

Supplemental Table 1. Exonic SPMs in protein-coding and untranslated genes.

Chr	Gene	Gene Function	Position	Mutation	Samples with Mutation
1	OMA1	protein coding	58946658	A>G	T1
1	NBPF1	protein coding	16902784	G>T	T2, T3, M2
1	NBPF14	protein coding	148010896	C>G	all
1	NBPF14	protein coding	148010901	A>C	all
1	HRNR	protein coding	152192327	T>A	T3, M1
1	NBPF1	protein coding	16902780	C>A	T2, T3, M2
1	NASP	protein coding	46073437	C>T	M1
1	IGSF3	protein coding	117158745	A>C	T1, T2, M1, M2
1	IGSF3	protein coding	117158748	A>G	all
2	OTOF	protein coding	26700372	C>T	all
2	C2orf73	protein coding	54587553	G>A	all
2	ADD2	protein coding	70890768	G>A	T1
2	PABPC1P2	untranslated	147345234	C>T	all
2	CHRNA1	protein coding	175618430	G>A	all
2	LOC151171	untranslated	239460128	C>T	T1, T2
2	KIF1A	protein coding	241727600	G>A	all
2	IFT172	protein coding	27667361	C>T	T3
3	ITIH4	protein coding	52857663	C>T	all
3	SUCLG2	protein coding	67659991	C>T	M1, M2
3	C3orf77	protein coding	44284557	G>A	all
3	OR5H14	protein coding	97868487	A>T	M1, M2
3	MUC4	protein coding	195515194	G>C	T1, T2
4	SH3BP2	protein coding	2833374	G>T	all
4	FAT4	protein coding	126371835	G>A	T3
4	SH3BP2	protein coding	2833373	T>A	all
4	OTOP1;OTO P1	protein coding	4198894	G>A	T1
4	OTOP1	protein coding	4198897	A>G	T1
4	OTOP1	protein coding	4204194	A>G	T2, M2
4	KCNIP4	protein coding	20852259	G>C	all
5	AMACR	protein coding	33989382	T>A	all
5	DDX46	protein coding	134146945	C>T	M1, M2
5	SLC45A2	protein coding	33944813	C>T	T2, T3, M1, M2
5	KIAA0825	protein coding	93752964	G>A	all
5	PDGFRB	protein coding	149512490	C>T	all
5	DOCK2	protein coding	169188532	G>C	all
6	MUC21	protein coding	30954459	G>C	T3, M1
7	TMEM140	protein coding	134849525	G>A	all
7	PAPOLB	protein coding	4900509	G>C	T3
7	FZD9	protein coding	72849199	G>A	all
7	PRSS2	protein coding	142481353	G>A	M1, M2
7	DPP6;DPP6	protein coding	154172024	C>T	all
8	RPL23AP53	untranslated	162614	A>G	M1, M2
8	FLJ10661	untranslated	8090323	G>A	M1
9	LOC642236	untranslated	68438597	C>T	all
9	TUBBP5	untranslated	141069199	A>G	T1, T3, M1, M2

9	PRSS3	protein coding	33798052	C>T	T3, M1, M2
10	PCDH15	protein coding	55587209	C>T	all
10	FLJ46300	untranslated	133605304	G>A	all
11	INPPL1	protein coding	71943698	C>T	T2
11	SF3B2	protein coding	65827308	A>T	all
11	ZNF202	protein coding	123596948	G>C	T2
11	LOC143666	untranslated	574178	A>G	all
11	OR51L1	protein coding	5021024	C>T	T3
11	TECTA	protein coding	121060548	G>A	T2, T3
12	WDR66	protein coding	122361695	T>G	all
13	ANKRD20A9 P	untranslated	19436166	C>G	T1, T2, M2
13	ANKRD20A9 P	untranslated	19437235	A>T	M1, M2
14	TRAF3	protein coding	103336726	C>T	all
15	HERC2	protein coding	28517429	G>A	T1, M2
15	LOC646214	untranslated	21932851	A>G	T2, M1
15	LOC503519	untranslated	26378119	G>A	all
15	LOC1005068 74	untranslated	85122517	A>T	T1, T2, M2
16	PMFBP1	protein coding	72153820	C>A	M1, M2
17	DHX8	protein coding	41598878	C>T	all
17	MYH10	protein coding	8434001	T>A	M1, M2
17	KRTAP4-11	protein coding	39274291	T>C	T3, M1, M2
18	ANKRD20A5 P	untranslated	14226328	C>G	M1, M2
18	LOC339298	untranslated	73835003	C>G	M1, M2
19	ZNF776	protein coding	58264776	G>A	all
19	MAP2K7	protein coding	7976169	C>T	T2, T3
19	HAMP	protein coding	35775900	C>T	M1, M2
19	ZNF14	protein coding	19823092	C>T	all
19	ZNF599	protein coding	35250883	G>A	all
20	JAG1	protein coding	10627665	C>T	all
20	MYLK2	protein coding	30419610	C>T	all
20	SNORA71D	untranslated	37062565	G>A	all
21	DSCAM-AS1	untranslated	41755879	G>A	M1, M2
22	SULT4A1	protein coding	44221961	C>T	all

Supplemental Table 2. List of 55 cancer genes containing validated somatic mutations. The sample (T1, T2, T3, M1, M2) columns show whether the sample contains validated somatic mutations in the gene (1) or not (0).

Gene Symbol	chr	Gene Start	Gene End	Sample				
				T1	T2	T3	M1	M2
PBX1	1	164523821	164855300	0	0	0	0	1
ABL2	1	179067462	179199819	0	0	0	0	1
SETD2	3	47056900	47206467	0	0	0	1	0
TRIM24	7	138143953	138275738	0	0	0	1	1
MYST3	8	41785997	41910508	0	0	1	0	0
EBF1	5	158121924	158527788	0	0	1	1	1
PRDM16	1	2984732	3356185	0	1	0	0	0
MAML2	11	95708762	96077382	0	1	0	0	0
CCND3	6	41901671	42019095	0	1	1	1	1
CDK6	7	92233235	92466908	0	1	1	1	1
KIAA1549	7	138515126	138667064	0	1	1	1	1
SEPT9 (MSF)	17	75276492	75497674	0	1	1	1	1
ZBTB16	11	113929315	114122398	1	0	0	0	0
TPM4	19	16177317	16214813	1	0	0	1	0
BCL10	1	85730931	85744771	1	0	1	0	0
SRGAP3	3	9021275	9405737	1	0	1	1	1
CDH11	16	64976656	65157101	1	1	0	0	1
NSD1	5	176559026	176728216	1	1	0	1	0
NFE2L2	2	178091323	178258425	1	1	1	0	1
FIP1L1	4	54242810	55162439	1	1	1	0	1
BRD3	9	136896963	136934657	1	1	1	0	1
NOTCH2	1	120453176	120613240	1	1	1	1	1
PDE4DIP	1	144835157	145077186	1	1	1	1	1
NTRK1	1	156784432	156852642	1	1	1	1	1
SDHC	1	161283047	161333984	1	1	1	1	1
ALK	2	29414640	30145432	1	1	1	1	1
EML4	2	42395490	42560688	1	1	1	1	1
MSH2	2	47629108	47790450	1	1	1	1	1
IGK@	2	89155874	89631436	1	1	1	1	1
AFF3	2	100162718	100760201	1	1	1	1	1
FHIT	3	59734036	61238133	1	1	1	1	1
FOXP1	3	71002844	71634140	1	1	1	1	1
LPP	3	187870072	188609460	1	1	1	1	1
AFF4	5	132210071	132300326	1	1	1	1	1
PDGFRB	5	149492400	149536423	1	1	1	1	1
EGFR	7	55085714	55325313	1	1	1	1	1
EZH2	7	148503475	148582413	1	1	1	1	1
MLL3	7	151831010	152134090	1	1	1	1	1

RUNX1T1	8	92966203	93116514	1	1	1	1	1
CD274	9	5449503	5471566	1	1	1	1	1
NFIB	9	14080842	14399982	1	1	1	1	1
ABL1	9	133588268	133764062	1	1	1	1	1
ZMYM2	13	20531810	20666968	1	1	1	1	1
KTN1	14	56045925	56152301	1	1	1	1	1
IGH@	14	106031614	107289051	1	1	1	1	1
CDH1	16	68770128	68870444	1	1	1	1	1
GAS7	17	9812926	10102868	1	1	1	1	1
NF1	17	29420945	29709905	1	1	1	1	1
COL1A1	17	48259650	48279993	1	1	1	1	1
MSI2	17	55332212	55758658	1	1	1	1	1
ALO17	17	78233667	78371086	1	1	1	1	1
BCL2	18	60789579	60988361	1	1	1	1	1
RUNX1	21	36159098	37358047	1	1	1	1	1
TMPRSS2	21	42835478	42904043	1	1	1	1	1
IGL@	22	22379474	23266085	1	1	1	1	1

References

- 1 Ewing, B. and Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res*, **8**, 186-194.
- 2 Wang, K., Li, M. and Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*, **38**, e164.
- 3 Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. and McVean, G.A. An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56-65.
- 4 Yachida, S., Jones, S., Bozic, I., Antal, T., Leary, R., Fu, B., Kamiyama, M., Hruban, R.H., Eshleman, J.R., Nowak, M.A. *et al.* Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature*, **467**, 1114-1117.

Supplement Table 3. List of validated somatic mutations within 1000bp of cancer genes. The sample (T1, T2, T3, M1, M2) columns show whether the sample contains the mutation (1) or not (0).

gene symbol	mutation type	position	T1	T2	T3	M1	M2
PRDM16	indel	chr1:3319602-3319632 CCCTCCTCTGAGTCT	0	1	0	0	0
BCL10	indel	chr1:85732582-85732582 G>-	1	0	1	0	0
NOTCH2	SPM	chr1:120514030-120514030 G>A	1	1	1	1	1
NOTCH2	indel	chr1:120613197-120613199 AAC>-	1	0	1	1	1
PDE4DIP	SPM	chr1:145052709-145052709 A>G	0	1	0	1	0
PDE4DIP	SPM	chr1:144842657-144842657 T>A	0	0	1	1	0
PDE4DIP	SPM	chr1:144840798-144840798 A>G	0	0	0	0	0
PDE4DIP	SPM	chr1:145027068-145027068 A>T	0	0	0	0	0
PDE4DIP	indel	chr1:144841031-144841033 CTT>-	1	0	1	1	1
PDE4DIP	indel	chr1:144853557-144853562 AAGTAA>-	0	0	0	0	0
PDE4DIP	indel	chr1:144991398-144991423 ACACACACACA	1	0	1	0	0
PDE4DIP	indel	chr1:145017838-145017847 TTCCTT>-	0	0	0	0	0
PDE4DIP	indel	chr1:145024577-145024590 ACACACACACA	0	0	0	0	0
PDE4DIP	indel	chr1:145048154-145048154 G>-	1	1	1	1	1
PDE4DIP	indel	chr1:145048529-145048529 C>-	1	1	1	1	1
PDE4DIP	indel	chr1:145048529-145048530 ->CC	1	0	1	1	1
PDE4DIP	indel	chr1:145054413-145054413 C>-	1	1	1	1	1
PDE4DIP	indel	chr1:145070603-145070603 A>-	1	1	1	1	1
PDE4DIP	indel	chr1:145070627-145070627 A>-	1	1	0	1	1
PDE4DIP	indel	chr1:145073546-145073547 AG>-	1	1	1	1	1
NTRK1	SPM	chr1:156842946-156842946 C>T	1	1	1	1	1
SDHC	SPM	chr1:161294185-161294185 G>A	1	1	1	1	1
PBX1	indel	chr1:164721444-164721449 TGTGTG>-	0	0	0	0	1
ABL2	SPM	chr1:179146550-179146550 C>T	0	0	0	0	1
ABL2	SPM	chr1:179089304-179089304 G>C	0	0	0	0	0
ALK	SPM	chr2:29465287-29465287 T>C	0	0	0	1	1
ALK	SPM	chr2:29978187-29978187 C>T	1	1	1	1	1
EML4	indel	chr2:42547465-42547465 T>-	1	1	1	1	1
MSH2	indel	chr2:47729626-47729629 ATGG>-	1	1	1	1	1
IGK@	SPM	chr2:89295050-89295050 C>G	0	1	1	1	1
IGK@	SPM	chr2:89297733-89297733 C>T	0	0	0	0	0
IGK@	indel	chr2:89249676-89249676 T>-	1	1	1	0	1
IGK@	indel	chr2:89249813-89249813 G>-	1	0	0	1	1
IGK@	indel	chr2:89298152-89298156 ACACC>-	0	0	1	1	0
AFF3	SPM	chr2:100461046-100461046 T>C	1	1	1	1	1
AFF3	indel	chr2:100314438-100314441 TGTT>-	1	1	1	1	1
NFE2L2	SPM	chr2:178130721-178130721 G>T	1	1	1	0	1
SRGAP3	SPM	chr3:9115691-9115691 T>C	0	0	0	1	1
SRGAP3	indel	chr3:9272552-9272561 AAAAGAAAAG>-	1	0	1	1	1
SETD2	SPM	chr3:47065973-47065973 T>C	0	0	0	1	0
FHIT	SPM	chr3:60291398-60291398 G>T	1	1	1	1	1
FHIT	SPM	chr3:60015625-60015625 G>C	0	0	1	1	0
FHIT	SPM	chr3:60811887-60811887 G>A	0	0	0	0	0

FHIT	SPM	chr3:60289463-60289463 T>C	1	1	1	1	1
FHIT	indel	chr3:60520017-60520172 TTAGCCGGGCGC	0	0	0	0	0
FHIT	indel	chr3:60545482-60545485 GCAT>-	1	1	1	1	1
FHIT	indel	chr3:60274296-60274303 GTGTGTGT>-	0	1	0	0	0
FOXP1	indel	chr3:71362686-71362686 T>-	1	1	1	1	1
LPP	SPM	chr3:188267504-188267504 C>T	1	1	1	1	1
FIP1L1	indel	chr4:54969762-54969763 GC>-	1	1	1	0	1
AFF4	SPM	chr5:132235649-132235649 C>A	1	1	1	1	1
PDGFRB	SPM	chr5:149512490-149512490 C>T	1	1	1	1	1
EBF1	SPM	chr5:158180399-158180399 C>T	0	0	0	1	1
EBF1	indel	chr5:158249272-158249275 GGAC>-	0	0	1	0	1
NSD1	indel	chr5:176679796-176679796 C>-	1	1	0	1	0
CCND3	SPM	chr6:41935228-41935228 T>C	0	1	1	1	1
EGFR	indel	chr7:55159541-55159544 ACAC>-	1	1	1	1	1
CDK6	SPM	chr7:92306707-92306707 G>A	0	1	1	1	1
TRIM24	SPM	chr7:138167346-138167346 G>A	0	0	0	1	1
KIAA1549	SPM	chr7:138610490-138610490 C>A	0	1	1	1	1
EZH2	indel	chr7:148520745-148520745 T>-	1	1	1	1	1
MLL3	SPM	chr7:152078131-152078131 G>A	0	0	0	0	0
MLL3	SPM	chr7:151928407-151928407 C>T	0	0	1	1	1
MLL3	SPM	chr7:152079940-152079940 G>T	1	0	0	1	0
MLL3	SPM	chr7:151980383-151980383 A>G	0	0	0	0	0
MLL3	SPM	chr7:151940427-151940427 G>A	0	0	0	1	0
MLL3	SPM	chr7:151983556-151983556 A>G	0	0	0	0	0
MLL3	SPM	chr7:152079794-152079794 G>T	0	0	0	1	0
MLL3	indel	chr7:151958091-151958094 GTAG>-	1	1	1	0	1
MLL3	indel	chr7:151972350-151972350 C>-	1	1	1	1	1
MLL3	indel	chr7:151976119-151976119 A>-	1	1	1	1	1
MLL3	indel	chr7:151978829-151978836 AAAAAAAAA>-	0	0	0	0	0
MLL3	indel	chr7:151979215-151979218 AAAA>-	1	1	1	1	1
MLL3	indel	chr7:151979439-151979440 AA>-	1	1	1	1	1
MLL3	indel	chr7:151981003-151981007 AACTC>-	1	1	1	1	1
MLL3	indel	chr7:151986595-151986595 T>-	1	1	1	1	1
MLL3	indel	chr7:152100795-152100818 ACACATACACA	1	0	0	0	0
MLL3	indel	chr7:152103531-152103534 AAA>-	1	1	1	1	1
MLL3	indel	chr7:152104757-152104757 A>-	1	1	1	1	1
MLL3	indel	chr7:152104782-152104784 AAA>-	1	1	1	1	1
MLL3	indel	chr7:152106492-152106495 AAGA>-	1	1	1	1	1
MLL3	indel	chr7:152109486-152109486 A>-	1	1	1	1	1
MLL3	indel	chr7:152111933-152111935 TTC>-	1	1	1	1	1
MYST3	indel	chr8:41804010-41804019 ATACACACAC>-	0	0	1	0	0
RUNX1T1	SPM	chr8:92995116-92995116 A>T	1	1	1	1	1
CD274	SPM	chr9:5468129-5468129 A>T	1	1	1	1	1
NFIB	SPM	chr9:14313235-14313235 G>A	1	1	1	1	1
ABL1	SPM	chr9:133617582-133617582 C>T	1	1	1	1	1
ABL1	indel	chr9:133684803-133684806 GAGT>-	1	1	1	1	1
BRD3	indel	chr9:136907769-136907770 AC>-	1	1	1	0	1

MAML2	SPM	chr11:95771033-95771033 A>G	0	1	0	0	0
ZBTB16	SPM	chr11:114037793-114037793 G>T	1	0	0	0	0
ZMYM2	SPM	chr13:20595524-20595524 G>A	0	1	1	1	1
ZMYM2	indel	chr13:20629136-20629136 G>-	1	1	1	1	1
KTN1	indel	chr14:56113175-56113176 TC>-	1	1	1	1	1
IGH@	SPM	chr14:106782048-106782048 T>A	1	0	1	0	1
IGH@	SPM	chr14:106172904-106172904 C>G	0	0	0	0	0
IGH@	SPM	chr14:106440767-106440767 A>G	0	0	0	0	0
IGH@	SPM	chr14:106312723-106312723 C>T	0	1	1	1	1
IGH@	SPM	chr14:106782132-106782132 C>G	0	0	0	1	0
IGH@	indel	chr14:106155518-106155554 CACACACACA	1	0	0	1	0
IGH@	indel	chr14:106440795-106440890 TCTGTCCCCT	0	0	0	0	0
IGH@	indel	chr14:106782023-106782023 C>-	1	1	1	1	1
IGH@	indel	chr14:106902128-106902128 G>-	0	1	0	1	1
IGH@	indel	chr14:107024556-107024557 GG>-	1	0	0	0	1
IGH@	indel	chr14:107148284-107148322 AGCATGGCC	0	1	0	0	1
IGH@	indel	chr14:107161695-107161695 G>-	0	0	0	1	0
IGH@	indel	chr14:107261784-107261787 TCAG>-	1	1	1	1	1
IGH@	indel	chr14:106372443-106372444 AG>-	1	1	1	1	1
CDH11	SPM	chr16:65096284-65096284 G>C	1	1	0	0	1
CDH1	SPM	chr16:68849960-68849960 G>A	1	1	1	1	1
GAS7	SPM	chr17:9862901-9862901 G>A	1	1	1	1	1
GAS7	indel	chr17:9908456-9908479 GCGCACGCATGCA	1	1	1	0	1
NF1	SPM	chr17:29653404-29653404 G>A	1	1	1	1	1
COL1A1	SPM	chr17:48265653-48265653 C>T	1	1	1	1	1
MSI2	indel	chr17:55660849-55660849 C>-	1	1	1	1	1
SEPT9	indel	chr17:75480513-75480521 TGGTGATGG>-	0	1	1	1	1
SEPT9	indel	chr17:75480585-75480590 TGGTGA>-	0	0	0	0	0
ALO17	SPM	chr17:78365227-78365227 C>T	1	1	1	1	1
BCL2	SPM	chr18:60829737-60829737 G>T	1	1	1	1	1
TPM4	indel	chr19:16180119-16180120 GT>-	1	0	0	1	0
TPM4	indel	chr19:16180137-16180138 GT>-	1	0	0	1	0
RUNX1	SPM	chr21:36478701-36478701 C>A	0	0	0	1	1
RUNX1	SPM	chr21:36271357-36271357 C>T	1	1	1	1	1
RUNX1	indel	chr21:36435846-36435857 AAACACACACA	0	0	0	0	0
RUNX1	indel	chr21:36809539-36809541 CTT>-	1	1	1	1	1
TMPRSS2	indel	chr21:42897867-42897867 C>-	1	1	1	1	1
IGL@	SPM	chr22:22656007-22656007 T>C	0	0	0	1	0
IGL@	SPM	chr22:22997406-22997406 A>G	0	0	1	0	0
IGL@	SPM	chr22:22644394-22644394 G>A	0	0	0	1	0
IGL@	SPM	chr22:22646300-22646300 G>A	0	0	0	0	0
IGL@	SPM	chr22:22644350-22644350 C>T	1	0	1	1	1
IGL@	SPM	chr22:22616403-22616403 G>A	0	0	0	0	0
IGL@	indel	chr22:22537437-22537472 TAGCTAGCTAG	0	0	0	0	0
IGL@	indel	chr22:22641073-22641105 TCATCATTTCAT	0	0	0	0	0
IGL@	indel	chr22:22704660-22704663 GCAA>-	1	1	1	1	1
IGL@	indel	chr22:22704995-22704995 C>-	1	1	1	1	1