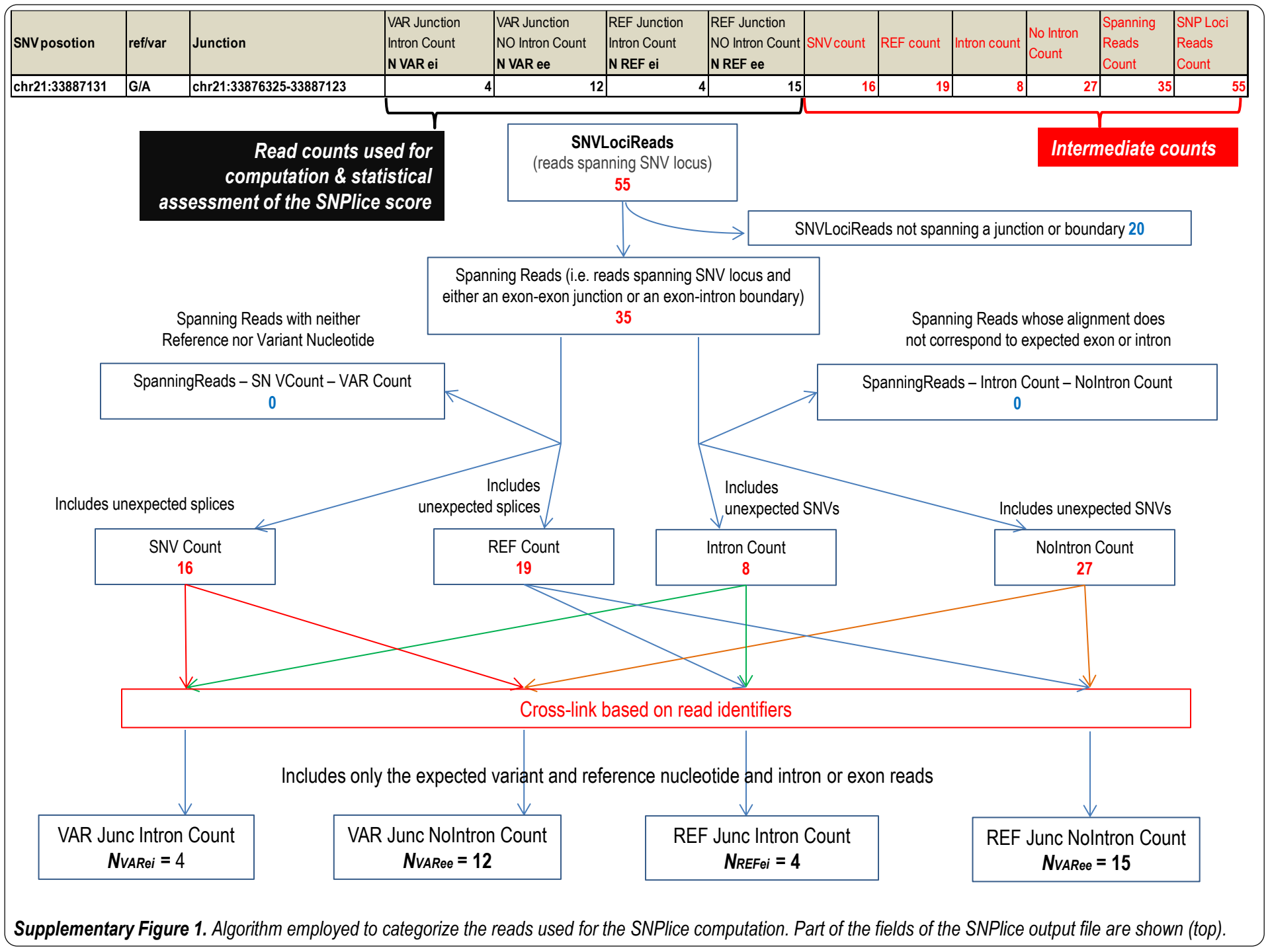
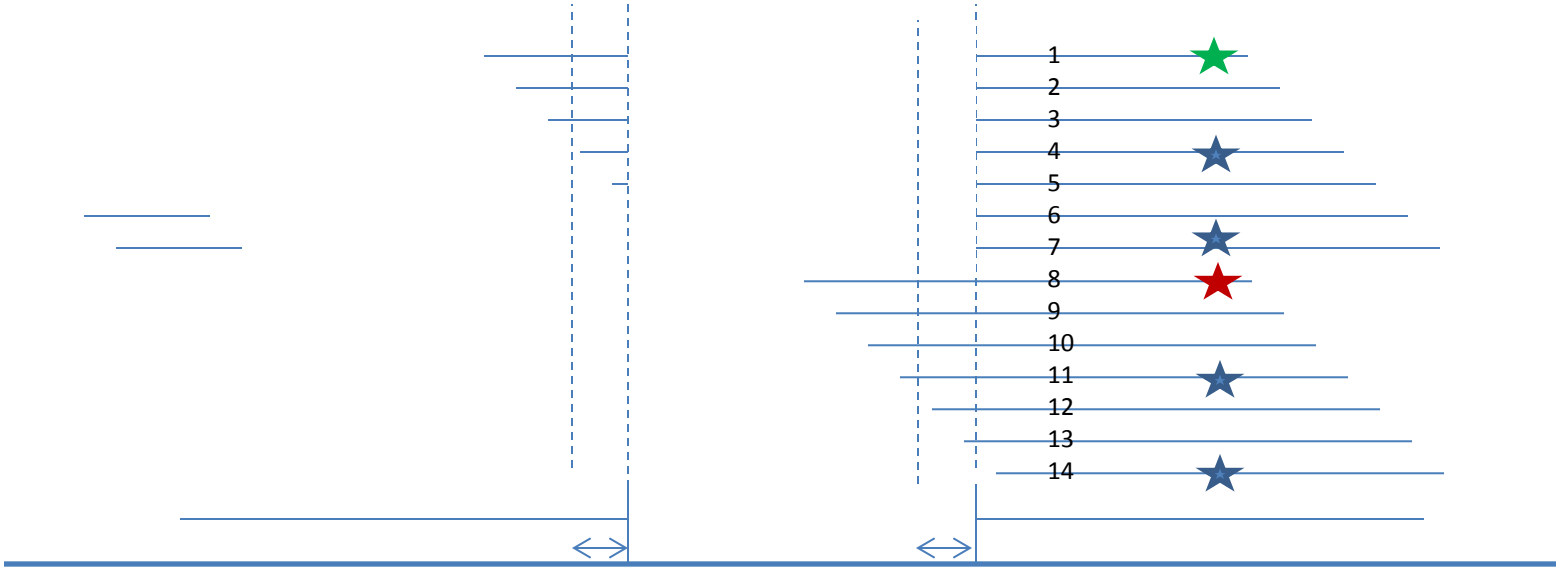
SNV posotion	ref/var	Junction	VAR Junction Intron Count **N VAR ei**	VAR Junction NO Intron Count **N VAR ee**	REF Junction Intron Count **N REF ei**	REF Junction NO Intron Count **N REF ee**	SNV count	REF count	Intron count	No Intron Count	Spanning Reads Count	SNP Loci Reads Count
chr21:33887131	G/A	chr21:33876325-33887123	4	12	4	15	16	19	8	27	35	55



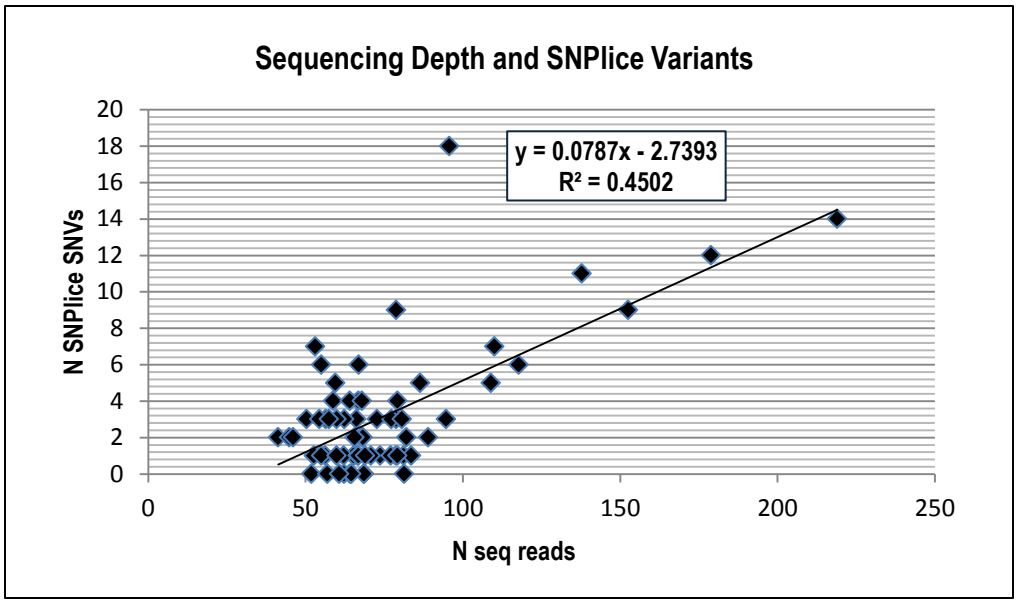
Supplementary Figure 1. *Algorithm employed to categorize the reads used for the SNPlice computation. Part of the fields of the SNPlice output file are shown (top).*

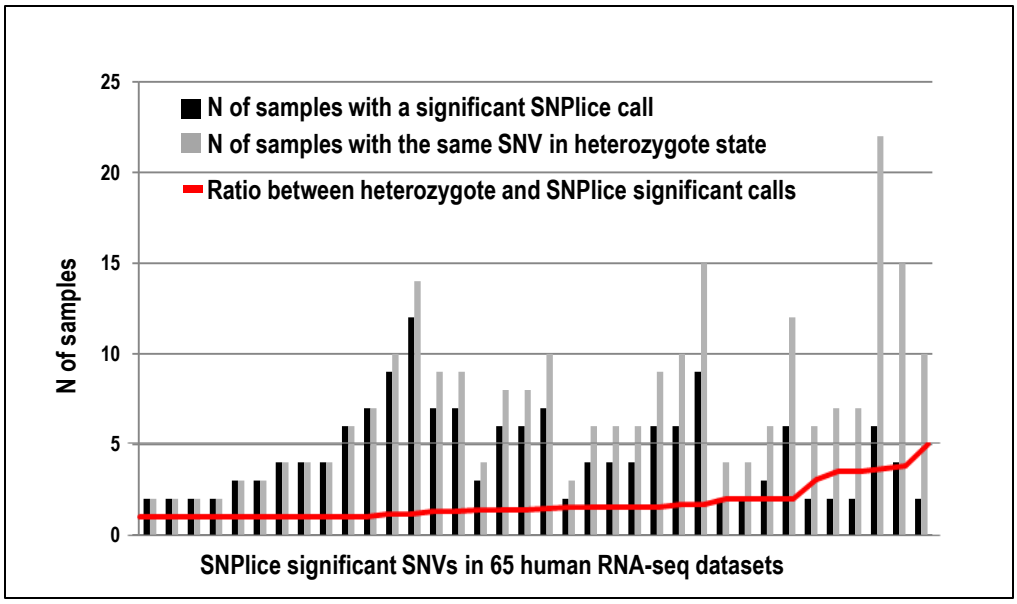
Diagram of cases

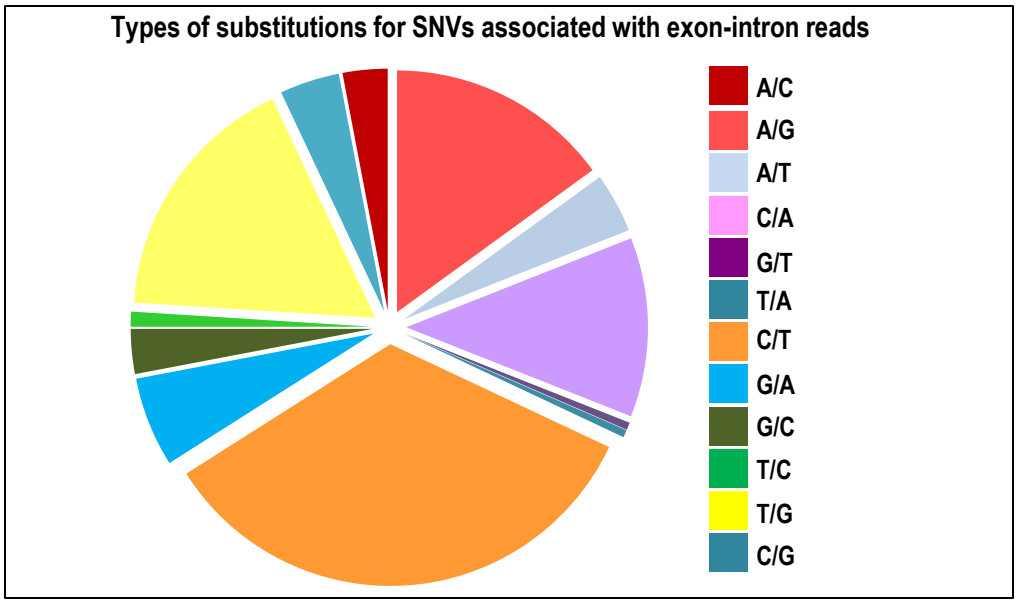
- No star: reference; blue star: expected SNV; other stars: unexpected SNVs.
- “Readthrough”↔(5bp)
- Reads 1,2,3,6,7,8,9,10,11 are considered “Spanning”, (i.e. reads that span SNV locus and either an exon-exon junction, or an exon-intron boundary).
- SNVCount: 2 (reads 7,11); NoSNVCount: 5 (reads 2,3,6,9,10);
  IntronCount: 4 (reads 8,9,10,11); NoIntronCount: 3 (reads 1,2,3)
- SNVIntron: 1 (read 11); SNVNoIntron: 0; NoSNVIntron: 2 (reads 9,10); NoSNVNoIntron 2 (2,3)



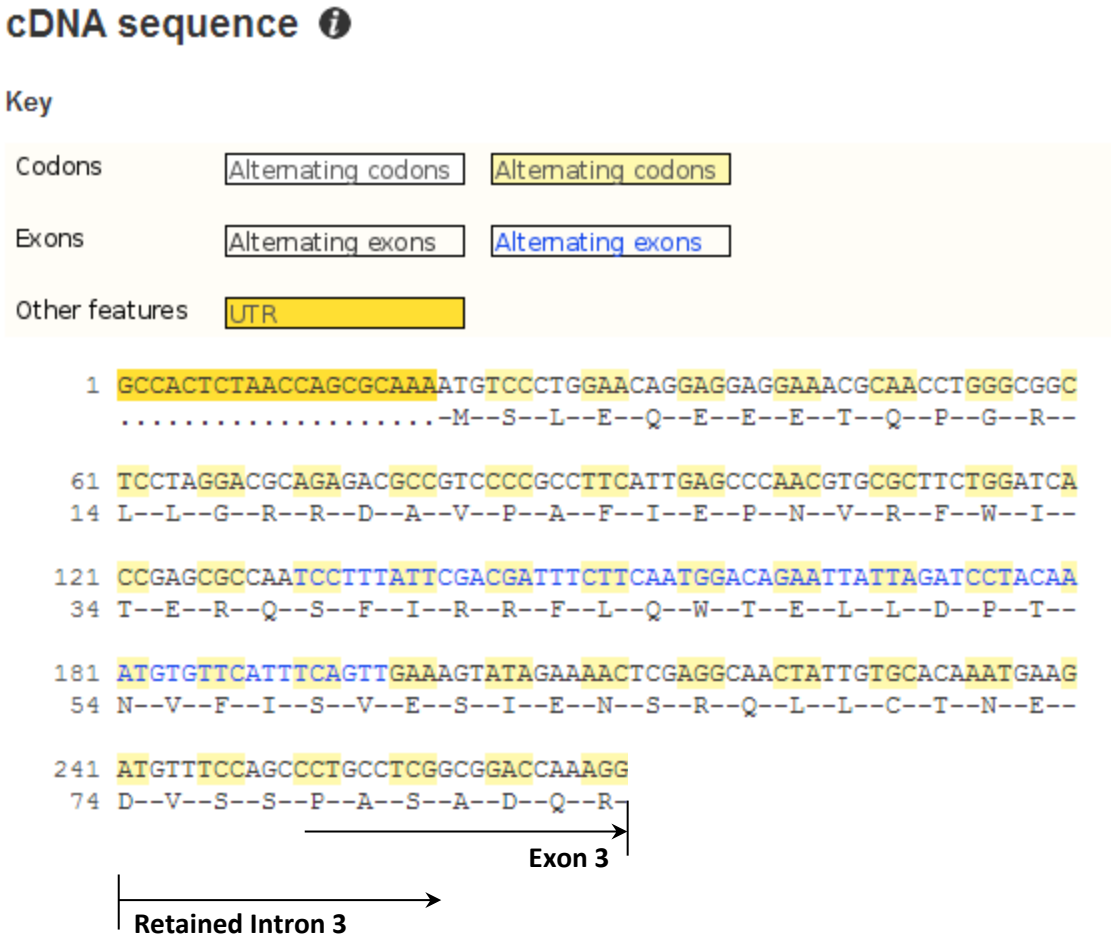
Supplementary Figure 2. *Diagram of reads' categories. Reads mapped less than 5 nt across exon-exon or exon-intron junction are removed.*



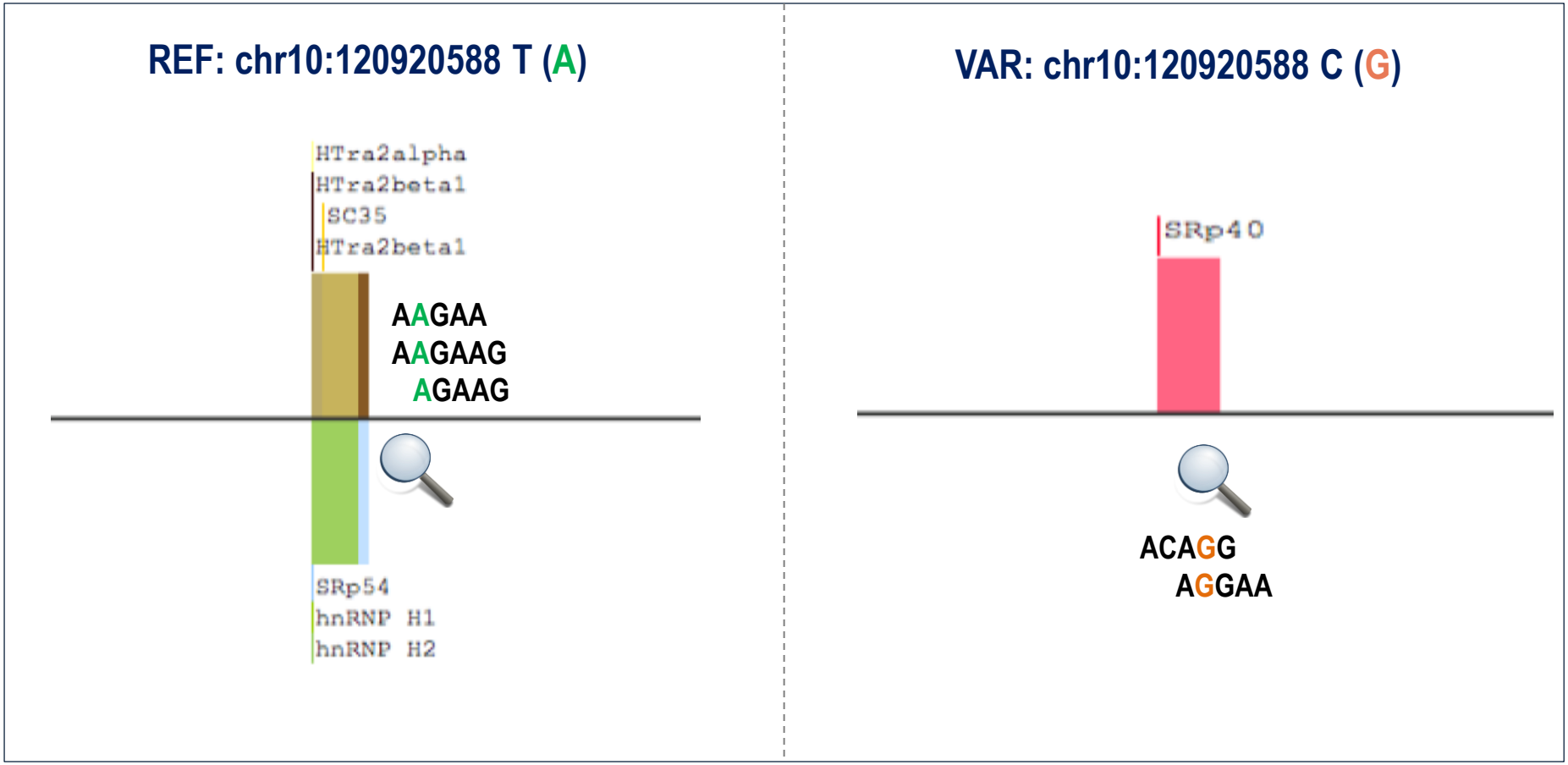




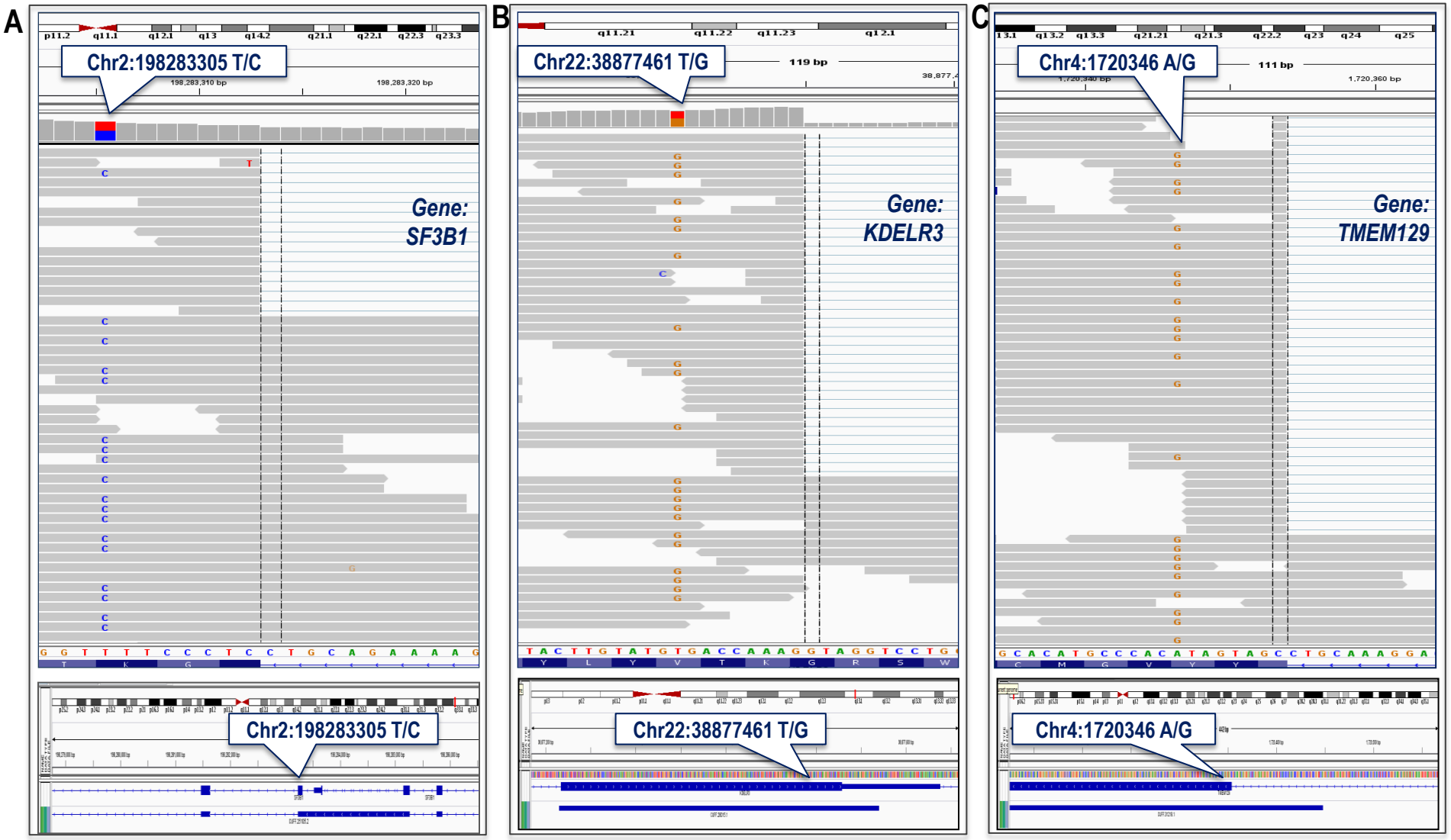
Supplementary Figure 3. SNPlice application on 65 human RNA-seq.
A. Positive correlation between number of sequencing reads and number of SNPlice significant calls (corrected $p<0.05$). The read numbers are presented in millions.
B. Consistency of SNPlice findings across samples. The analysis includes 36 significant variants called in more than one unrelated samples. The red line indicates the ratio between the number of the samples with a heterozygote variant and the number of samples in which the variant was called significant through SNPlice. The ratio ranged between 1 (all the samples heterozygote for the corresponding variant present with significant SNPlice call) and 5 (20% of the heterozygote samples present with a significant SNPlice call for the corresponding variant).
C. Types substitutions highlighted by SNPlice; the most common being C>T (34%). The substitutions are in regards to the sense orientation of the gene open reading frame.



Supplementary Figure 4. Predicted changes in the open reading frame (ORF) and the protein structure of SFXN4 as a result of intron retention on both sides of exon 4. The predicted ORF expands 50 codons downstream of the in-frame exon 3 into intron 3, until it reaches an in-frame stop codon (TAA).

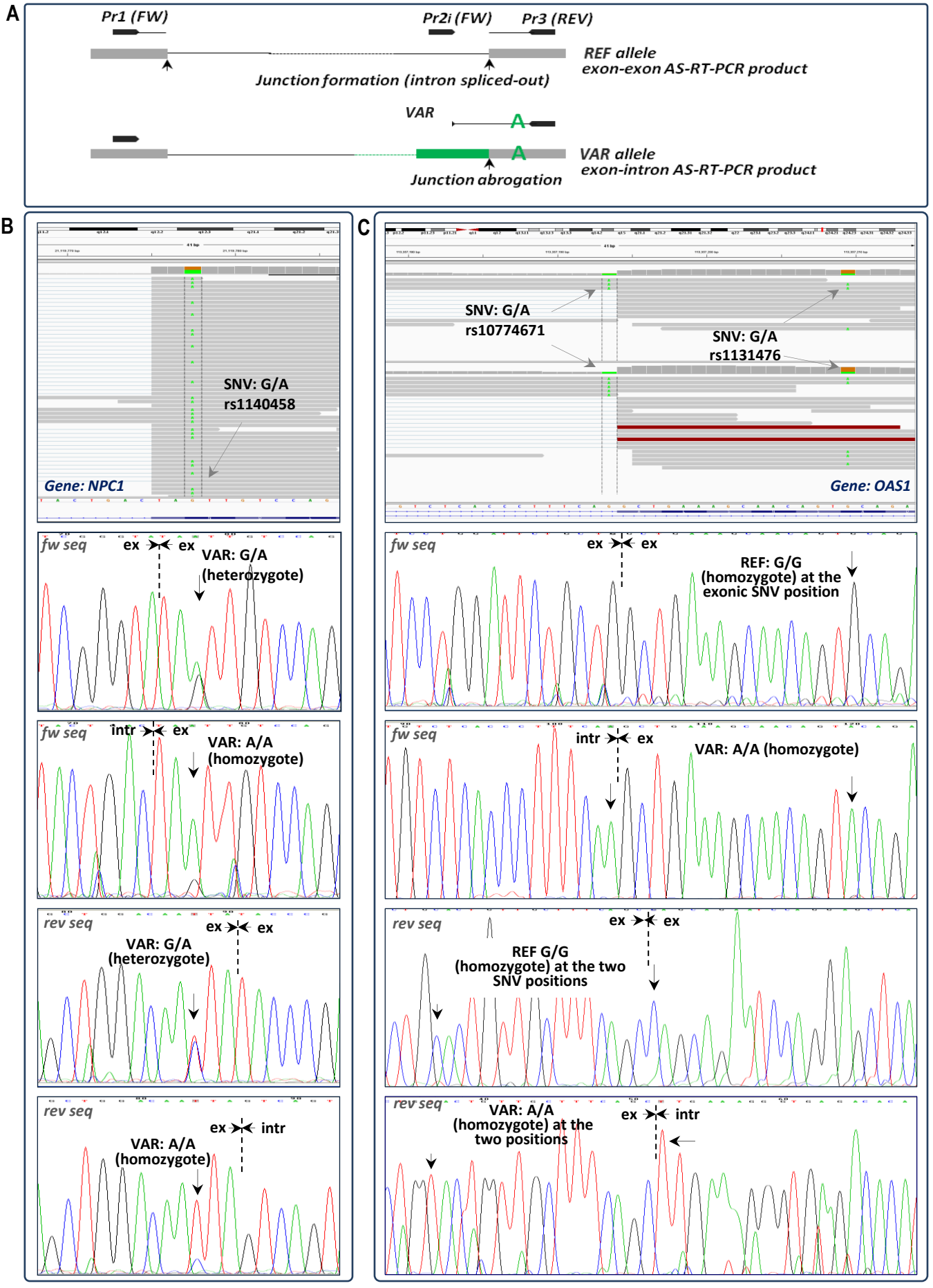


Supplementary Figure 5. Comparison of the binding sites for splice-regulatory molecules recognizing the wild type (left) and the variant (right) sequence for rs10749291 in SFXN4, as modeled by SpliceAid. Splice associated molecules SFRS2, SFRS10, SFRS11, TRA2A, HNRNPH1 and HNRNPH bind only the reference sequence, and two different ones (SFRS5 and SFRS9) bind the variant only. The SpliceAid sequence is in sense (in regards to the SFXN4 ORF direction).



Supplementary Figure 6. A. Upper panel: Read alignment in the region of rs788023 (exon 5, *SF3B1*). Exclusive residing of the variant nucleotide within reads encompassing the acceptor junctions (7 nucleotides from the SNV), is seen. **Lower Panel:** IGV visualization of the Cufflinks assembly of the region of rs788023, showing the presence of an isoform retaining the intron on the acceptor side of exon. The downstream positioned exon-intron-exon structure is also retained in the isoform. **B. Upper panel:** Read alignment in the region of rs12004 (exon 4 of *KDELR3*) shows prevalent residing of the variant nucleotide within reads encompassing the acceptor site (8 nucleotides from the SNV) is seen. **Lower panel:** IGV visualization of the Cufflinks assembly of the region of rs12004 showing an alternative last exon of *KDELR3*. **C. Upper panel:** Read alignment in the region of rs11552262 (ex 2 of *TMEM129*) shows prevalent residing of the variant nucleotide within reads encompassing the acceptor junction (7 nucleotides from the SNV) is seen. This SNV' locus, containing either the reference, or the variant nucleotide, is not a binding site for known splice related molecules. **Lower panel:** IGV visualization of the reads assembly in the region of rs11552262 showing partial intron retention on the acceptor site of exon 2 of *TMEM129*.

Supplementary Figure 7. Allele-Specific Sanger Sequencing



Supplementary Figure 7 (cont.). A. Schematic presentation of the AS-RT-PCR designed to detect co-allelic variant nucleotide and exon-intron boundary. For each allele-specific PCR, three primers were designed: a common forward exonic primer to amplify the SNV locus, and two reverse primers hybridizing in the downstream exon or intron, respectively. The example shows a variant nucleotide (A, in green) which, if splice modulating, is expected to over-dominate the reference allele in the PCR-amplicon containing the exon-intron junction. **B.** Sanger sequencing chromatograms of the allele specific amplification of the region flanking SNV rs1140458 near 3'SS of in the gene *NPC1*: The top two chromatograms show forward, and the bottom two - reverse primer Sanger sequencing of the exon-intron and exon-exon bearing RNA molecules. It is well seen that the alleles with the variant nucleotide (indicated with an arrow) predominate in the PCR amplifying exon-intron junction. In contrast, the amplicon of the canonically spliced exon-exon region (2nd from the top), shows an equal proportion of the variant and the WT allele. **C.** Sanger sequencing chromatograms of the region flanking two closely positioned SNVs in the *OAS1* gene: rs1131476 (16nt in the exon) and rs10774671 (intronic 3' SS). Both SNVs are homozygote variant in the PCR encompassing the exon-intron junction, indicating strong splice modulating potential of the variant allele. In contrast, the canonically spliced product harboring exon-exon junction, shows only the WT allele, indicating complete absence of correctly spliced molecules harboring the variant nucleotide(s). The result is supported by both forward (top two chromatograms), and reverse sequencing. The exonic SNV, without necessarily being involved in the splice modulation, is indicative for alteration of the junction.

Supplementary Table 1. Samples used to test SNPlice Performance, RNA-seq characteristics.

#	Sample Code	TCGA sample code	Sample Origine	Total N Reads	Overal Mapping (%)	Concordant Alignments (%)
1	RPE1	na	primary cell line	137745484	0.969	0.872
2	RPE2	na	primary cell line	178865492	0.971	0.869
3	RPE3	na	primary cell line	109990651	0.97	0.887
4	RPE4	na	primary cell line	152566009	0.966	0.883
5	RPE5	na	primary cell line	218971451	0.959	0.882
6	NT1	TCGA-BH-A0BT	normal breast tissue	59486476	0.96	0.862
7	NT2	TCGA-BH-A0B3	normal breast tissue	66759317	0.94	0.842
8	NT3	TCGA-BH-A0BA	normal breast tissue	53082144	0.95	0.862
9	NT4	TCGA-BH-A0BJ	normal breast tissue	56191732	0.93	0.827
10	NT5	TCGA-BH-A0BQ	normal breast tissue	56448462	0.94	0.847
11	NT6	TCGA-BH-A0BS	normal breast tissue	68587587	0.95	0.839
12	NT7	TCGA-BH-A0C3	normal breast tissue	65443987	0.96	0.855
13	NT8	TCGA-A7-A0D9	normal breast tissue	73756226	0.97	0.884
14	NT9	TCGA-BH-A0DD	normal breast tissue	67885549	0.97	0.885
15	NT10	TCGA-BH-A0DV	normal breast tissue	72730922	0.95	0.86
16	NT11	TCGA-BH-A0E1	normal breast tissue	51892081	0.94	0.825
17	NT12	TCGA-BH-A1ET	normal breast tissue	95645995	0.96	0.881
18	NT13	TCGA-BH-A0HA	normal breast tissue	62227947	0.96	0.853
19	NT14	TCGA-BH-A1EU	normal breast tissue	41295091	0.96	0.87
20	NT15	TCGA-BH-A1FM	normal breast tissue	81172903	0.95	0.852
21	NT16	TCGA-BH-A1FU	normal breast tissue	44882440	0.96	0.874
22	NT17	TCGA-BH-A2FF	normal breast tissue	88931561	0.95	0.851
23	NT18	TCGA-BH-A18N	normal breast tissue	78830417	0.94	0.849
24	NT19	TCGA-BH-A0C0	normal breast tissue	77311482	0.94	0.842
25	NT20	TCGA-BH-A0E0	normal breast tissue	58775971	0.92	0.81
26	NT21	TCGA-BH-A0DL	normal breast tissue	78830417	0.94	0.849
27	NT22	TCGA-BH-A0D0	normal breast tissue	66233339	0.96	0.864
28	NT23	TCGA-BH-A1FJ	normal breast tissue	94684175	0.95	0.867
29	NT24	TCGA-BH-A18S	normal breast tissue	50296853	0.95	0.871
30	NT25	TCGA-BH-A18R	normal breast tissue	80615943	0.96	0.879
31	NT26	TCGA-BH-A0AY	normal breast tissue	54373442	0.91	0.804
32	NT27	TCGA-BH-A0DB	normal breast tissue	65303269	0.94	0.853
33	NT28	TCGA-BH-A0BV	normal breast tissue	78745529	0.94	0.847
34	NT29	TCGA-BH-A0DH	normal breast tissue	66945837	0.94	0.835
35	NT30	TCGA-BH-A0DQ	normal breast tissue	83623515	0.94	0.847
36	TP1	TCGA-BH-A0BT	tumor breast tissue	55004971	0.95	0.854
37	TP2	TCGA-BH-A0B3	tumor breast tissue	68221259	0.94	0.843
38	TP3	TCGA-BH-A0BA	tumor breast tissue	65552046	0.95	0.866
39	TP4	TCGA-BH-A0BJ	tumor breast tissue	64112433	0.96	0.87
40	TP5	TCGA-BH-A0BQ	tumor breast tissue	52660375	0.96	0.873
41	TP6	TCGA-BH-A0BS	tumor breast tissue	70793519	0.96	0.867
42	TP7	TCGA-BH-A0C3	tumor breast tissue	62082509	0.96	0.871
43	TP8	TCGA-A7-A0D9	tumor breast tissue	62308445	0.94	0.832
44	TP9	TCGA-BH-A0DD	tumor breast tissue	64107647	0.95	0.859
45	TP10	TCGA-BH-A0DV	tumor breast tissue	86417764	0.97	0.883
46	TP11	TCGA-BH-A0E1	tumor breast tissue	59934510	0.95	0.867
47	TP12	TCGA-BH-A1ET	tumor breast tissue	108899128	0.96	0.883
48	TP13	TCGA-BH-A0HA	tumor breast tissue	56969739	0.94	0.848
49	TP14	TCGA-BH-A1EU	tumor breast tissue	117745484	0.96	0.88
50	TP15	TCGA-BH-A1FM	tumor breast tissue	77091719	0.95	0.867
51	TP16	TCGA-BH-A1FU	tumor breast tissue	55004971	0.95	0.854
52	TP17	TCGA-BH-A2FF	tumor breast tissue	79237177	0.95	0.851
53	TP18	TCGA-BH-A18N	tumor breast tissue	66610964	0.96	0.879
54	TP19	TCGA-BH-A0C0	tumor breast tissue	57426535	0.96	0.871
55	TP20	TCGA-BH-A0E0	tumor breast tissue	46047258	0.93	0.828
56	TP21	TCGA-BH-A0DL	tumor breast tissue	81338216	0.92	0.826
57	TP22	TCGA-BH-A0D0	tumor breast tissue	66843177	0.96	0.88
58	TP23	TCGA-BH-A1FJ	tumor breast tissue	79248775	0.95	0.86
59	TP24	TCGA-BH-A18S	tumor breast tissue	67890588	0.96	0.866
60	TP25	TCGA-BH-A18R	tumor breast tissue	82063135	0.95	0.851
61	TP26	TCGA-BH-A0AY	tumor breast tissue	68989695	0.90	0.798
62	TP27	TCGA-BH-A0DB	tumor breast tissue	68515944	0.92	0.826
63	TP28	TCGA-BH-A0BV	tumor breast tissue	59906501	0.96	0.866
64	TP29	TCGA-BH-A0DH	tumor breast tissue	64608301	0.95	0.859
65	TP30	TCGA-BH-A0DQ	tumor breast tissue	60724531	0.94	0.855

Supplementary Table 2. Primer sequences and amplicon sizes for AS-RT-PCR amplification of the surrounding exon-exon and exon-intron regions of two selected SNVs suggested modulating splicing.

SNV	Forward (FW) Primer	Reverse (REV) Primer	AS-RT-PCR size (bp)
18:21119777 G/A	Pr1: Forward Exonic (FW_E_NPC1)	Pr3: Reverse Exonic (REV_E_NPC1)	223
	TGCAGAACTGGTCAGTGATATTG		
	Pr2: Forward Intronic (FW_I_NPC1)	GAGGAAGGGCACGACTACAC	220
	TTCCAAGAAAGTCTATTTTCAGTGAG		
chr12:113357193 G/A	Pr1: Forward Exonic (FW_E_OAS1)	Pr3: Reverse Exonic (REV_E_OAS1)	288
	CGGACCCTACAGGAAACTTG		
chr12:113357209 G/A	Pr2: Forward Intronic (FW_I_OAS1)	GGGTACTCATGTGTTCCAATGT	218
	AGCAGCTGGGGCTTGTTAGT		

Supplementary Table 3. SNVs assessed significant through SNPlice for association with intron retention. The annotation is performed using SeattleSeq v8.01, dbSNP build 138. DTB – Distance to Boundary.

Chr:location (hg19)	nt	rsID	SNPlice						Annotation						
			N VAR ei	N VAR ee	N REF ei	N REF ee	LOD	FDR (q)	Gene	Transcript Accession	Function	AA change	Protein Position/ Length	cDNA Position	DTB
chr19:54704760	A/C	3810232	43	0	0	26	12.17	1E-16	RPS9	NM_001013.3	intron	none	NA	NA	4
chr1:223718651	G/A	61823553	14	7	4	103	5.47	1E-07	CAPN8	NM_001143962.1	missense	THR,MET	492/604	1475	11
chr7:99955866	G/A	61735533	13	0	0	16	9.80	3E-06	PILRB	NM_178238.3	5-prime-UTR	none	NA	NA	24
chr17:5289580	A/G	11209	24	5	7	29	4.13	3E-05	NUP88	NM_002532.4	synonymous	HIS	724/742	2172	10
chr7:128034629	C/T	2228075	8	0	0	22	9.58	6E-05	IMPDH1	XM_005250316.1	synonymous	ALA	319/394	957	25
chr22:21830934	G/A	469205	18	7	2	31	4.96	8E-05	PI4KAP2	NR_003700.1	noncoding exon	none	NA	1620	2
chr16:70164334	A/G	10852462	8	0	0	20	9.45	1E-04	PDPR	XM_005256019.1	5-prime-UTR	none	NA	NA	9
chr17:73588058	G/T	736522	9	1	0	19	7.95	0.0006	MYO15B	NR_003587.2	noncoding exon	none	NA	2359	1
chr19:4511140	C/G	7255715	31	0	0	5	9.44	0.0008	PLIN4	XM_005259641.1	synonymous	GLY	944/1372	2832	12
chr19:11688460	T/C	2071484	15	3	0	12	6.79	0.0014	ACP5	XM_005259939.1	5-prime-UTR	none	NA	NA	35
chr22:22049783	C/T	12484060	18	0	10	15	5.77	0.0043	PPIL2	NM_014337.3	3-prime-UTR	none	NA	NA	3
chr17:27000391	A/G	9904043	13	1	14	23	3.87	0.0081	SUPT6H	NM_003170.3	5-prime-UTR	none	NA	NA	3
chr15:75130093	T/C	12898397	3	0	0	46	9.35	0.0092	ULK3	XM_005254291.1	intron	none	NA	NA	2
chr17:42084067	T/C	55708447	16	33	0	38	5.25	0.0095	NAGS	NM_153006.2	synonymous	PHE	362/535	1086	11
chr6:31938120	C/T	45531831	12	2	2	14	4.86	0.0114	DXO	NM_005510.3	synonymous	ARG	316/397	948	1
chr22:21830934	G/A	469205	8	6	3	52	4.29	0.0125	PI4KAP2	NR_003700.1	noncoding exon	none	NA	1620	2
chr2:113943470	A/G	2241976	6	4	0	24	6.15	0.0183	PSD4	XM_005263634.1	synonymous	GLY	422/1057	1266	17
chr20:52788190	G/T	35873579	3	0	0	28	8.64	0.0204	CYP24A1	NM_000782.4	synonymous	ARG	157/515	469	20
chr17:73519413	C/T	8064529	12	5	1	15	4.55	0.0297	TSEN54	NM_207346.2	missense	ALA,VAL	437/527	1310	4
chr17:80585094	C/T	2291393	4	0	1	22	7.08	0.0328	WDR45B	NM_019613.3	synonymous	LYS	106/345	318	15