

**Table S4.** Prediction of functional impact for somatic missense mutations.

GeneName	RefSeq ID	AccessionProt	AAchange	change_string	Prediction	GOSS_score	logR.E-value_score	pfam_domain_name	pfam_domain_start	pfam_domain_end	SIFT_score	SIFT_median_sequence_information
BCL2	NM_000657.2	NP_000648	S203N	GGWVGALGDV[S/N]LG	Not determined	12.86	NA	No domain altered	NA	NA	0	4.32
BCL2	NM_000657.2	NP_000648	*206R	GGWVGALGDVSLG[*R]	Not determined	13.86	NA	No domain altered	NA	NA	NA	NA
GLI2	NM_005270.3	NP_005261	G1083R	HASGALDEGT[G/R]QVYPTSTGTF	Not determined	16	NA	No domain altered	NA	NA	0.72	2.94
TLN2	NM_015059.1	NP_055874	T588M	CAITTISSNL[T/M]EMSKGVKLLA	Likely not-cancer	5.7	0.24	NAD_Gly3P_dh_N	576	697	0.06	3.09
GCLC	NM_001498.2	NP_001489	R423K	PPPPNSDIGW[R/K]VEFRPMEVQL	Likely not-cancer	1.96	1.74	GCS	236	610	0	2.81
GPR112	NM_153834.3	NP_722576	W71R	LVFMDDNSRY[W/R]MAFYSITNNA	Not determined	1.83	0.12	Pentaxin	31	220	0	4.32
SHANK2	NM_012309.1	NP_036441	R374Q	YSLDSEDLYS[R/Q]NAGPQANFRN	Likely not-cancer	2.38	-0.01	Extensin_2	121	487	0.28	2.87
TAIP-2	NM_024969.2	NP_079245	V12A	SGILKRFEE[V/A]DGSSPCCSVR	Not determined	6.24	NA	No domain altered	NA	NA	0.01	3.45
TIGD6	NM_030953.2	NP_112215	C307Y	RIQVGYLPSN[C/Y]TAVLQPLNLG	Likely not-cancer	9.84	1.57	DDE	170	372	0.12	2.61
GPR116	NM_015234.3	NP_056049	S494I	EGQNFSIKCI[S/I]DVSNDVYVW	Not determined	1.83	0.15	ig	485	547	0.01	4.32
C10orf120	NM_001010912.1	NP_001010912	Q181H	MRLARALGNH[Q/H]PLPYIERFTR	Not determined	NA	NA	No domain altered	NA	NA	0.3	3.48
HIST1H3G	NM_003534.2	NP_003525	A115T	FEDTNLCAIH[A/T]KRVIMPDKI	Not determined	13	0.4	Histone	58	132	0	4.19
EP300	NM_001429	NP_001420	V1148_F1149del	YKYCSKLSE[del_VF]EQEIDPVMQS	Not determined	23.82	NA	DUF902	1145	1205	NA	NA
KLHL6	NM_130446	NP_569713	K485_T486insK	GGPNGKLATDK[ins_K]TQCYPDSTNK	Not determined	1.4	NA	Kelch_1	457	503	NA	NA
CTSS	NM_004079.3	NP_004070	M185V	YGNKGCNGGF[W/V]TTAFQYIIDN	Likely not-cancer	-0.55	1.6	Peptidase_C1	115	330	0	2.86
RAFTLIN	NM_015150.1	NP_055965	V254M	GDGGELSPQG[V/M]SKTLDGPESN	Not determined	NA	NA	No domain altered	NA	NA	0.15	3.34
ATP6V1B2	NM_001693.3	NP_001684	R400Q	PPINVLPSLS[R/Q]LMKSAIGEGM	Not determined	-1.87	NA	No domain altered	NA	NA	0	2.85
HMCN1	NM_031935.1	NP_114141	T5167M	GQDCDN[T/M]IGSYRCVVRGSG	Not determined	6.38	NA	NA	NA	NA	NA	NA
S1151T	NM_016341.2	NP_001159451	S1151T	PNPLPSRRAH[S/T]LTTAGSPNLA	Not determined	6.48	-0.15	MCPVI	1040	1206	0.04	3.45
FGF23	NM_020638.2	NP_065689	A12T	LGARLRLWVC[A/T]LCSVCSMSVL	Not determined	1.24	NA	No domain altered	NA	NA	0.08	2.86
ARID1A	NM_139135.1	NP_624361	R1276*	PGLGNVAMGP[R/*]QHYPYGGPYD	Not determined	16.28	NA	No domain altered	NA	NA	NA	NA

**Analysis software:** CanPredict (<http://www.cgl.ucsf.edu/Research/genentech/canpredict/index.html>)  
**Cancer Res. 2007 Jan 15;67(2):465-73**

**SIFT Score:**

Multiple alignment information to predict tolerated and deleterious substitutions  
 Substitutions with a SIFT Score (normalized probability) of less than 0.05 are predicted to be deleterious by SIFT

**SIFT Median Information Content:**

Determined by the SIFT algorithm. Calculates the information at each position in the alignment and then obtaining the median over all positions  
 Only scores with a median sequence information content < 3.25 were used in this analysis

**Pfam Domain Affected:**

Indicates the name of the domain altered by a particular mutation

**Wildtype Expect/Mutant Expect:**

Provided by the HMMER 2.3.2 software. The Wildtype Expect is the E-value obtained when searching the Pfam database using the wildtype version of the provided protein. The Mutant Expect is the E-value obtained when searching the Pfam database using the mutated version of the provided protein. These two values are used to calculate the LogR.E-value score.

**LogR.E-value score:**

Calculated as  $\text{LogR.E-value} = \log_{10}(\text{E-value variant} / \text{E-value wildtype})$ . LogR.E-value scores > 0.5 are likely to indicate deleterious changes

**GOSS score:**

The Gene Ontology Similarity Score (GOSS) provides an indication of how similar a particular gene is to known cancer genes  
 GOSS score ~ 14.04 for the cancer genes and ~3.23 for the non-cancer genes