking	2205	proportion of AT divergent sites relative to total num divergent sites (no gaps) in 2 kb flank
692	percentage LINE1 in 10 kb downstream	1449	frequency of GCCCG in 5 kb flanking	2206	proportion of CG divergent sites relative to total num divergent sites (no gaps) in 2 kb flank
693	number LINE2 in 10 kb downstream	1450	frequency of GCCCT in 5 kb flanking	2207	proportion of CT divergent sites relative to total num divergent sites (no gaps) in 2 kb flank
694	percentage LINE2 in 10 kb downstream	1451	frequency of GCCGA in 5 kb flanking	2208	proportion of GT divergent sites relative to total num divergent sites (no gaps) in 2 kb flank
695	number L3/CR1 in 10 kb downstream	1452	frequency of GCCGC in 5 kb flanking	2209	AC divergent sites per bp in 2 kb flank
696	percentage L3/CR1 in 10 kb downstream	1453	frequency of GCCGG in 5 kb flanking	2210	AG divergent sites per bp in 2 kb flank
697	number LTR in 10 kb downstream	1454	frequency of GCCGT in 5 kb flanking	2211	AT divergent sites per bp in 2 kb flank
698	percentage LTR in 10 kb downstream	1455	frequency of GCCTA in 5 kb flanking	2212	CG divergent sites per bp in 2 kb flank
699	number MaLRs in 10 kb downstream	1456	frequency of GCCTC in 5 kb flanking	2213	CT divergent sites per bp in 2 kb flank
700	percentage MaLRs in 10 kb downstream	1457	frequency of GCCTG in 5 kb flanking	2214	GT divergent sites per bp in 2 kb flank
701	number ERVL in 10 kb downstream	1458	frequency of GCCTT in 5 kb flanking	2215	proportion unalignable (gaps in human or other species) per bp in 2 kb flank
702	percentage ERVL in 10 kb downstream	1459	frequency of GCGAA in 5 kb flanking	2216	proportion unalignable (gaps in human) per bp in 2 kb flank
703	number ERV_classI in 10 kb downstream	1460	frequency of GCGAC in 5 kb flanking	2217	proportion unalignable (gaps in other species) per bp in 2 kb flank
704	percentage ERV_classI in 10 kb downstream	1461	frequency of GCGAG in 5 kb flanking	2218	proportion of divergent sites per bp in 5 kb flank
705	number ERV_classII in 10 kb downstream	1462	frequency of GCGAT in 5 kb flanking	2219	number of divergent sites in total in 5 kb flank
706	percentage ERV_classII in 10 kb downstream	1463	frequency of GCGCA in 5 kb flanking	2220	proportion of divergent sites per bp, including gaps in 5 kb flank
707	number MER1_type in 10 kb downstream	1464	frequency of GCGCC in 5 kb flanking	2221	number of divergent sites in total, including gaps in 5 kb flank
708	percentage MER1_type in 10 kb downstream	1465	frequency of GCGCG in 5 kb flanking	2222	proportion of divergent transitions per bp in 5 kb flank
709	number MER2_type in 10 kb downstream	1466	frequency of GCGCT in 5 kb flanking	2223	number of divergent transitions in 5 kb flank
710	percentage MER2_type in 10 kb downstream	1467	frequency of GCGGA in 5 kb flanking	2224	proportion of divergent transversions per bp in 5 kb flank
711	number unclassified in 10 kb downstream	1468	frequency of GCGGC in 5 kb flanking	2225	number of divergent transversions in 5 kb flank
712	percentage unclassified in 10 kb downstream	1469	frequency of GCGGG in 5 kb flanking	2226	proportion of divergent sites (no gaps) that are transitions in 5 kb flank
713	number small RNA in 10 kb downstream	1470	frequency of GCGGT in 5 kb flanking	2227	proportion of divergent sites that (no gaps) are transversions in 5 kb flank
714	percentage small RNA in 10 kb downstream	1471	frequency of GCGTA in 5 kb flanking	2228	proportion of total divergent sites that are gaps, including gaps in 5 kb flank
715	number satellites in 10 kb downstream	1472	frequency of GCGTC in 5 kb flanking	2229	proportion of AC divergent sites relative to total num divergent sites (no gaps) in 5 kb flank
716	percentage satellites in 10 kb downstream	1473	frequency of GCGTG in 5 kb flanking	2230	proportion of AG divergent sites relative to total num divergent sites (no gaps) in 5 kb flank

717	number simple repeats in 10 kb downstream	1474	frequency of GCGTT in 5 kb flanking	2231	proportion of AT divergent sites relative to total num divergent sites (no gaps) in 5 kb flank
718	percentage simple repeats in 10 kb downstream	1475	frequency of GCTAA in 5 kb flanking	2232	proportion of CG divergent sites relative to total num divergent sites (no gaps) in 5 kb flank
719	number low complexity in 10 kb downstream	1476	frequency of GCTAC in 5 kb flanking	2233	proportion of CT divergent sites relative to total num divergent sites (no gaps) in 5 kb flank
720	percentage low complexity in 10 kb downstream	1477	frequency of GCTAG in 5 kb flanking	2234	proportion of GT divergent sites relative to total num divergent sites (no gaps) in 5 kb flank
721	total repeats (number SINEs + LINEs + LTR + DNA elements) in 10 kb downstream	1478	frequency of GCTAT in 5 kb flanking	2235	AC divergent sites per bp in 5 kb flank
722	GC content in ccds plus 2kb flanking	1479	frequency of GCTCA in 5 kb flanking	2236	AG divergent sites per bp in 5 kb flank
723	total % bases masked in ccds plus 2kb flanking	1480	frequency of GCTCC in 5 kb flanking	2237	AT divergent sites per bp in 5 kb flank
724	number SINEs in ccds plus 2kb flanking	1481	frequency of GCTCG in 5 kb flanking	2238	CG divergent sites per bp in 5 kb flank
725	percentage SINEs in ccds plus 2kb flanking	1482	frequency of GCTCT in 5 kb flanking	2239	CT divergent sites per bp in 5 kb flank
726	number ALUs in ccds plus 2kb flanking	1483	frequency of GCTGA in 5 kb flanking	2240	GT divergent sites per bp in 5 kb flank
727	percentage ALUs in ccds plus 2kb flanking	1484	frequency of GCTGC in 5 kb flanking	2241	proportion unalignable (gaps in human or other species) per bp in 5 kb flank
728	number MIRs in ccds plus 2kb flanking	1485	frequency of GCTGG in 5 kb flanking	2242	proportion unalignable (gaps in human) per bp in 5 kb flank
729	percentage MIRs in ccds plus 2kb flanking	1486	frequency of GCTGT in 5 kb flanking	2243	proportion unalignable (gaps in other species) per bp in 5 kb flank
730	number LINEs in ccds plus 2kb flanking	1487	frequency of GCTTA in 5 kb flanking	2244	proportion of divergent sites per bp in 10 kb flank
731	percentage LINEs in ccds plus 2kb flanking	1488	frequency of GCTTC in 5 kb flanking	2245	number of divergent sites in total in 10 kb flank
732	number LINE1 in ccds plus 2kb flanking	1489	frequency of GCTTG in 5 kb flanking	2246	proportion of divergent sites per bp, including gaps in 10 kb flank
733	percentage LINE1 in ccds plus 2kb flanking	1490	frequency of GCTTT in 5 kb flanking	2247	number of divergent sites in total, including gaps in 10 kb flank
734	number LINE2 in ccds plus 2kb flanking	1491	frequency of GGAAA in 5 kb flanking	2248	proportion of divergent transitions per bp in 10 kb flank
735	percentage LINE2 in ccds plus 2kb flanking	1492	frequency of GGAAC in 5 kb flanking	2249	number of divergent transitions in 10 kb flank
736	number L3/CR1 in ccds plus 2kb flanking	1493	frequency of GGAAG in 5 kb flanking	2250	proportion of divergent transversions per bp in 10 kb flank
737	percentage L3/CR1 in ccds plus 2kb flanking	1494	frequency of GGAAT in 5 kb flanking	2251	number of divergent transversions in 10 kb flank
738	number LTR in ccds plus 2kb flanking	1495	frequency of GGACA in 5 kb flanking	2252	proportion of divergent sites (no gaps) that are transitions in 10 kb flank
739	percentage LTR in ccds plus 2kb flanking	1496	frequency of GGACC in 5 kb flanking	2253	proportion of divergent sites that (no gaps) are transversions in 10 kb flank
740	number MaLRs in ccds plus 2kb flanking	1497	frequency of GGACG in 5 kb flanking	2254	proportion of total divergent sites that are gaps, including gaps in 10 kb flank
741	percentage MaLRs in ccds plus 2kb flanking	1498	frequency of GGACT in 5 kb flanking	2255	proportion of AC divergent sites relative to total num divergent sites (no gaps) in 10 kb flank
742	number ERVL in ccds plus 2kb flanking	1499	frequency of GGAGA in 5 kb flanking	2256	proportion of AG divergent sites relative to total num divergent sites (no gaps) in 10 kb flank

743	percentage ERVL in ccds plus 2kb flanking	1500	frequency of GGAGC in 5 kb flanking	2257	proportion of AT divergent sites relative to total num divergent sites (no gaps) in 10 kb flank
744	number ERV_classI in ccds plus 2kb flanking	1501	frequency of GGAGG in 5 kb flanking	2258	proportion of CG divergent sites relative to total num divergent sites (no gaps) in 10 kb flank
745	percentage ERV_classI in ccds plus 2kb flanking	1502	frequency of GGAGT in 5 kb flanking	2259	proportion of CT divergent sites relative to total num divergent sites (no gaps) in 10 kb flank
746	number ERV_classII in ccds plus 2kb flanking	1503	frequency of GGATA in 5 kb flanking	2260	proportion of GT divergent sites relative to total num divergent sites (no gaps) in 10 kb flank
747	percentage ERV_classII in ccds plus 2kb flanking	1504	frequency of GGATC in 5 kb flanking	2261	AC divergent sites per bp in 10 kb flank
748	number DNA elements in ccds plus 2kb flanking	1505	frequency of GGATG in 5 kb flanking	2262	AG divergent sites per bp in 10 kb flank
749	percentage DNA elements in ccds plus 2kb flanking	1506	frequency of GGATT in 5 kb flanking	2263	AT divergent sites per bp in 10 kb flank
750	number MER1_type in ccds plus 2kb flanking	1507	frequency of GGCAA in 5 kb flanking	2264	CG divergent sites per bp in 10 kb flank
751	percentage MER1_type in ccds plus 2kb flanking	1508	frequency of GGCAC in 5 kb flanking	2265	CT divergent sites per bp in 10 kb flank
752	number MER2_type in ccds plus 2kb flanking	1509	frequency of GGCAG in 5 kb flanking	2266	GT divergent sites per bp in 10 kb flank
753	percentage MER2_type in ccds plus 2kb flanking	1510	frequency of GGCAT in 5 kb flanking	2267	proportion unalignable (gaps in human or other species) per bp in 10 kb flank
754	number unclassified in ccds plus 2kb flanking	1511	frequency of GGCCA in 5 kb flanking	2268	proportion unalignable (gaps in human) per bp in 10 kb flank
755	percentage unclassified in ccds plus 2kb flanking	1512	frequency of GGCCC in 5 kb flanking	2269	proportion unalignable (gaps in other species) per bp in 10 kb flank
756	number small RNA in ccds plus 2kb flanking	1513	frequency of GGCCG in 5 kb flanking		
757	percentage small RNA in ccds plus 2kb flanking	1514	frequency of GGCCT in 5 kb flanking		

Table S3: 500 features most commonly retained in the 469 feature model. Features are ranked by the proportion of cases in which they were included in the 469 feature model, over 287 instances of leave one out cross-validation. Thus, features listed with a proportion of 1 were included in the final 469 feature model in every set of cross-validation feature eliminations.

Feature ID	Proportion	Feature ID	Proportion	Feature ID	Proportion
proportion of transitions in the CCDS relative to total number of SNPs	100	frequency of GGACC in 5 kb flanking	100.00	proportion of CG divergent sites relative to total num divergent sites (no gaps) in ccds	97.56
number of A-T SNPs in exons	100	frequency of GGAGG in 5 kb flanking	100.00	percentage unclassified in 5 kb downstream	97.21
number of C-T SNPs in exons	100	frequency of GGATG in 5 kb flanking	100.00	frequency of TATCT in 5 kb flanking	97.21
number of transversions in exons	100	frequency of GGCCC in 5 kb flanking	100.00	percentage unclassified in ccds plus 2kb flanking	96.86
proportion of C-G SNPs in exons relative to total num SNPs	100	frequency of GGGTT in 5 kb flanking	100.00	frequency of TAGGG in 5 kb flanking	96.86
proportion of G-T SNPs in introns relative to total num SNPs	100	frequency of GGTA A in 5 kb flanking	100.00	frequency of TCGAT in 5 kb flanking	96.86
proportion of C-T SNPs in 2 kb up relative to total num SNPs	100	frequency of GGTAG in 5 kb flanking	100.00	frequency of ATGTA in 5 kb flanking	96.52
C-G SNPs per bp in 2 kb up	100	frequency of GGTAT in 5 kb flanking	100.00	frequency of GCGAT in 5 kb flanking	96.52
proportion of A-C SNPs in 2 kb down relative to total num SNPs	100	frequency of GGTCG in 5 kb flanking	100.00	number of A-T SNPs in 2 kb up	96.17
transitions per bp in 2 kb down	100	frequency of GGTGT in 5 kb flanking	100.00	C-G SNPs per bp in ccds plus 10kb flanking	96.17
proportion of C-T SNPs in 5 kb up relative to total num SNPs	100	frequency of GTAAG in 5 kb flanking	100.00	frequency of AATGA in 5 kb flanking	96.17
number of A-G SNPs in 5 kb down	100	frequency of GTACA in 5 kb flanking	100.00	frequency of ATAAT in 5 kb flanking	96.17
proportion of transitions in 5 kb down relative to total num of SNPs	100	frequency of GTAGG in 5 kb flanking	100.00	frequency of GCTGA in 5 kb flanking	96.17
proportion of A-C SNPs in 5 kb down relative to total num SNPs	100	frequency of GTATA in 5 kb flanking	100.00	CT divergent sites per bp in ccds	96.17
A-G SNPs per bp in 5 kb down	100	frequency of GTCCG in 5 kb flanking	100.00	frequency of GTAGT in 5 kb flanking	95.82
total snps per bp in 5 kb down	100	frequency of GTCGG in 5 kb flanking	100.00	number ALUs in 10 kb upstream	95.47
number of A-G SNPs in 10 kb up	100	frequency of GTGAT in 5 kb flanking	100.00	frequency of ACGTT in 5 kb flanking	95.47
number of A-T SNPs in 10 kb up	100	frequency of GTGCA in 5 kb flanking	100.00	frequency of CCACC in 5 kb flanking	95.47
number of C-G SNPs in 10 kb up	100	frequency of GTGCG in 5 kb flanking	100.00	number of exons in the gene	95.47
number of G-T SNPs in 10 kb up	100	frequency of GTGTG in 5 kb flanking	100.00	A-T SNPs per bp in 2 kb up	95.12
number of total SNPS in 10 kb up	100	frequency of GTTCC in 5 kb flanking	100.00	frequency of CAAAT in 5 kb flanking	95.12
number of transitions in 10 kb up	100	frequency of TAACC in 5 kb flanking	100.00	proportion of CT divergent sites relative to total num divergent sites (no gaps) in 5 kb up	95.12
number of transversions in 10 kb up	100	frequency of TAATA in 5 kb flanking	100.00	transversions per bp in exons	94.77
proportion of transitions in 10 kb up relative to total num of SNPs	100	frequency of TACGA in 5 kb flanking	100.00	percentage SINEs in 2 kb downstream	94.43
proportion of A-C SNPs in 10 kb up relative to total num SNPs	100	frequency of TACTC in 5 kb flanking	100.00	number of C-T SNPs in 2 kb up	93.73
proportion of G-T SNPs in 10 kb up relative to total num SNPs	100	frequency of TAGCC in 5 kb flanking	100.00	C-T SNPs per bp in 2 kb up	93.73
A-G SNPs per bp in 10 kb up	100	frequency of TAGCG in 5 kb flanking	100.00	frequency of AAGAG in 5 kb flanking	93.73

A-T SNPs per bp in 10 kb up	100	frequency of TAGCT in 5 kb flanking	100.00	frequency of TGCTG in 5 kb flanking	93.73
C-G SNPs per bp in 10 kb up	100	frequency of TAGTG in 5 kb flanking	100.00	transitions per bp in the CCDS	93.38
G-T SNPs per bp in 10 kb up	100	frequency of TATGG in 5 kb flanking	100.00	percentage ERVL in 2kb upstream	93.38
transitions per bp in 10 kb up	100	frequency of TCAAG in 5 kb flanking	100.00	percentage satellites in 10 kb downstream	93.38
transversions per bp 10 kb up	100	frequency of TCACC in 5 kb flanking	100.00	proportion of AT divergent sites relative to total num divergent sites (no gaps) in 10 kb up	93.38
total snps per bp in 10 kb up	100	frequency of TCAGT in 5 kb flanking	100.00	AT divergent sites per bp in 2 kb flank	93.38
proportion of G-T SNPs in 10 kb down relative to total num SNPs	100	frequency of TCATT in 5 kb flanking	100.00	number LINE1 in ccds plus 5 kb flanking	93.03
proportion of A-C SNPs in ccds plus 10kb flanking relative to total num SNPs	100	frequency of TCCCC in 5 kb flanking	100.00	frequency of TCGTT in 5 kb flanking	93.03
A-T SNPs per bp in ccds plus 10kb flanking	100	frequency of TCGCC in 5 kb flanking	100.00	CG divergent sites per bp in introns	92.68
number of C-T SNPs in 2kb flanking	100	frequency of TCGTC in 5 kb flanking	100.00	frequency of CCTTC in 5 kb flanking	92.33
C-T SNPs per bp in 2kb flanking	100	frequency of TCTTG in 5 kb flanking	100.00	frequency of CGGAG in 5 kb flanking	92.33
number of A-G SNPs in 5kb flanking	100	frequency of TGAAA in 5 kb flanking	100.00	number L3/CR1 in 5 kb downstream	91.99
A-G SNPs per bp in 5kb flanking	100	frequency of TGATT in 5 kb flanking	100.00	frequency of AAGCT in 5 kb flanking	91.99
number of C-G SNPs in 10kb flanking	100	frequency of TGCCT in 5 kb flanking	100.00	frequency of AGGCT in 5 kb flanking	91.99
C-G SNPs per bp in 10kb flanking	100	frequency of TGGCG in 5 kb flanking	100.00	frequency of TCGGC in 5 kb flanking	91.99
proportion of transitions in introns plus 5 kb flanking relative to total num of SNPs	100	frequency of TGCTC in 5 kb flanking	100.00	frequency of AACTG in 5 kb flanking	91.64
proportion of G-T SNPs in introns plus 5 kb flanking relative to total num SNPs	100	frequency of TTAAA in 5 kb flanking	100.00	frequency of CCTGT in 5 kb flanking	91.64
number of synonymous snps	100	frequency of TTAAT in 5 kb flanking	100.00	frequency of GCGGT in 5 kb flanking	91.64
number ERV classI in ccds	100	frequency of TTATA in 5 kb flanking	100.00	frequency of GTAGA in 5 kb flanking	90.94
total repeats in ccds	100	frequency of TTATC in 5 kb flanking	100.00	frequency of TAATG in 5 kb flanking	90.94
percentage ALUs in 2kb upstream	100	frequency of TTATT in 5 kb flanking	100.00	percentage MER1_type in 2kb upstream	89.90
number MIRs in 2kb upstream	100	frequency of TTCGC in 5 kb flanking	100.00	frequency of TGACC in 5 kb flanking	89.20
percentage L3/CR1 in 2kb upstream	100	frequency of TTGCT in 5 kb flanking	100.00	frequency of TTTGT in 5 kb flanking	89.20
percentage low complexity in 2kb upstream	100	frequency of TTGGC in 5 kb flanking	100.00	number of transversions in introns	88.85
number SINEs in 2 kb downstream	100	frequency of TTGGG in 5 kb flanking	100.00	frequency of CTCCA in 5 kb flanking	88.85
number ALUs in 2 kb downstream	100	frequency of TTGGT in 5 kb flanking	100.00	proportion of AT divergent sites relative to total num divergent sites (no gaps) in 5 kb up	88.50
percentage MIRs in 2 kb downstream	100	frequency of TTGTA in 5 kb flanking	100.00	frequency of CGAGA in 5 kb flanking	88.15
percentage low complexity in 2 kb downstream	100	frequency of TTTCC in 5 kb flanking	100.00	frequency of CCTTG in 5 kb flanking	87.46
percentage LINE2 in 5 kb upstream	100	frequency of TTTCT in 5 kb flanking	100.00	frequency of TCGGA in 5 kb flanking	87.46
percentage LTR in 5 kb upstream	100	frequency of TTTGG in 5 kb flanking	100.00	frequency of AAGTT in 5 kb flanking	87.11

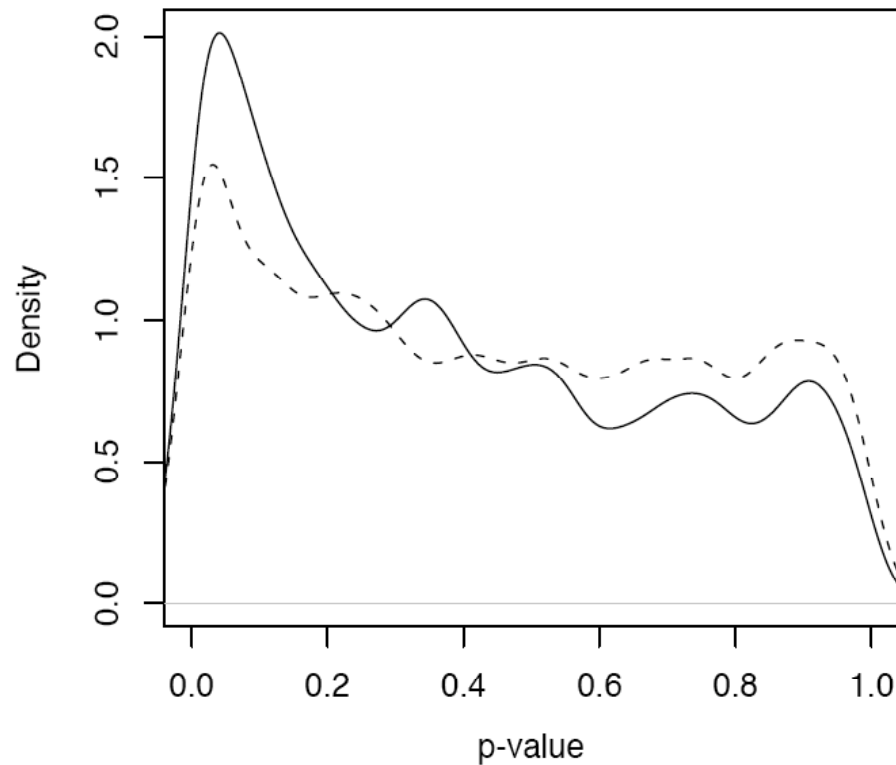


		flanking		flanking	
percentage simple repeats in 5 kb upstream	100	frequency of TTTTC in 5 kb flanking	100.00	frequency of CGGGT in 5 kb flanking	87.11
number MIRs in 5 kb downstream	100	frequency of TTTTG in 5 kb flanking	100.00	frequency of ACTTG in 5 kb flanking	86.41
percentage MIRs in 5 kb downstream	100	number of divergent transversions in exons	100.00	frequency of TTGTT in 5 kb flanking	86.41
percentage LINEs in 5 kb downstream	100	proportion of CG divergent sites relative to total num divergent sites (no gaps) in exons	100.00	frequency of TAAGC in 5 kb flanking	85.37
number SINEs in 10 kb upstream	100	CG divergent sites per bp in exons	100.00	frequency of TAGCA in 5 kb flanking	84.67
number LINEs in 10 kb upstream	100	proportion of AG divergent sites relative to total num divergent sites (no gaps) in introns	100.00	A-T SNPs per bp in introns plus 2 kb flanking	84.32
number LINE1 in 10 kb upstream	100	proportion of AT divergent sites relative to total num divergent sites (no gaps) in 2 kb up	100.00	frequency of ACACG in 5 kb flanking	84.32
percentage unclassified in 10 kb upstream	100	AT divergent sites per bp in 2 kb up	100.00	frequency of ACTCG in 5 kb flanking	84.32
percentage low complexity in 10 kb upstream	100	proportion of divergent sites that (no gaps) are transversions in 2 kb down	100.00	proportion of transitions in 2 kb up relative to total num of SNPs	83.97
percentage MIRs in 10 kb downstream	100	proportion of AG divergent sites relative to total num divergent sites (no gaps) in 2 kb down	100.00	proportion of GT divergent sites relative to total num divergent sites (no gaps) in ccds + 10 kb flank	83.97
number LINEs in 10 kb downstream	100	number of divergent transitions in 5 kb up	100.00	number MaLRs in 5 kb upstream	83.28
percentage LINEs in 10 kb downstream	100	proportion of divergent transversions per bp in 5 kb up	100.00	frequency of TCTCC in 5 kb flanking	82.93
percentage LTR in 10 kb downstream	100	proportion of GT divergent sites relative to total num divergent sites (no gaps) in 5 kb up	100.00	frequency of TATTC in 5 kb flanking	82.58
number MaLRs in 10 kb downstream	100	AT divergent sites per bp in 5 kb up	100.00	percentage unclassified in 5 kb upstream	82.23
percentage unclassified in 10 kb downstream	100	proportion of divergent sites that (no gaps) are transversions in 5 kb down	100.00	frequency of AAATG in 5 kb flanking	82.23
percentage unclassified in ccds plus 5 kb flanking	100	AT divergent sites per bp in 5 kb down	100.00	proportion unalignable (gaps in human or other species) per bp in introns	81.88
number MaLRs in ccds plus 10 kb flanking	100	proportion of CT divergent sites relative to total num divergent sites (no gaps) in 10 kb up	100.00	number L3/CR1 in ccds plus 10 kb flanking	79.44
total repeats in ccds plus 10 kb flanking	100	proportion of GT divergent sites relative to total num divergent sites (no gaps) in 10 kb up	100.00	frequency of CCTAC in 5 kb flanking	79.44
frequency of AAGAC in 5 kb flanking	100	proportion of divergent sites that (no gaps) are transversions in 10 kb down	100.00	frequency of TCAAC in 5 kb flanking	79.44
frequency of AAGCA in 5 kb flanking	100	AC divergent sites per bp in 10 kb down	100.00	number LINEs in ccds plus 2kb flanking	79.09
frequency of AAGCG in 5 kb flanking	100	proportion of divergent sites that (no gaps) are transversions in ccds + 10 kb flank	100.00	frequency of GTTGA in 5 kb flanking	78.40
frequency of AAGGA in 5 kb flanking	100	number of divergent transitions in 5 kb flank	100.00	number of G-T SNPs in 5 kb up	78.05
frequency of AATGG in 5 kb flanking	100	proportion of divergent transversions per bp in 5 kb flank	100.00	transversions per bp 5 kb up	78.05
frequency of AAAAA in 5 kb flanking	100	proportion of AT divergent sites relative to total num divergent sites (no gaps) in 5 kb flank	100.00	proportion of C-G SNPs in 5 kb down relative to total num SNPs	78.05

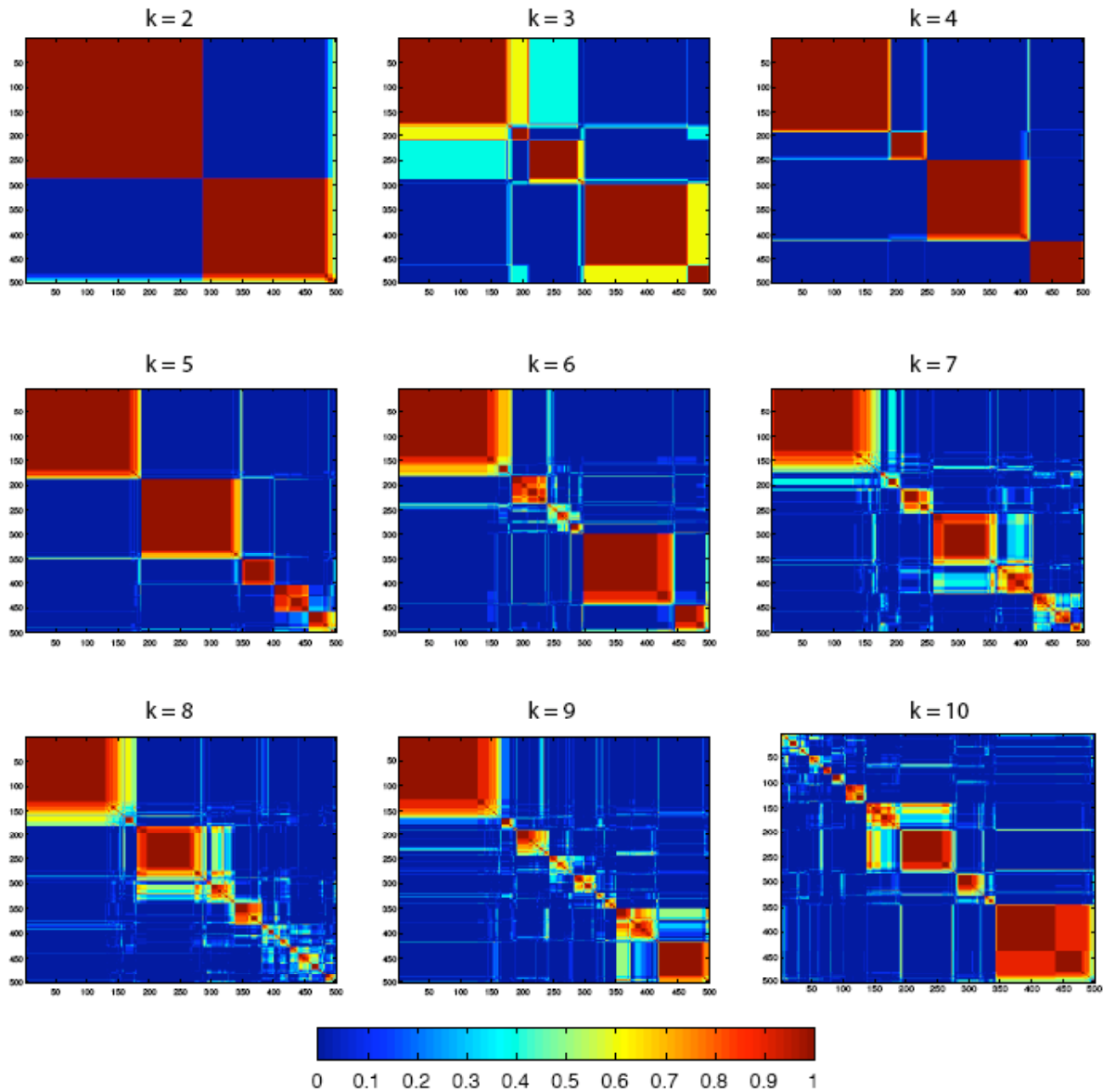
frequency of ACAAG in 5 kb flanking	100	proportion of CT divergent sites relative to total num divergent sites (no gaps) in 5 kb flank	100.00	A-T SNPs per bp in introns plus 5 kb flanking	77.70
frequency of ACATC in 5 kb flanking	100	AT divergent sites per bp in 5 kb flank	100.00	frequency of ATCTC in 5 kb flanking	77.00
frequency of ACCAC in 5 kb flanking	100	proportion of divergent sites that (no gaps) are transversions in 10 kb flank	100.00	frequency of CGGAT in 5 kb flanking	75.96
frequency of ACCCC in 5 kb flanking	100	AT divergent sites per bp in 10 kb flank	100.00	frequency of TTGAA in 5 kb flanking	75.96
frequency of ACGAC in 5 kb flanking	100	proportion unalignable (gaps in other species) per bp in 10 kb flank	100.00	frequency of CACTG in 5 kb flanking	75.26
frequency of ACGCC in 5 kb flanking	100	proportion of C-T SNPs in exons relative to total num SNPs	99.65	frequency of CGGTA in 5 kb flanking	75.26
frequency of ACGGG in 5 kb flanking	100	proportion of transitions in introns relative to total number of SNPs	99.65	frequency of AATTC in 5 kb flanking	74.91
frequency of ACGTG in 5 kb flanking	100	proportion of C-G SNPs in introns relative to total num SNPs	99.65	frequency of TCTTT in 5 kb flanking	74.22
frequency of ACTAA in 5 kb flanking	100	number of A-C SNPs in 2 kb down	99.65	percentage MER1_type in ccds plus 10 kb flanking	73.52
frequency of ACTAC in 5 kb flanking	100	A-C SNPs per bp in 2 kb down	99.65	frequency of AGGTC in 5 kb flanking	73.52
frequency of ACTAG in 5 kb flanking	100	proportion of G-T SNPs in introns plus 2 kb flanking relative to total num SNPs	99.65	frequency of TATCA in 5 kb flanking	73.17
frequency of AGAAT in 5 kb flanking	100	number of nonsynonymous snps	99.65	frequency of AGCAT in 5 kb flanking	72.13
frequency of AGACT in 5 kb flanking	100	percentage simple repeats in 2kb upstream	99.65	frequency of CAGCC in 5 kb flanking	71.78
frequency of AGAGC in 5 kb flanking	100	percentage LINE1 in 2 kb downstream	99.65	frequency of ACTCA in 5 kb flanking	71.43
frequency of AGATC in 5 kb flanking	100	number MaLRs in 2 kb downstream	99.65	transitions per bp in introns plus 2 kb flanking	70.73
frequency of AGCGT in 5 kb flanking	100	percentage L3/CR1 in 5 kb upstream	99.65	frequency of GAATG in 5 kb flanking	70.73
frequency of AGTAG in 5 kb flanking	100	percentage low complexity in 5 kb downstream	99.65	frequency of TACAC in 5 kb flanking	70.73
frequency of AGTCG in 5 kb flanking	100	total % bases masked in 10 kb upstream	99.65	percentage LINE1 in 10 kb downstream	69.69
frequency of ATAAC in 5 kb flanking	100	number ERV_classI in ccds plus 2kb flanking	99.65	number of A-T SNPs in 10kb flanking	69.34
frequency of ATAGG in 5 kb flanking	100	number L3/CR1 in ccds plus 5 kb flanking	99.65	frequency of TAGGC in 5 kb flanking	68.99
frequency of ATCAT in 5 kb flanking	100	frequency of AGCTC in 5 kb flanking	99.65	A-T SNPs per bp in 10kb flanking	68.64
frequency of ATCGT in 5 kb flanking	100	frequency of AGTGT in 5 kb flanking	99.65	frequency of CTATA in 5 kb flanking	68.29
frequency of ATGGT in 5 kb flanking	100	frequency of ATAGC in 5 kb flanking	99.65	CG divergent sites per bp in ccds	67.60
frequency of ATTCC in 5 kb flanking	100	frequency of ATCCG in 5 kb flanking	99.65	AT divergent sites per bp in 10 kb up	67.60
frequency of ATTGA in 5 kb flanking	100	frequency of CACTC in 5 kb flanking	99.65	frequency of TCATA in 5 kb flanking	65.51
frequency of ATTGC in 5 kb flanking	100	frequency of CCGAG in 5 kb flanking	99.65	percentage LINE1 in 5 kb downstream	65.16
frequency of CAACC in 5 kb flanking	100	frequency of CCGGC in 5 kb flanking	99.65	frequency of ATGCG in 5 kb flanking	64.81
frequency of CAACT in 5 kb flanking	100	frequency of CGAAA in 5 kb flanking	99.65	frequency of GCGTC in 5 kb flanking	64.81
frequency of CAAGT in 5 kb flanking	100	frequency of GAGAG in 5 kb flanking	99.65	A-G SNPs per bp in introns plus 2 kb flanking	64.46
frequency of CAATA in 5 kb flanking	100	frequency of GGACG in 5 kb flanking	99.65	frequency of ACCCG in 5 kb flanking	64.46

frequency of CAATC in 5 kb flanking	100	frequency of TAGGT in 5 kb flanking	99.65	frequency of TACCC in 5 kb flanking	63.76
frequency of CAATG in 5 kb flanking	100	proportion of CG divergent sites relative to total num divergent sites (no gaps) in 10 kb down	99.65	frequency of TTAGC in 5 kb flanking	62.72
frequency of CAATT in 5 kb flanking	100	proportion of AT divergent sites relative to total num divergent sites (no gaps) in 2 kb flank	99.65	frequency of CGCAC in 5 kb flanking	61.67
frequency of CACAG in 5 kb flanking	100	number of C-G SNPs in exons	99.30	AT divergent sites per bp in 10 kb down	61.32
frequency of CACGG in 5 kb flanking	100	number of C-G SNPs in 2 kb up	99.30	number LINE1 in ccds plus 2kb flanking	60.28
frequency of CACTT in 5 kb flanking	100	number of transitions in 2 kb down	99.30	frequency of TGTCC in 5 kb flanking	60.28
frequency of CAGTG in 5 kb flanking	100	number of transversions in 5 kb down	99.30	frequency of CTGTG in 5 kb flanking	59.93
frequency of CCAAG in 5 kb flanking	100	proportion of A-G SNPs in 10 kb up relative to total num SNPs	99.30	proportion of C-T SNPs in 10 kb down relative to total num SNPs	58.54
frequency of CCAGA in 5 kb flanking	100	C-G SNPs per bp in introns plus 5 kb flanking	99.30	percentage unclassified in ccds	58.19
frequency of CCATC in 5 kb flanking	100	number low complexity in 5 kb upstream	99.30	frequency of CTATT in 5 kb flanking	58.19
frequency of CCCGC in 5 kb flanking	100	frequency of GGCTT in 5 kb flanking	99.30	percentage LINEs in 2 kb downstream	57.49
frequency of CCGCT in 5 kb flanking	100	frequency of GTTAC in 5 kb flanking	99.30	frequency of ATTAT in 5 kb flanking	57.14
frequency of CCTGG in 5 kb flanking	100	proportion of GT divergent sites relative to total num divergent sites (no gaps) in ccds	99.30	frequency of CATT A in 5 kb flanking	57.14
frequency of CGAAG in 5 kb flanking	100	number of total SNPs in exons	98.95	frequency of ATCCT in 5 kb flanking	56.10
frequency of CGAGG in 5 kb flanking	100	A-G SNPs per bp in introns plus 5 kb flanking	98.95	frequency of TTTGC in 5 kb flanking	56.10
frequency of CGCCT in 5 kb flanking	100	transitions per bp in introns plus 5 kb flanking	98.95	number LINEs in ccds plus 5 kb flanking	54.70
frequency of CGGCG in 5 kb flanking	100	transversions per bp in introns plus 5 kb flanking	98.95	proportion of divergent sites per bp in 5 kb up	52.96
frequency of CGGTG in 5 kb flanking	100	total % bases masked in 2kb upstream	98.95	proportion unalignable (gaps in other species) per bp in 2 kb down	52.61
frequency of CGGTT in 5 kb flanking	100	percentage MaLRs in 2 kb downstream	98.95	percentage SINEs in 10 kb upstream	51.57
frequency of CGTGT in 5 kb flanking	100	percentage LTR in 5 kb downstream	98.95	frequency of ATGTG in 5 kb flanking	51.57
frequency of CGTTC in 5 kb flanking	100	frequency of CGCTG in 5 kb flanking	98.95	AG divergent sites per bp in introns	51.57
frequency of CTACA in 5 kb flanking	100	frequency of CGGGC in 5 kb flanking	98.95	percentage ERV_classII in 2kb upstream	50.87
frequency of CTACC in 5 kb flanking	100	frequency of GGCTC in 5 kb flanking	98.95	frequency of CGTCC in 5 kb flanking	50.87
frequency of CTACG in 5 kb flanking	100	frequency of GTCTG in 5 kb flanking	98.95	frequency of CTCTG in 5 kb flanking	50.87
frequency of CTAGA in 5 kb flanking	100	frequency of TCTCT in 5 kb flanking	98.95	frequency of TGGTA in 5 kb flanking	50.52
frequency of CTCGT in 5 kb flanking	100	number of G-T SNPs in exons	98.61	proportion of divergent sites that (no gaps) are transversions in ccds + 2kb flank	50.52
frequency of CTGCG in 5 kb flanking	100	number of A-G SNPs in 10kb flanking	98.61	frequency of CGAGC in 5 kb flanking	50.17
frequency of CTGTT in 5 kb flanking	100	proportion of A-G SNPs in introns plus 2 kb flanking relative to total num SNPs	98.61	frequency of GGGTG in 5 kb flanking	50.17
frequency of CTTAA in 5 kb flanking	100	A-C SNPs per bp in introns plus 2 kb flanking	98.61	frequency of TTGAC in 5 kb flanking	49.48

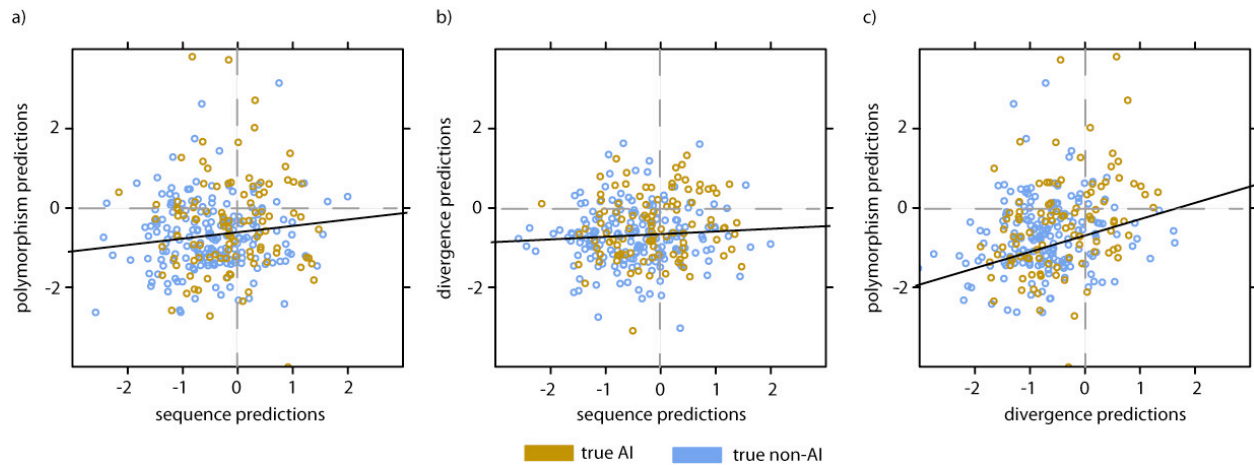
frequency of CTTAC in 5 kb flanking	100	G-T SNPs per bp in introns plus 5 kb flanking	98.61	GC content in ccds	48.08
frequency of CTTGA in 5 kb flanking	100	percentage unclassified in 2kb upstream	98.61	frequency of CTAGC in 5 kb flanking	48.08
frequency of CTTGC in 5 kb flanking	100	percentage LTR in 2 kb downstream	98.61	frequency of CAGGG in 5 kb flanking	47.39
frequency of CTTGT in 5 kb flanking	100	frequency of CAACG in 5 kb flanking	98.61	proportion of A-T SNPs in 5 kb up relative to total num SNPs	45.99
frequency of CTTTG in 5 kb flanking	100	frequency of CAGAT in 5 kb flanking	98.61	frequency of GGTC in 5 kb flanking	45.99
frequency of GAAGG in 5 kb flanking	100	proportion of C-G SNPs in 2 kb up relative to total num SNPs	98.26	frequency of GTACC in 5 kb flanking	45.99
frequency of GACCC in 5 kb flanking	100	A-G SNPs per bp in 10kb flanking	98.26	proportion unalignable (gaps in human or other species) per bp in ccds + 2kb flank	45.64
frequency of GACGG in 5 kb flanking	100	C-G SNPs per bp in introns plus 2 kb flanking	98.26	proportion of C-G SNPs in 2 kb down relative to total num SNPs	45.30
frequency of GACGT in 5 kb flanking	100	G-T SNPs per bp in introns plus 2 kb flanking	98.26	proportion of transitions in ccds plus 5kb flanking relative to total num of SNPs	44.60
frequency of GAGAC in 5 kb flanking	100	transversions per bp introns plus 2 kb flanking	98.26	proportion of AC divergent sites relative to total num divergent sites (no gaps) in ccds	44.25
frequency of GAGCT in 5 kb flanking	100	percentage ALUs in 5 kb downstream	98.26	total % bases masked in 2 kb downstream	43.21
frequency of GAGTA in 5 kb flanking	100	frequency of GTCCC in 5 kb flanking	98.26	frequency of GAACA in 5 kb flanking	42.86
frequency of GAGTT in 5 kb flanking	100	frequency of TACAG in 5 kb flanking	98.26	frequency of GGCGC in 5 kb flanking	42.86
frequency of GATAC in 5 kb flanking	100	percentage unclassified in ccds plus 10 kb flanking	97.91	total snps per bp in 2 kb up	42.16
frequency of GATCA in 5 kb flanking	100	frequency of ACCGG in 5 kb flanking	97.91	proportion of G-T SNPs in 2 kb down relative to total num SNPs	41.81
frequency of GATCC in 5 kb flanking	100	frequency of ACTCT in 5 kb flanking	97.91	frequency of CGTTT in 5 kb flanking	41.46
frequency of GATTC in 5 kb flanking	100	frequency of CGTCA in 5 kb flanking	97.91	frequency of CTCTT in 5 kb flanking	41.11
frequency of GCAAA in 5 kb flanking	100	frequency of GCCCC in 5 kb flanking	97.91	proportion of G-T SNPs in ccds plus 5kb flanking relative to total num SNPs	40.42
frequency of GCAAC in 5 kb flanking	100	frequency of TTCTC in 5 kb flanking	97.91	frequency of TCGAG in 5 kb flanking	40.42
frequency of GCAGG in 5 kb flanking	100	proportion of AG divergent sites relative to total num divergent sites (no gaps) in exons	97.91	percentage ERV_classII in 2 kb downstream	39.72
frequency of GCCTT in 5 kb flanking	100	proportion of total divergent sites that are gaps, including gaps in 2 kb down	97.91	frequency of CCGAC in 5 kb flanking	39.37
frequency of GCGAC in 5 kb flanking	100	proportion of transitions in 2 kb down relative to total num of SNPs	97.56	frequency of ATGAT in 5 kb flanking	38.33
frequency of GCGCC in 5 kb flanking	100	number simple repeats in 2 kb downstream	97.56	frequency of CTCGA in 5 kb flanking	37.63
frequency of GCGGG in 5 kb flanking	100	percentage LINE2 in 5 kb downstream	97.56	frequency of ACTGA in 5 kb flanking	35.19
frequency of GCGTA in 5 kb flanking	100	number LINE1 in 10 kb downstream	97.56		
frequency of GCTCC in 5 kb flanking	100	frequency of ACATT in 5 kb flanking	97.56		
frequency of GCTGC in 5 kb flanking	100	frequency of AGTGA in 5 kb flanking	97.56		
frequency of GCTTG in 5 kb flanking	100	frequency of GGATT in 5 kb flanking	97.56		



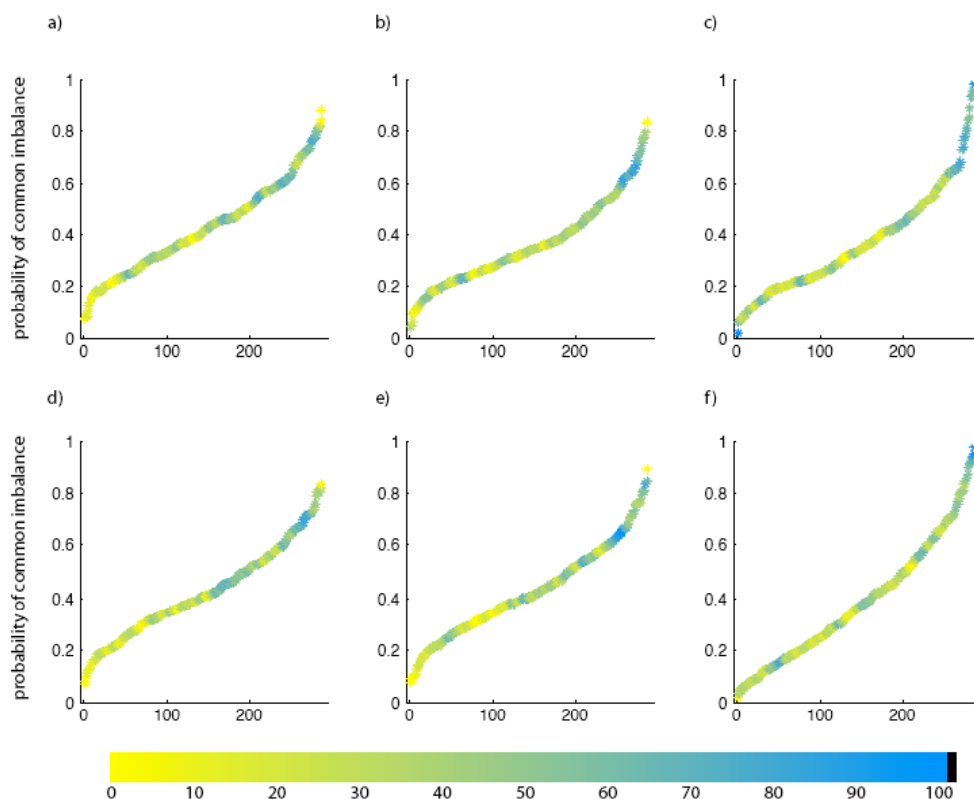
**Figure S1.** Five-mers from gene flanking regions differentiate between the allelic imbalance class and the non-allelic imbalance class better than five-mers from the coding sequence. Lines show the smoothed histograms of p-values from Wilcoxon summed ranks tests for all 1024 possible five-mers, where each test compares the values of a five-mer feature for the AI class and the non-AI class. The solid line represents five-mers from concatenated region 5 kb upstream of the start of translation and 5 kb downstream of the translation stop site; the dashed line represents five-mers obtained from the coding sequence (exons + introns). Because the p-values for the 5 kb flanking region were more skewed towards low (i.e., nominally significant) values, we used only these features when subsequently fitting the full model.



**Figure S2.** NMF consensus clustering matrices for varying number of clusters  $k$ . The matrix displays the frequency that the 500 features are found to be in the same cluster when the NMF algorithm is applied with different random initialization values.

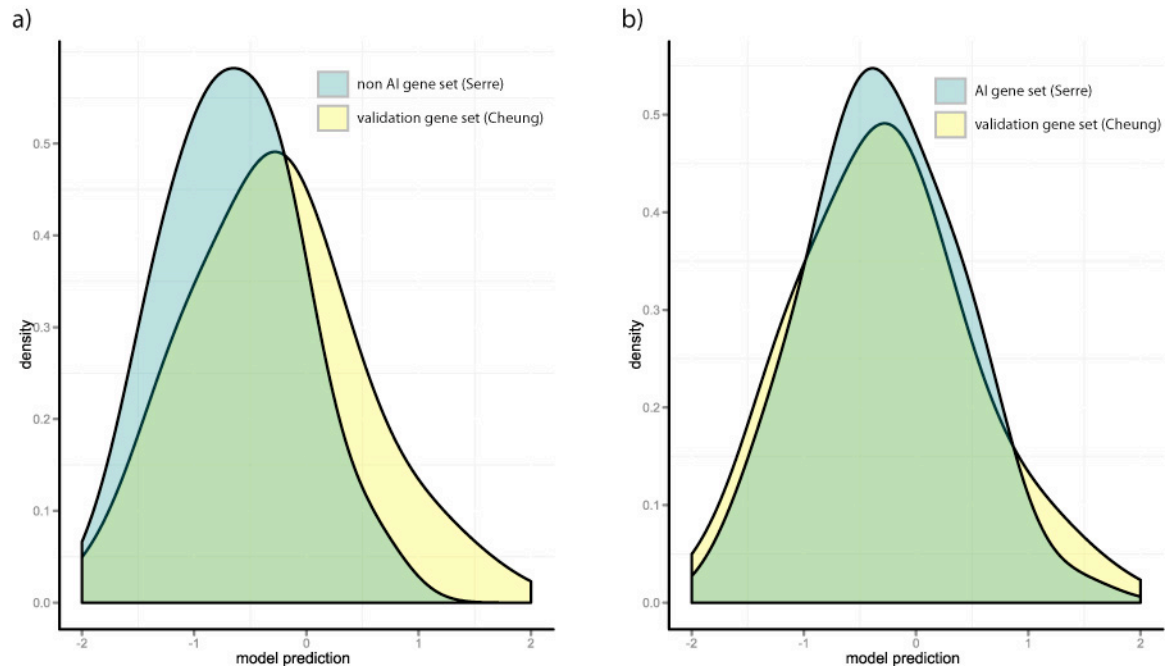


**Figure S3.** Correlation between predictions based on single feature subsets. Positive values on the x and y axes correspond to a prediction that the gene falls in the AI class; negative values correspond to a prediction that the gene falls in the non-AI class. Neither polymorphism- nor divergence-based predictions are correlated with sequence-based predictions: (a) Sequence predictions versus polymorphism predictions (Spearman's  $\rho = 0.051$ ,  $p = 0.389$ ); (b) Sequence predictions versus divergence predictions (Spearman's  $\rho = 0.085$ ,  $p = 0.150$ ). However, predictions derived from polymorphism features and divergence features are significantly positively correlated, suggesting that polymorphism data and divergence data contain partially redundant information about allelic imbalance: (c) Divergence predictions versus polymorphism predictions (Spearman's  $\rho = 0.250$ ,  $p = 1.93 \times 10^{-5}$ ).



**Figure S4.** Genes with more extreme predicted values are more likely to be predicted correctly. Predictions from different models were passed through a logit link function to produce a predicted probability of common imbalance at each gene. All 287 probability values are plotted, ranked from lowest probability of common imbalance to highest probability of common imbalance. True imbalance class is reflected by the color bar: yellow represents non-AI genes and blue represents AI genes. The color for each dot represents the degree to which model predictions were correct for a window size of eight genes around a given gene. Non-AI genes are predicted as commonly imbalanced with lower probability (lower left of each panel); AI genes are predicted as commonly imbalanced with higher probability (upper right of each panel). (a) Sequence features only; (b) Divergence features only; (c) Polymorphism features only; (d) Sequence + Divergence features; (e) Sequence + Polymorphism; (f) Polymorphism + Divergence.





**Figure S5.** Raw predictions from the full model for genes that exhibit allelic imbalance in the Cheung et al. (2008) dataset are significantly different from predictions for the non-AI gene set ( $p = 4.70 \times 10^{-6}$ ) derived from Serre et al. (2008), but not significantly different from predictions for the AI gene set from Serre et al. ( $p = 0.506$ ).

## Supplementary Methods:

### *Radial basis function kernel for the SVM*

We explored whether use of an alternative kernel function would improve the classification accuracy of the full model using the radial basis function kernel implemented in the program SVM<sup>light</sup> (the ‘-t 2’ option: Joachims 1999). The gamma parameter for the RBF kernel was estimated as the inverse of the median value of the sum of squared pairwise differences among all 287 genes in the sample. We explored settings of gamma over four orders of magnitude, centered on this estimate, and also over four orders of magnitude of the regularization parameter  $C$ . In no case did classification accuracy (as assessed by leave-one-out cross validation) improve over use of the default linear kernel. Hence, for all subsequent models and all results reported in the main text, we used a linear kernel.

### *Relationship between magnitude of imbalance and classification accuracy*

We analyzed the correlation between the mean allelic imbalance from the Serre *et al* (2008) dataset for genes in the allelic imbalance class and the predictions from the full model (for both  $C = 0.05$  and  $C = 1$ ). Where multiple values were given for the same gene (corresponding to different allelic imbalance marker SNPs within the gene), we always used the most extreme mean imbalance. All mean imbalance values were  $\log_2$  transformed prior to analysis.

We found some evidence for a relationship between the mean magnitude of imbalance and classification accuracy for predictions made with  $C = 0.05$  (Spearman’s  $\rho = 0.229$ ,  $p = 0.026$ ). Specifically, we identified a positive correlation between mean imbalance and the value of the prediction made by the full model when we used the continuous value of  $f(x)$  from the model (more positive values of  $f(x)$  correspond to greater confidence that a gene is truly commonly imbalanced; more negative values of  $f(x)$  correspond to greater confidence that a gene is truly not imbalanced). This result suggests that genes that exhibit more extreme mean imbalance may be easier to predict using our approach than genes that exhibit more modest levels of imbalance. However, it is difficult to disentangle whether means in this case truly represent more extreme values when imbalance occurs or, instead, imbalance that occurs more commonly in the population. When we conducted the same analysis with predictions made with  $C = 1$ , we found no relationship between mean imbalance and the value of the predictions (Spearman’s  $\rho = -0.056$ ,  $p = 0.586$ ). Hence, it remains unclear whether heterogeneity in the predictive ability of our model can generally be accounted for by differences in magnitude of mean imbalance.