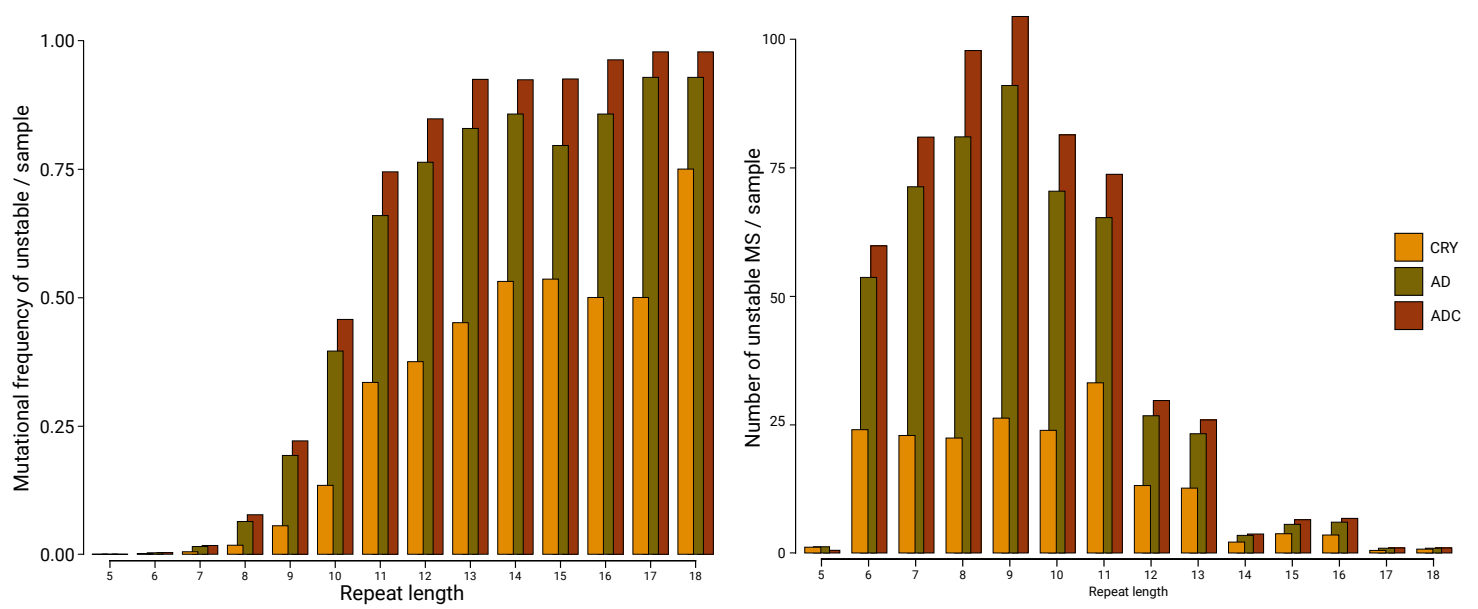
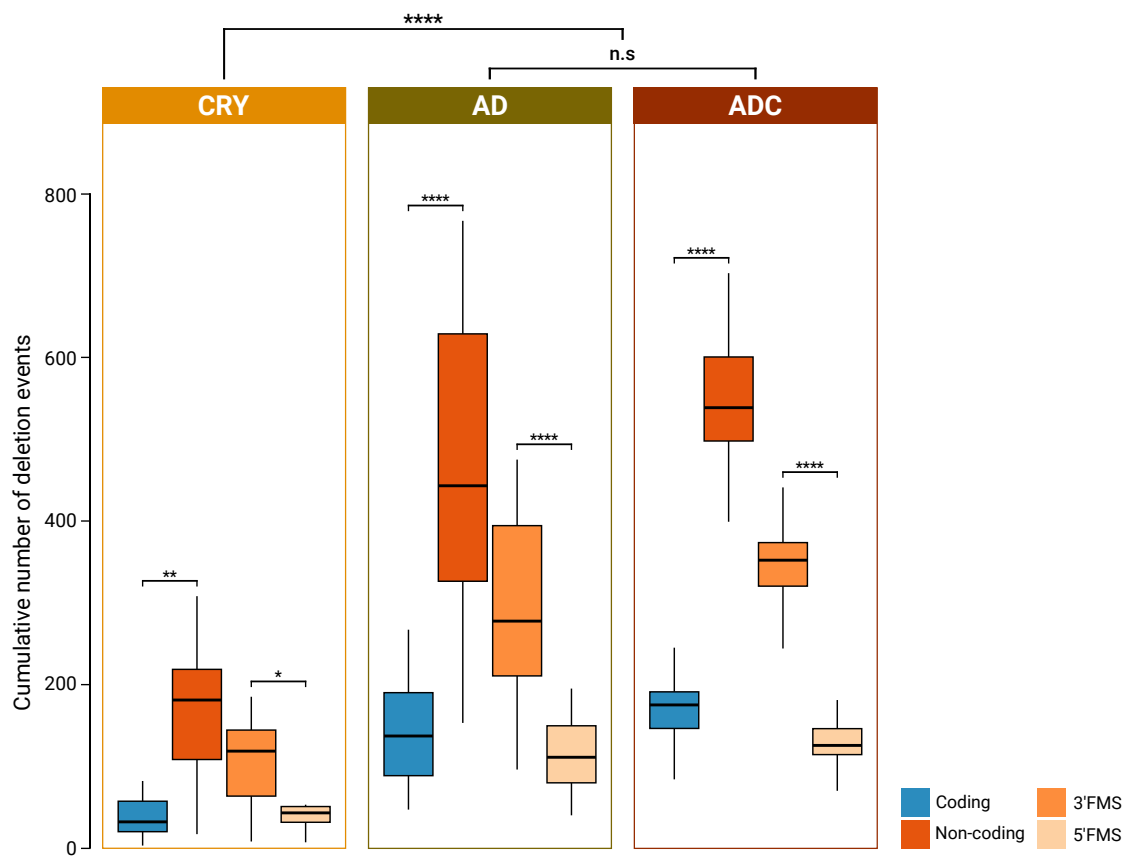


A**B**

Supplementary Figure S1

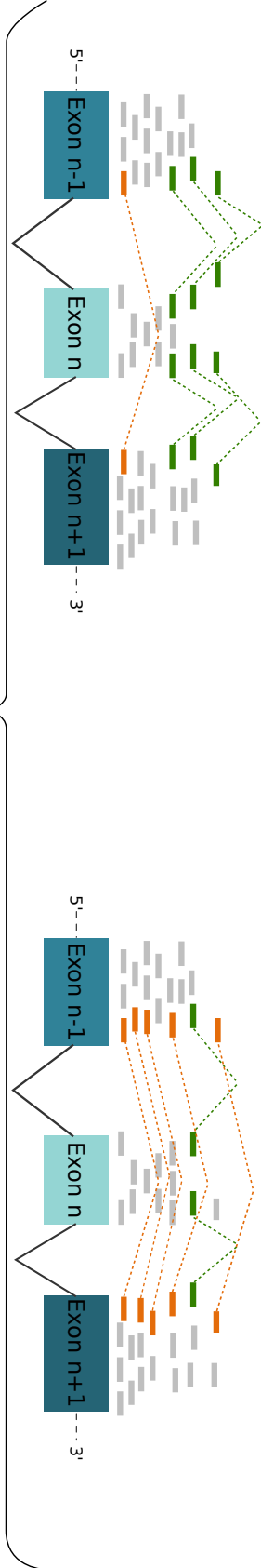
A

NORMAL TISSUE

TUMORAL TISSUE

$$V_i \in [1, n_samples]$$
$$PSI_i^{Normal} = \frac{Inclusion\ Reads_i}{Inclusion\ Reads_i + Exclusion\ Reads_i}$$

$$PSI_i^{Tumoral} = \frac{Inclusion\ Reads_i}{Inclusion\ Reads_i + Exclusion\ Reads_i}$$



Adjusted t-test
↓
Significant Exon
Exclusion/ Inclusion

B

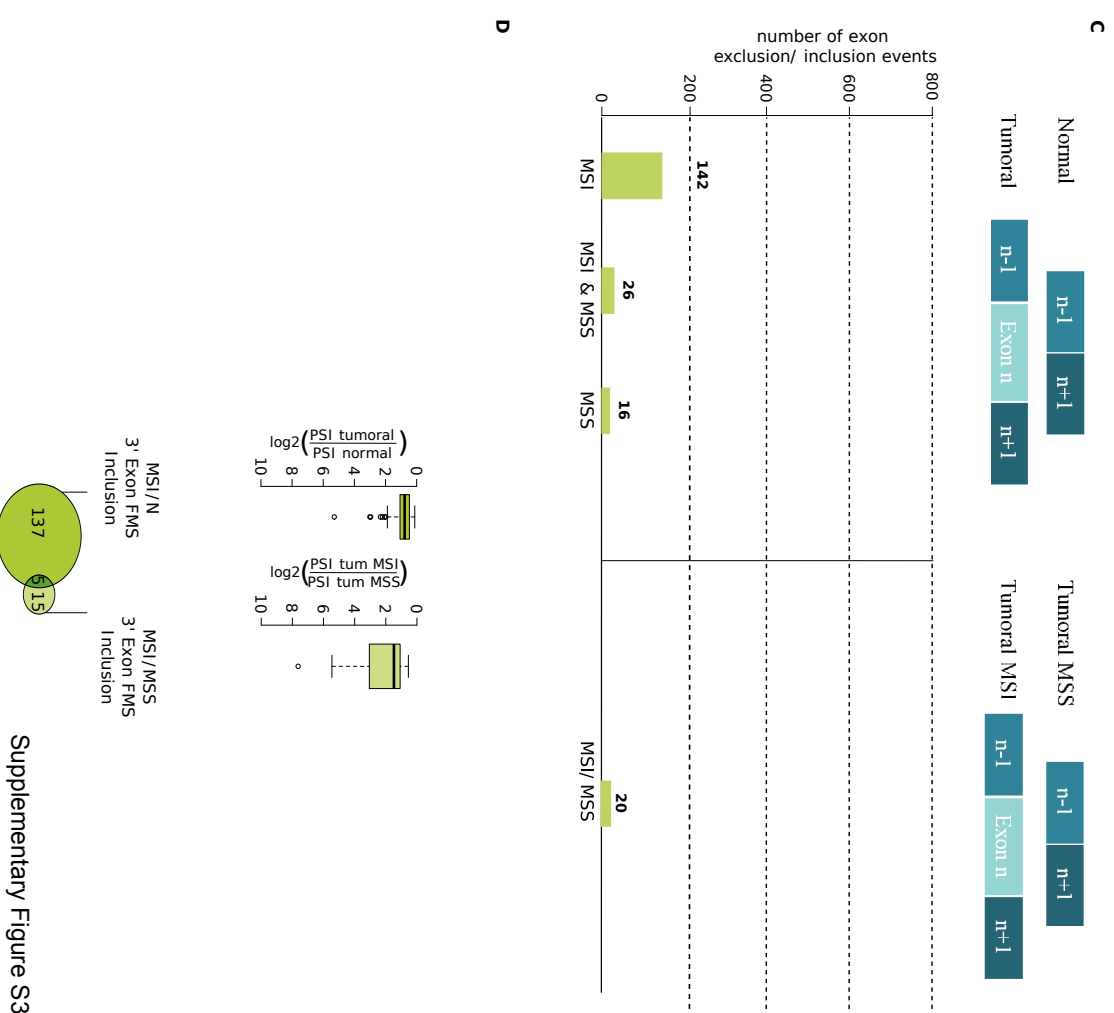
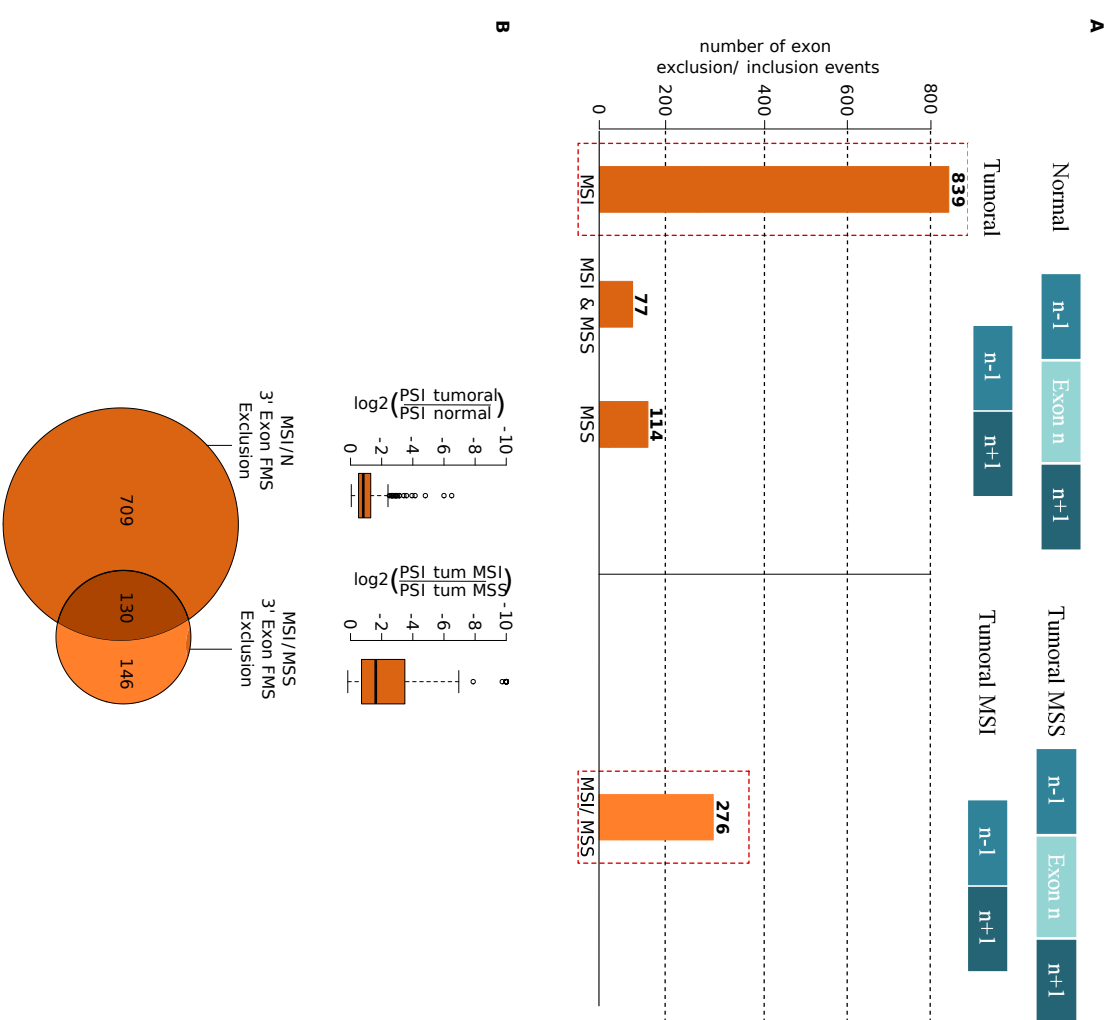
TUMORAL MSS

TUMORAL MSI

$$V_i \in [1, n_samples]$$
$$PSI_i^{Tum\ MSS} = \frac{Inclusion\ Reads_i}{Inclusion\ Reads_i + Exclusion\ Reads_i}$$

$$PSI_i^{Tum\ MSI} = \frac{Inclusion\ Reads_i}{Inclusion\ Reads_i + Exclusion\ Reads_i}$$

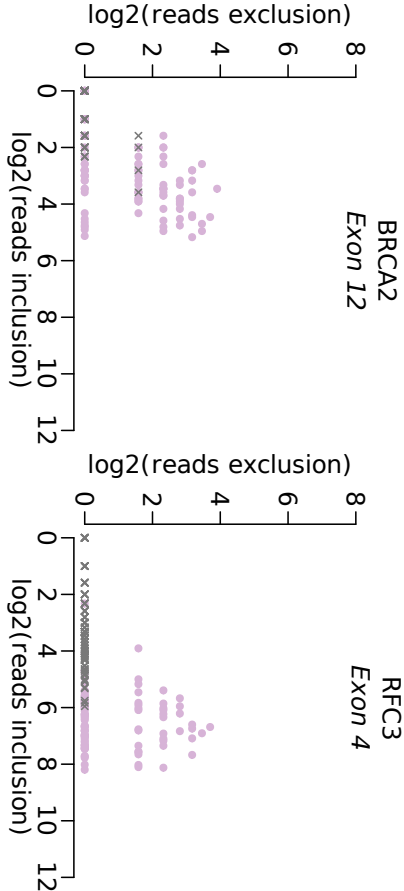
Adjusted t-test
↓
Significant Exon
Exclusion/ Inclusion



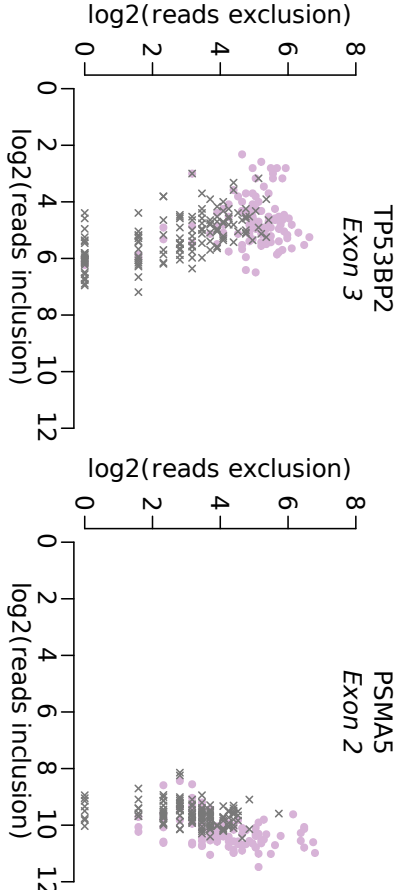
Supplementary Figure S3

A

Constitutive exons
Frequency (Percentage of tumors)

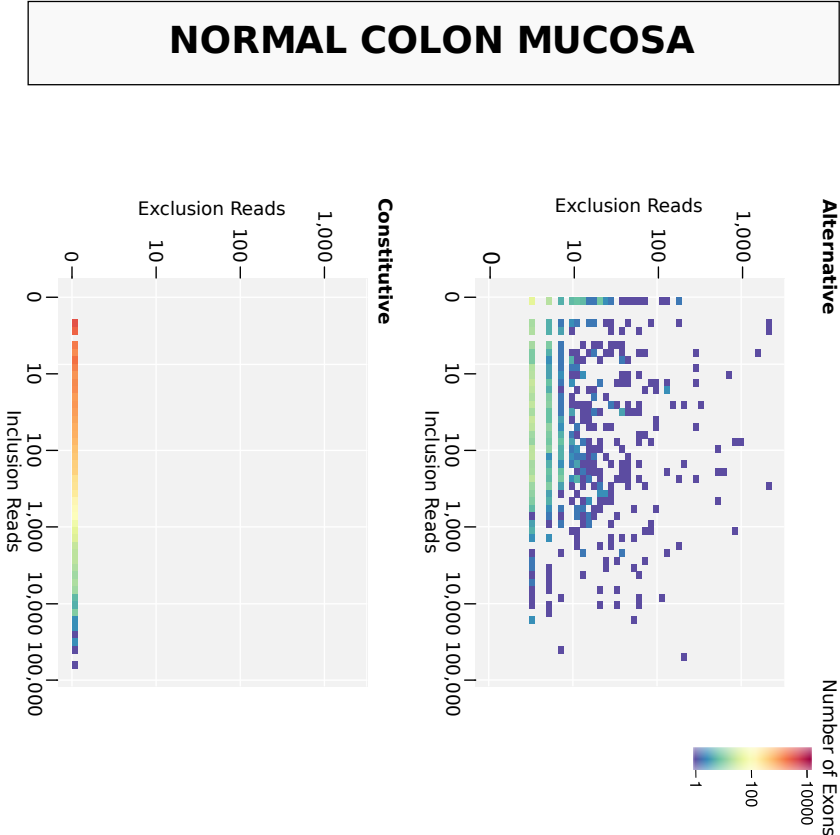


Alternative exons
Frequency (Percentage of tumors)

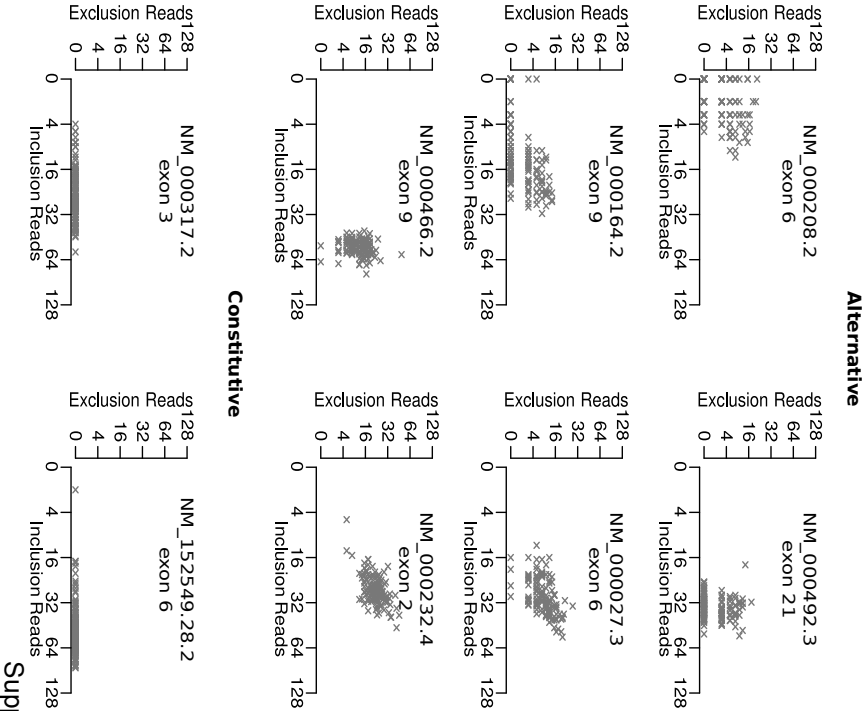


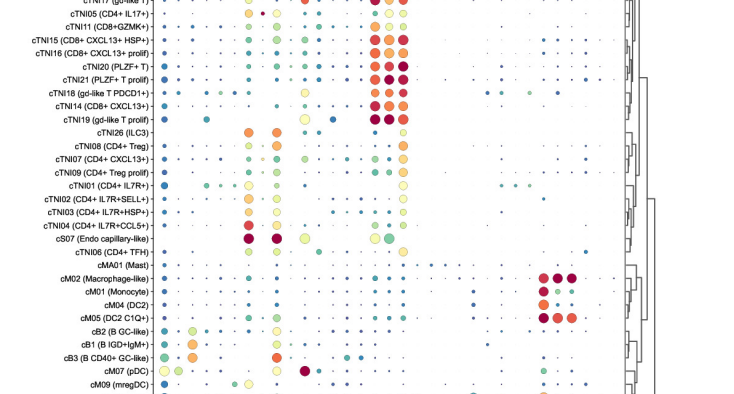
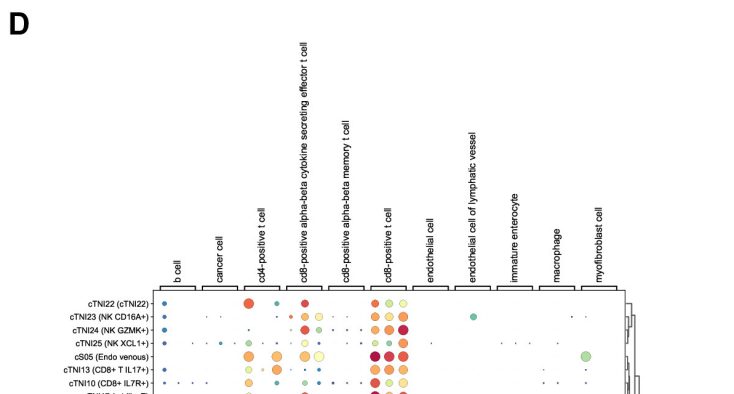
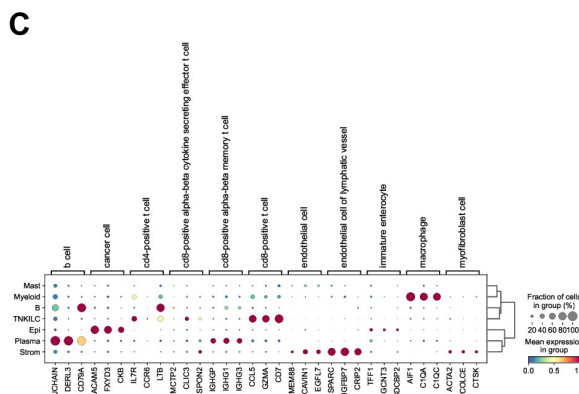
B

● Paired normal x Tumoral MSI



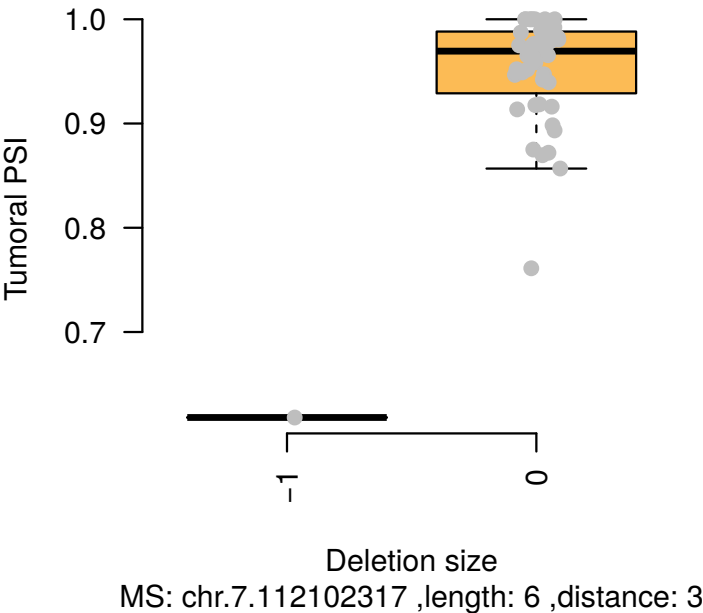
C



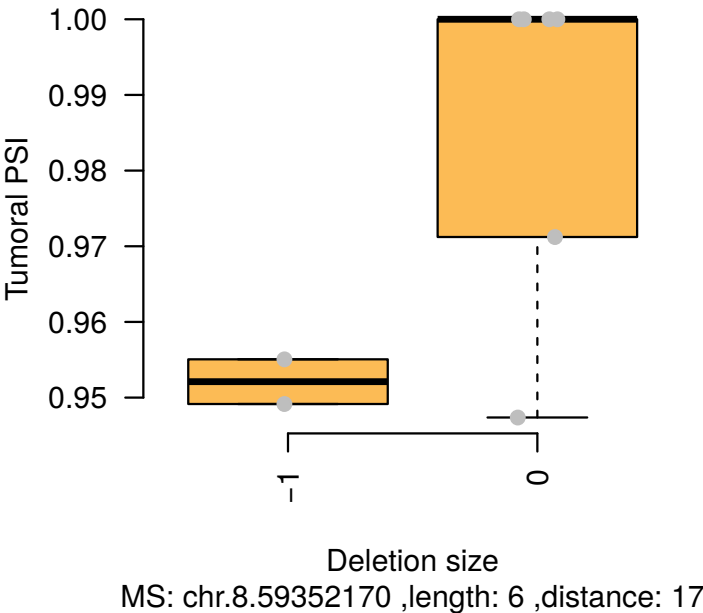


Supplementary Figure S6

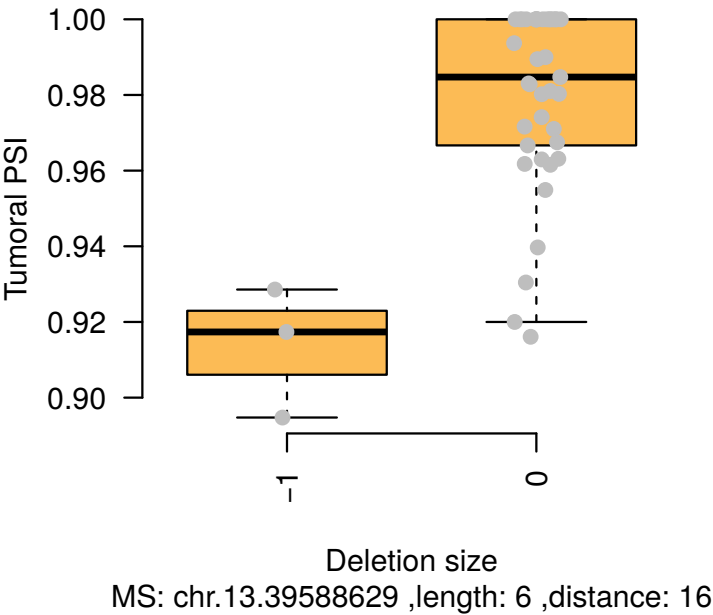
gene: IFRD1 ,exon: 146 ,type: Alternative



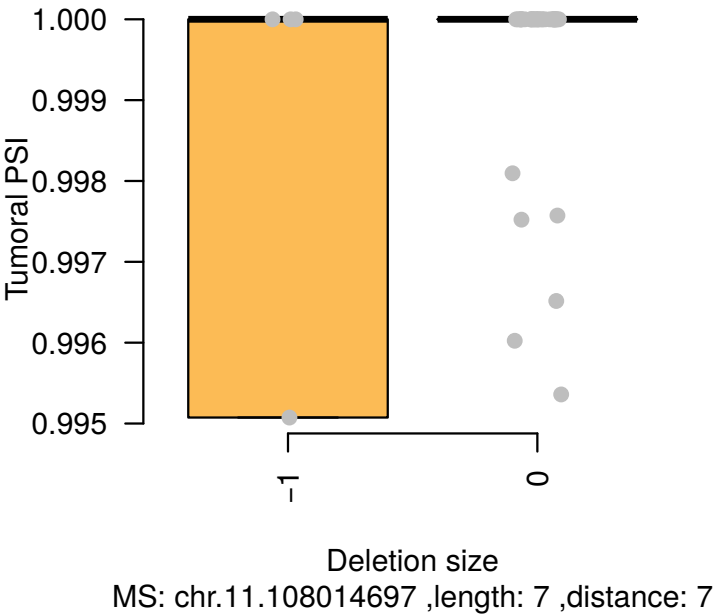
gene: UBXN2B ,exon: 113 ,type: Alternative



gene: C13orf23 ,exon: 12 ,type: Alternative

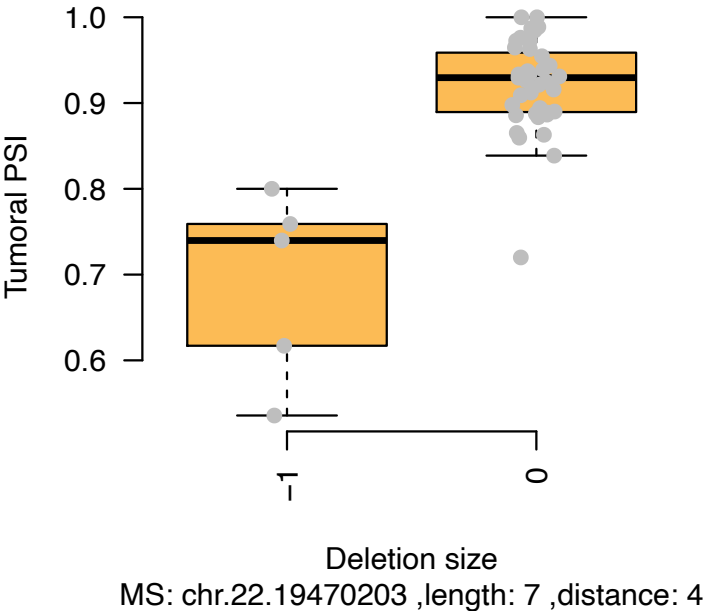


gene: ACAT1 ,exon: 1 ,type: Alternative

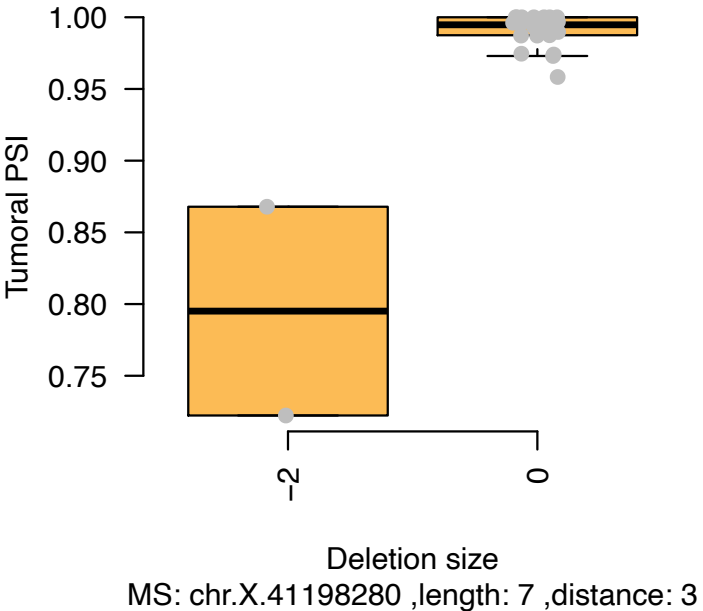


Supplementary Figure S6

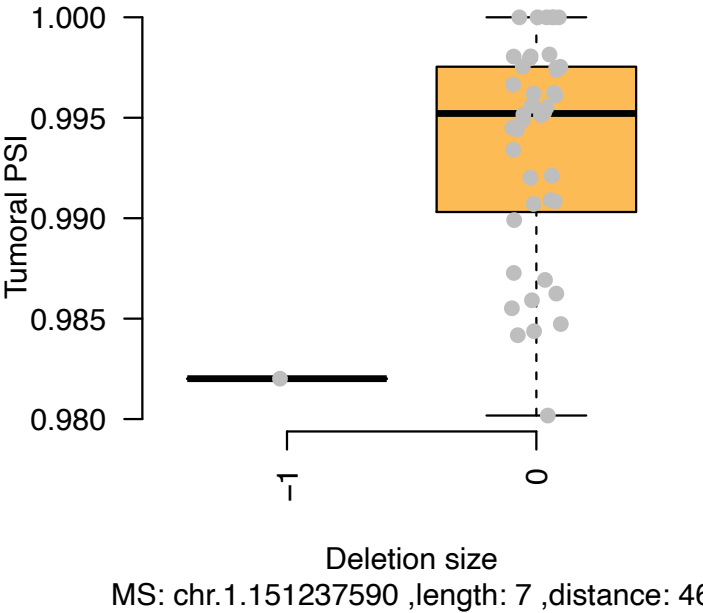
gene: CDC45L ,exon: 91 ,type: Alternative



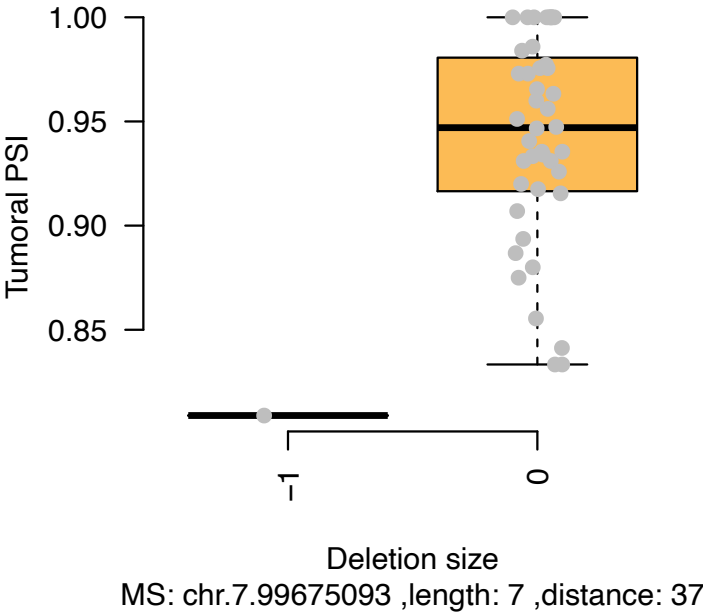
gene: DDX3X ,exon: 80 ,type: Alternative



gene: PSMD4 ,exon: 102 ,type: Alternative

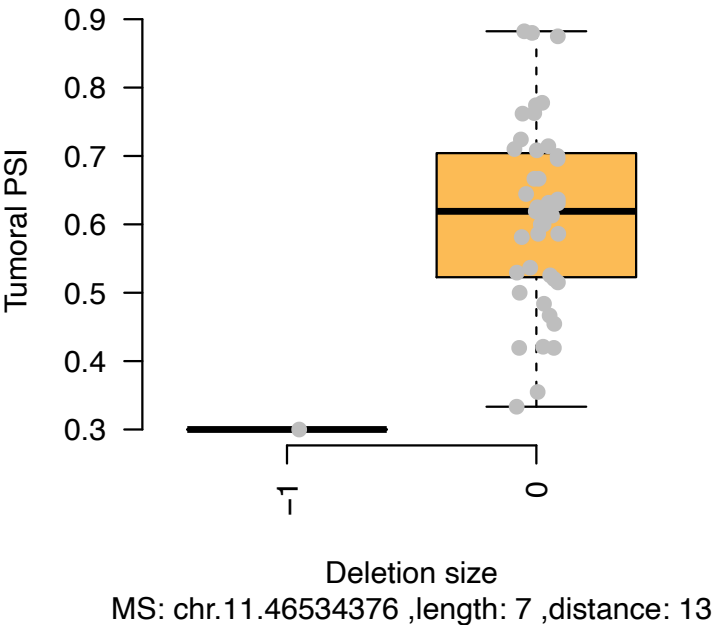


gene: ZNF3 ,exon: 80 ,type: Alternative

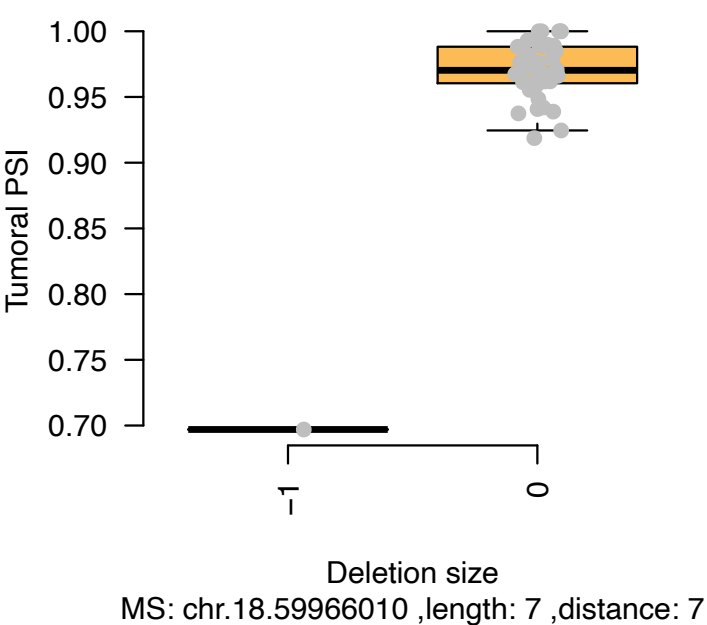


Supplementary Figure S6

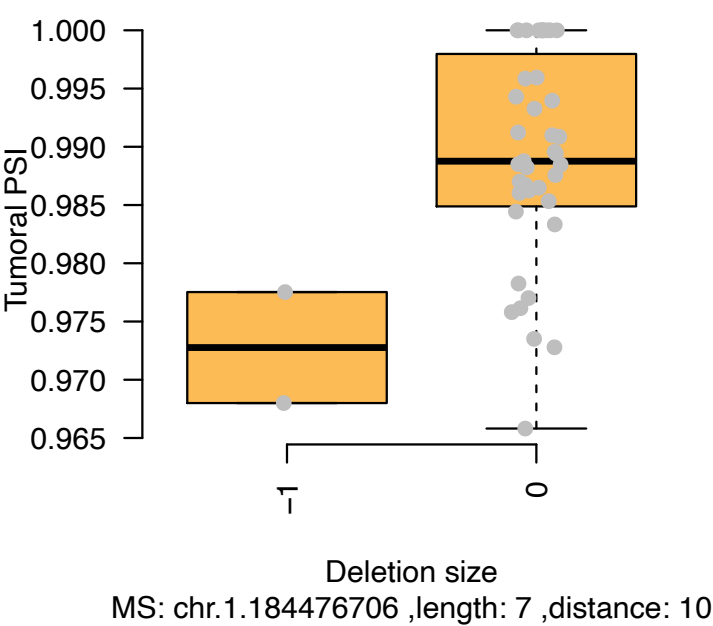
gene: **AMBRA1** ,exon: 146 ,type: Alternative



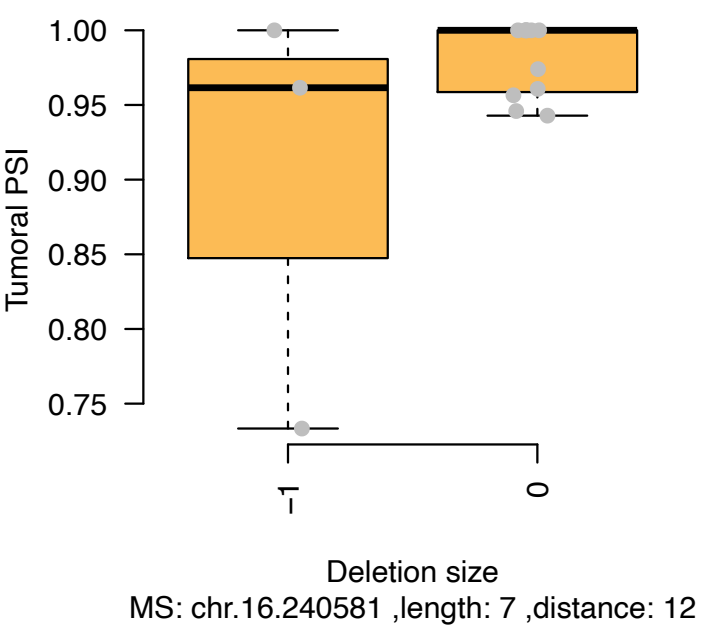
gene: **KIAA1468** ,exon: 78 ,type: Alternative



gene: **C1orf21** ,exon: 80 ,type: Alternative

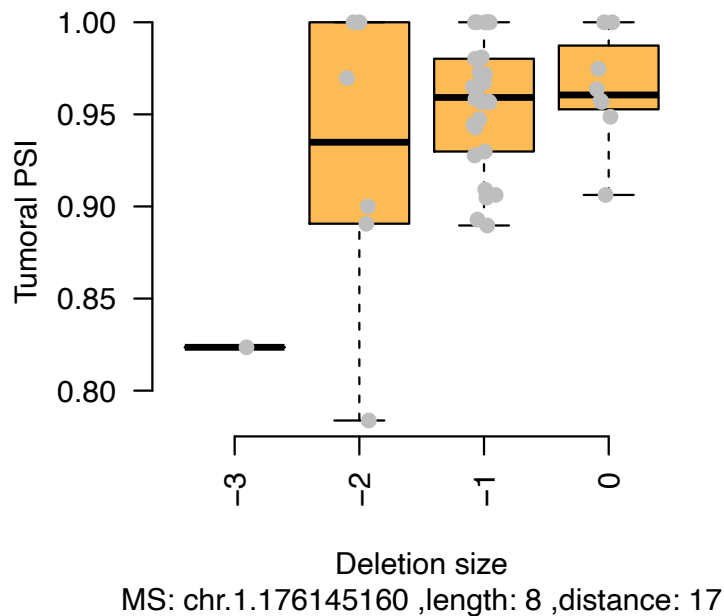


gene: **LUC7L** ,exon: 135 ,type: Alternative

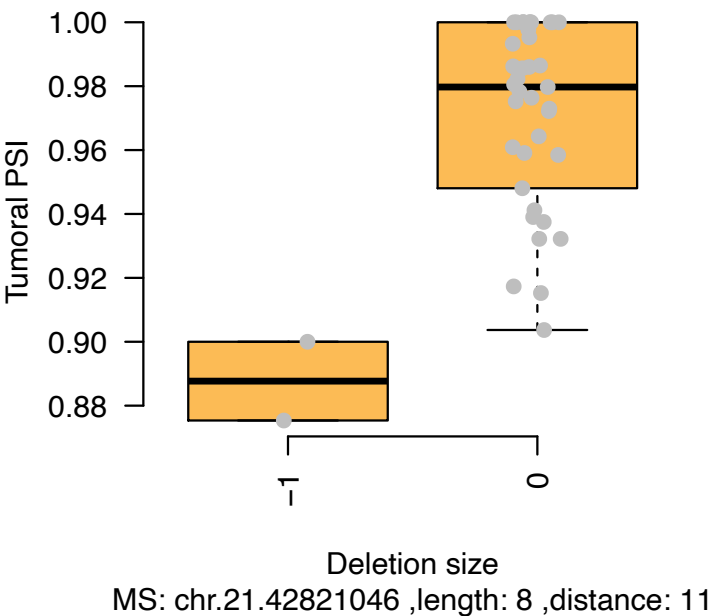


Supplementary Figure S6

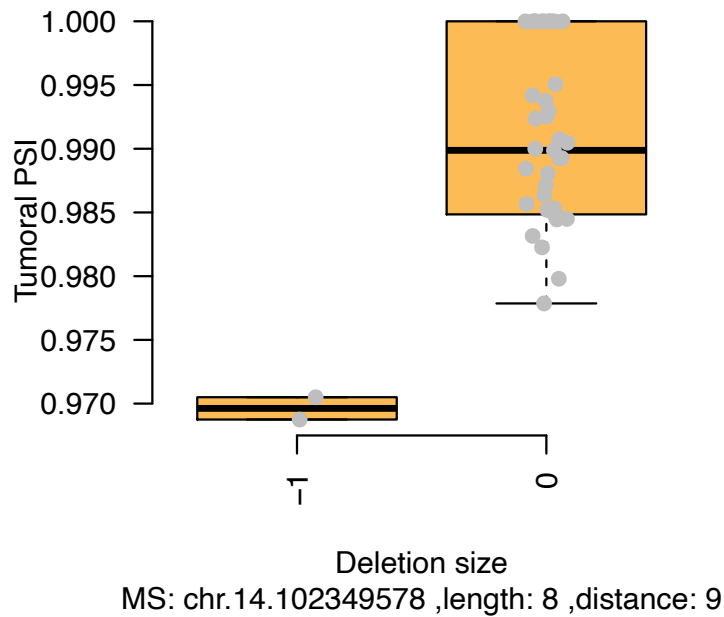
gene: RFWD2 ,exon: 80 ,type: Alternative



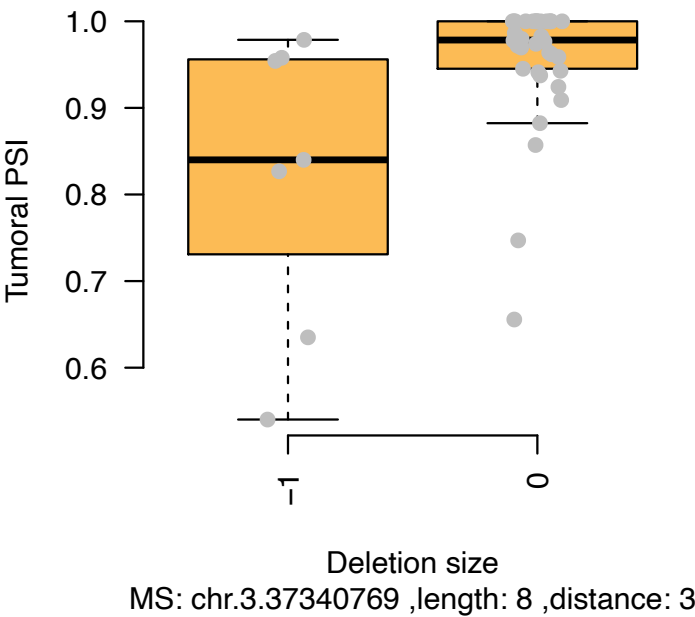
gene: MX1 ,exon: 48 ,type: Alternative



gene: PPP2R5C ,exon: 113 ,type: Alternative

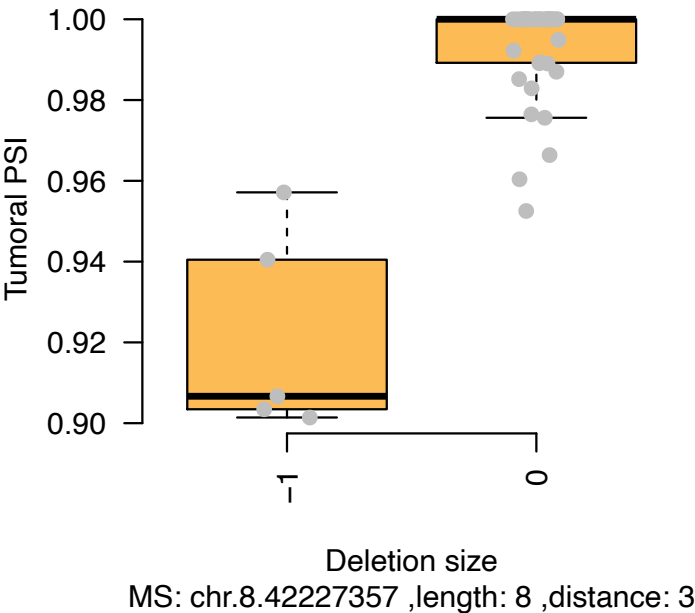


gene: GOLGA4 ,exon: 146 ,type: Alternative

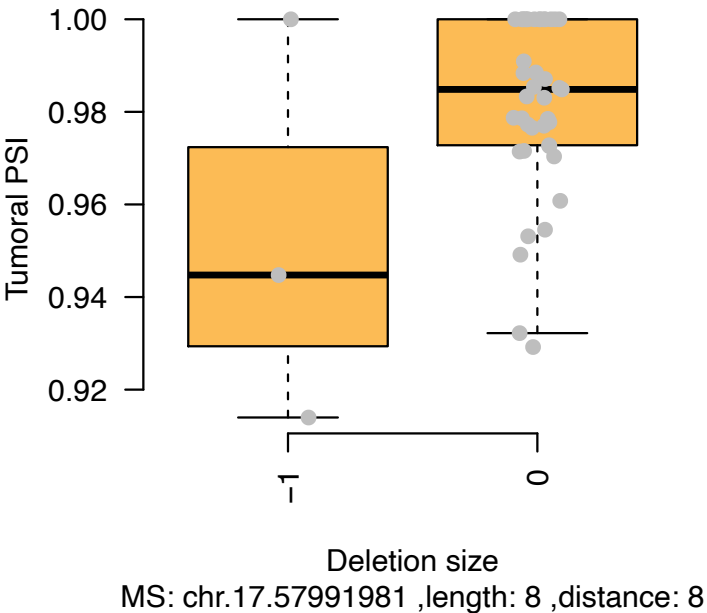


Supplementary Figure S6

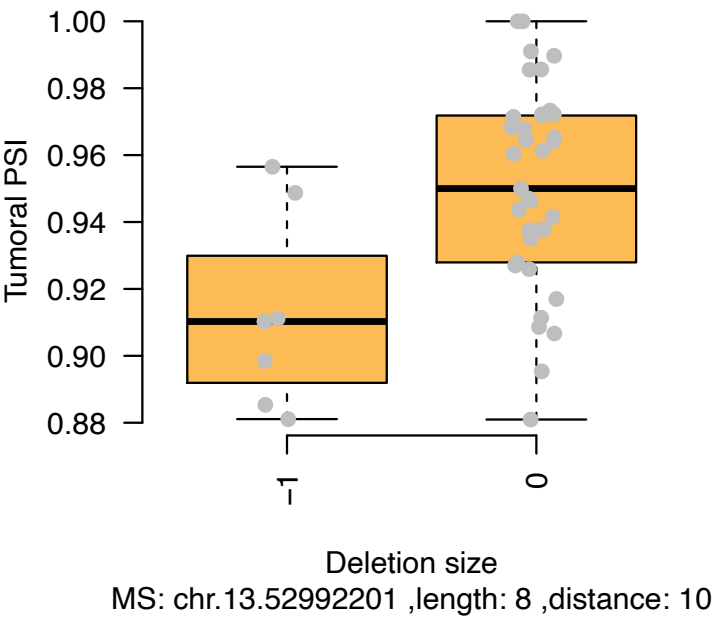
gene: POLB ,exon: 30 ,type: Alternative



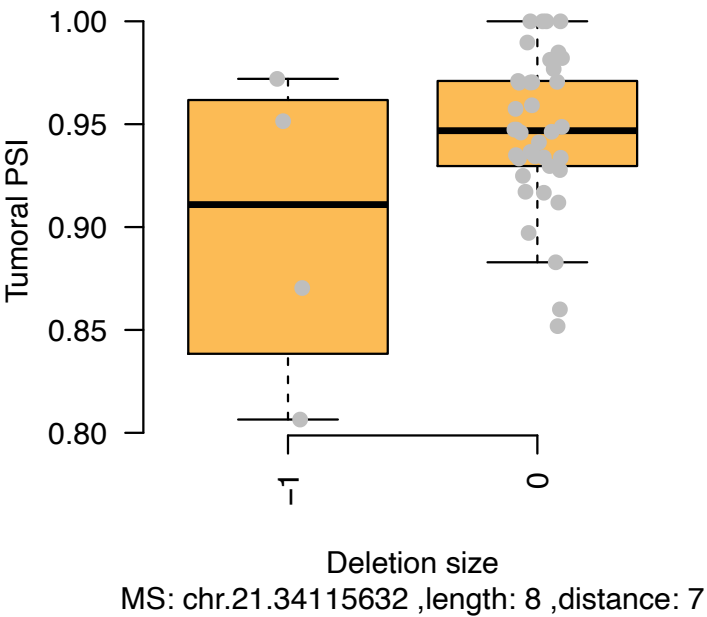
gene: RPS6KB1 ,exon: 91 ,type: Alternative



gene: VPS36 ,exon: 12 ,type: Alternative

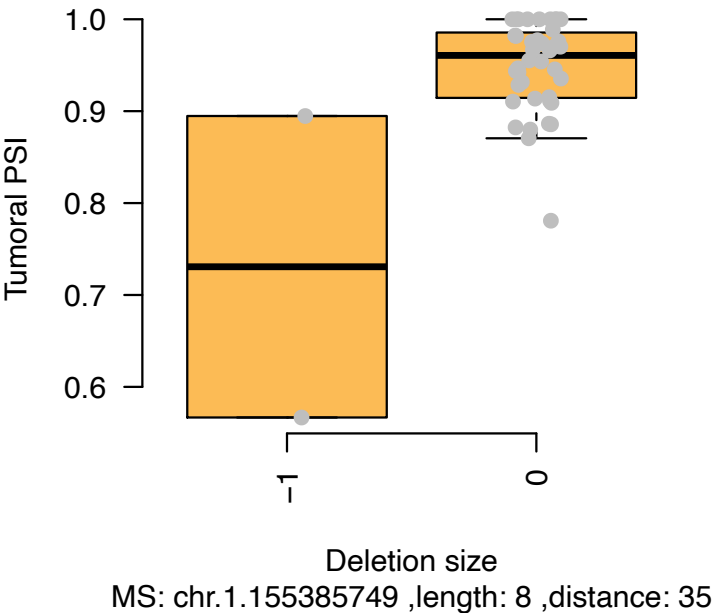


gene: C21orf66 ,exon: 43 ,type: Alternative

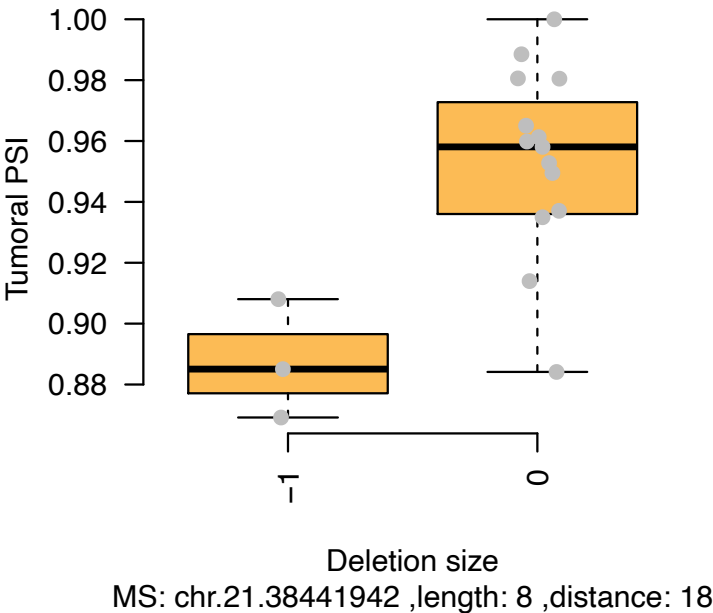


Supplementary Figure S6

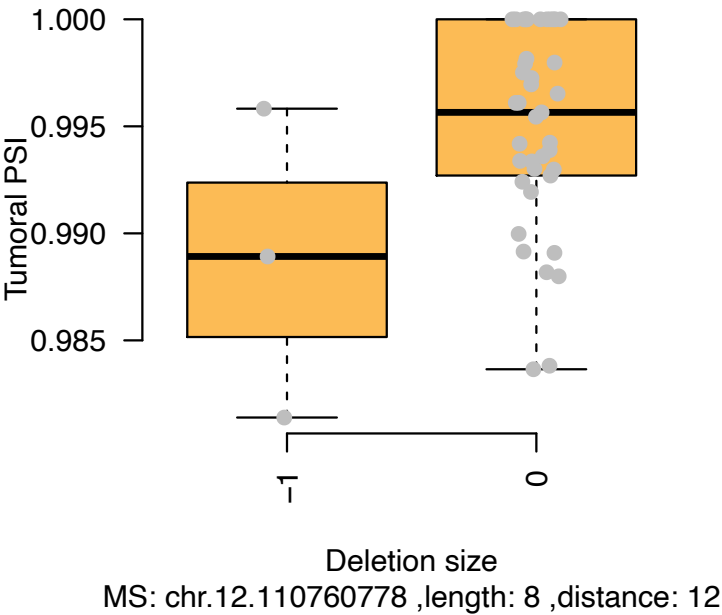
gene: **ASH1L** ,exon: 113 ,type: Alternative



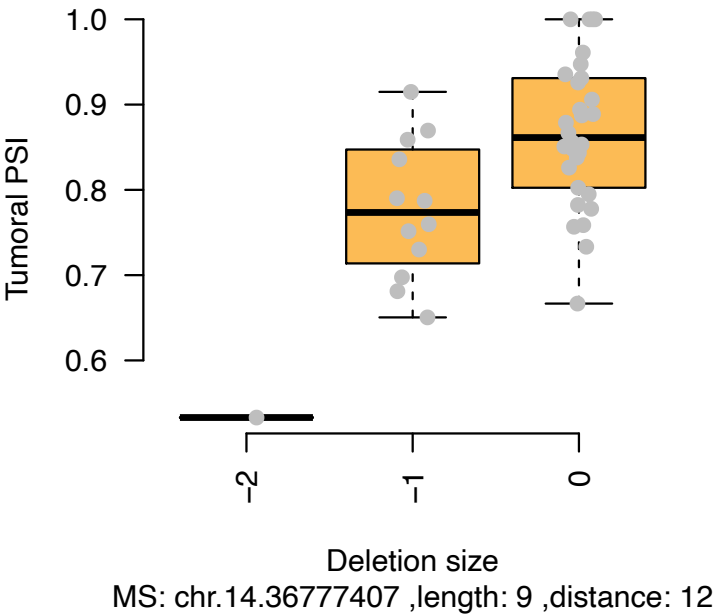
gene: **PIGP** ,exon: 69 ,type: Alternative



gene: **ATP2A2** ,exon: 113 ,type: Alternative

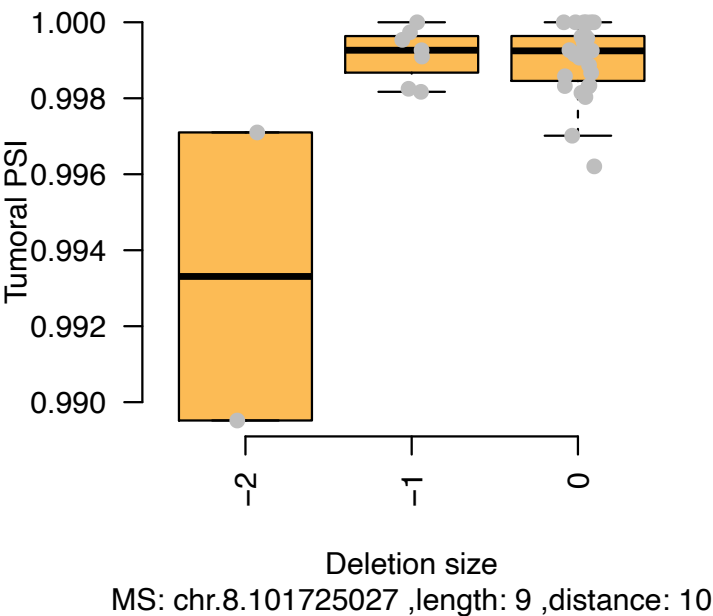


gene: **MBIP** ,exon: 124 ,type: Alternative

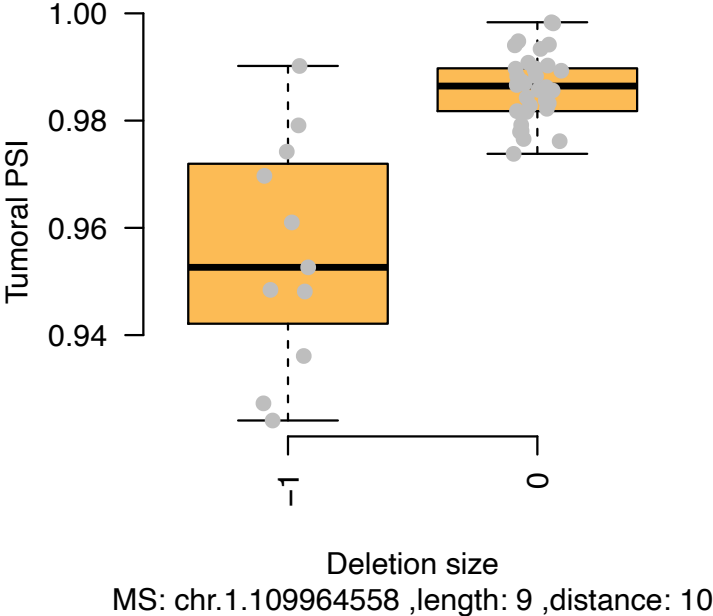


Supplementary Figure S6

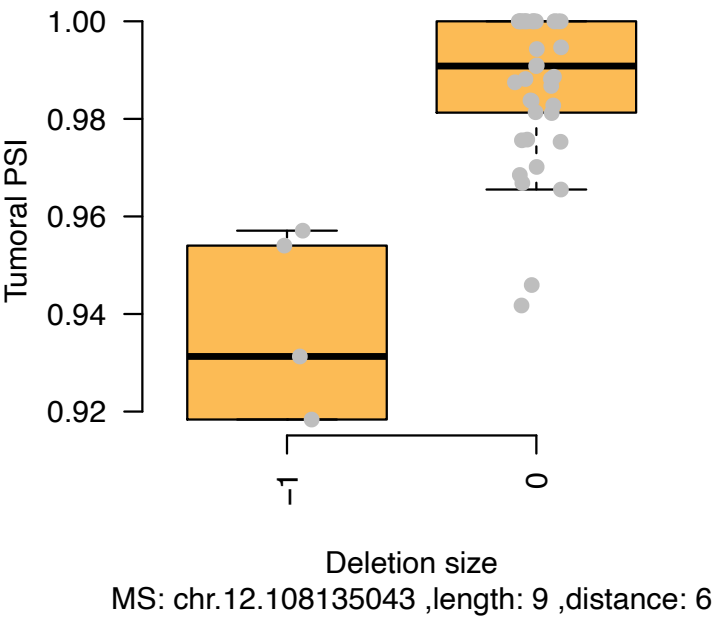
gene: PABPC1 ,exon: 113 ,type: Alternative



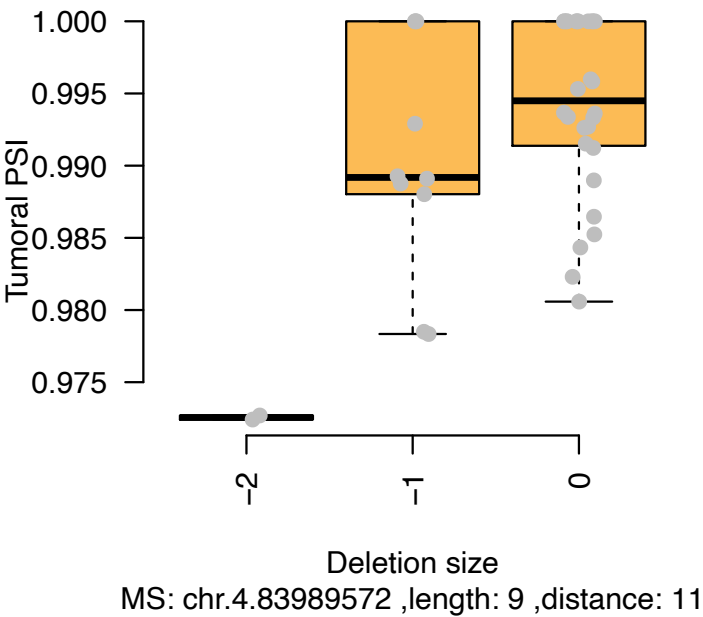
gene: PSMA5 ,exon: 69 ,type: Alternative



gene: PRDM4 ,exon: 1 ,type: Alternative

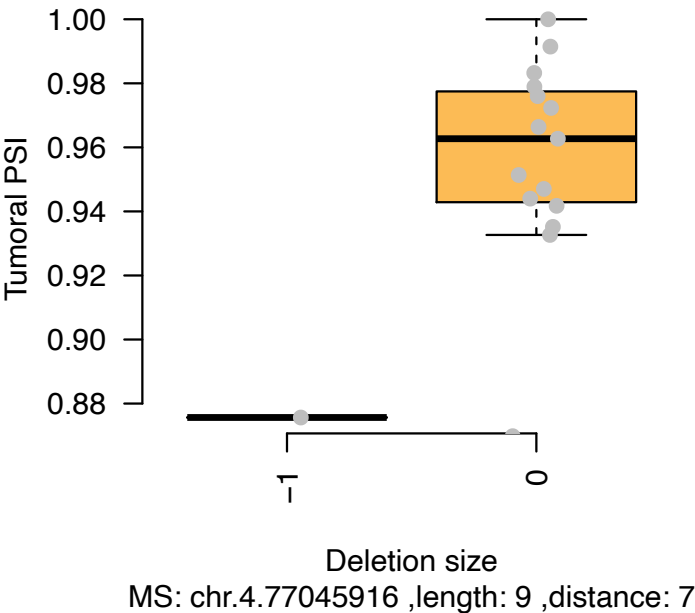


gene: COPS4 ,exon: 146 ,type: Alternative

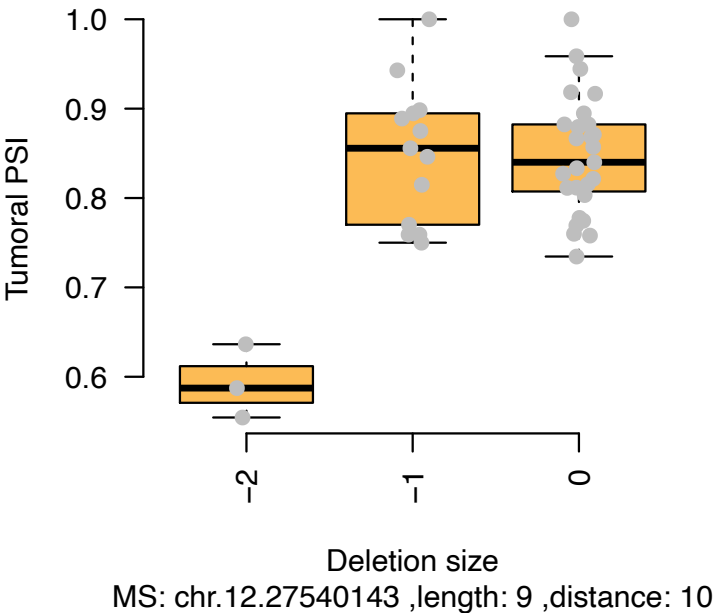


Supplementary Figure S6

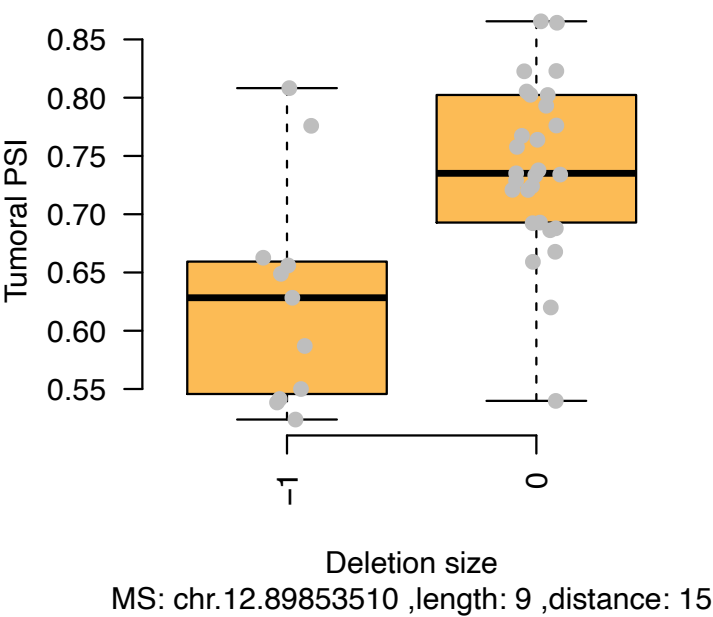
gene: NUP54 ,exon: 146 ,type: Alternative



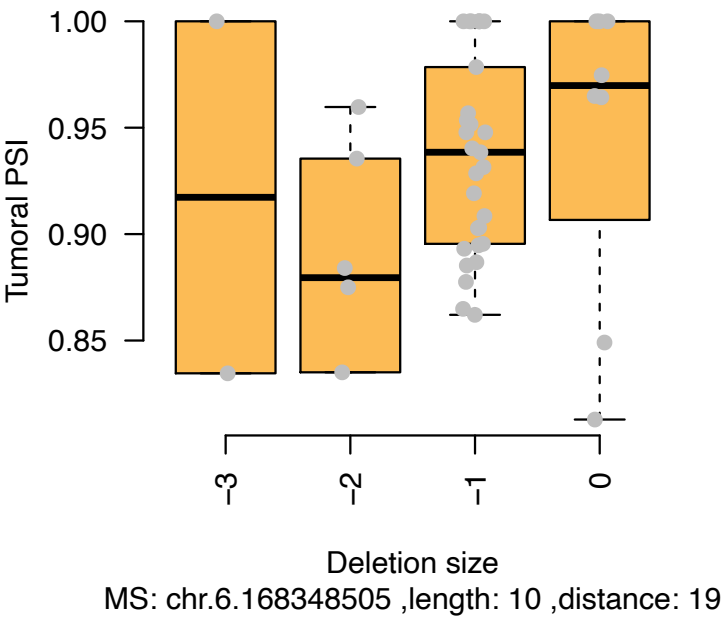
gene: ARNTL2 ,exon: 124 ,type: Alternative



gene: WDR51B ,exon: 1 ,type: Alternative

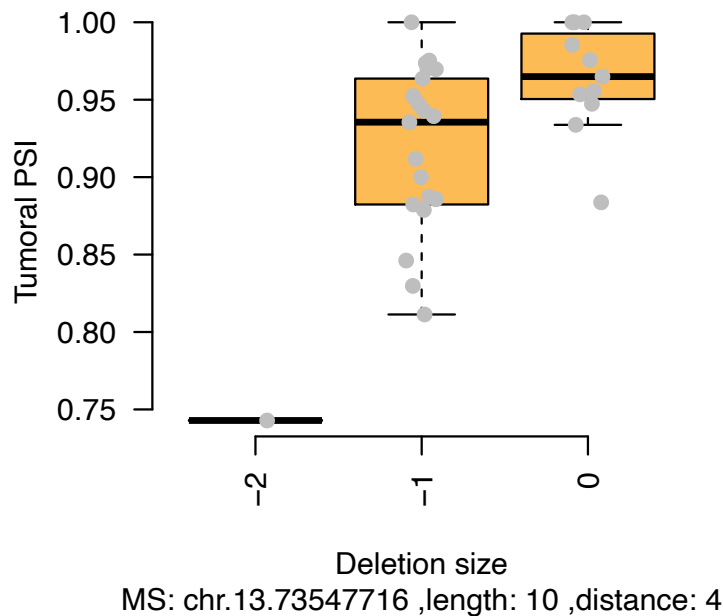


gene: MLLT4 ,exon: 77 ,type: Alternative

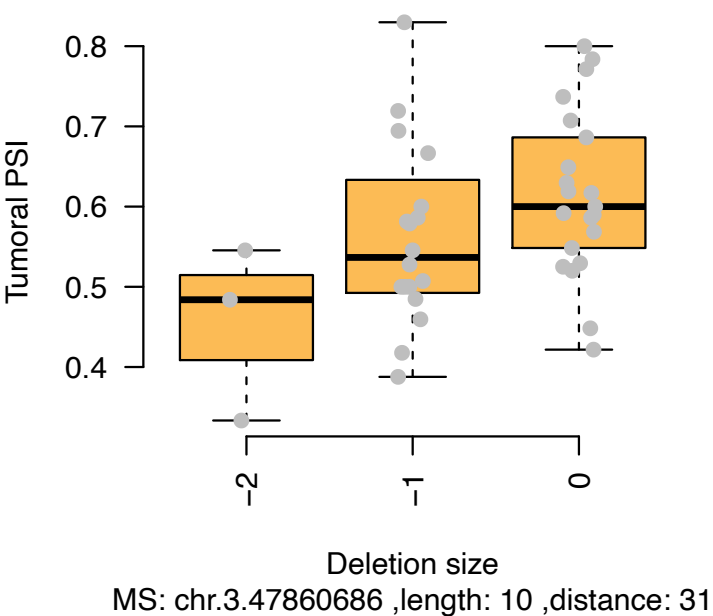


Supplementary Figure S6

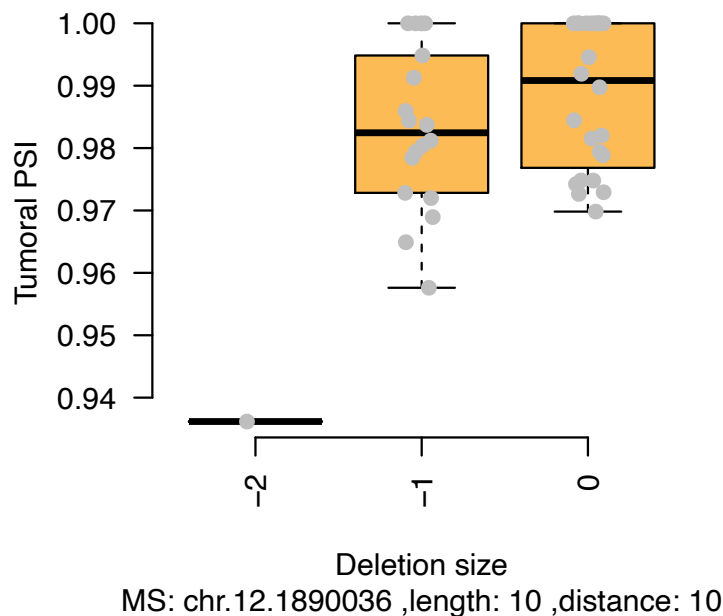
gene: PIBF1 ,exon: 48 ,type: Alternative



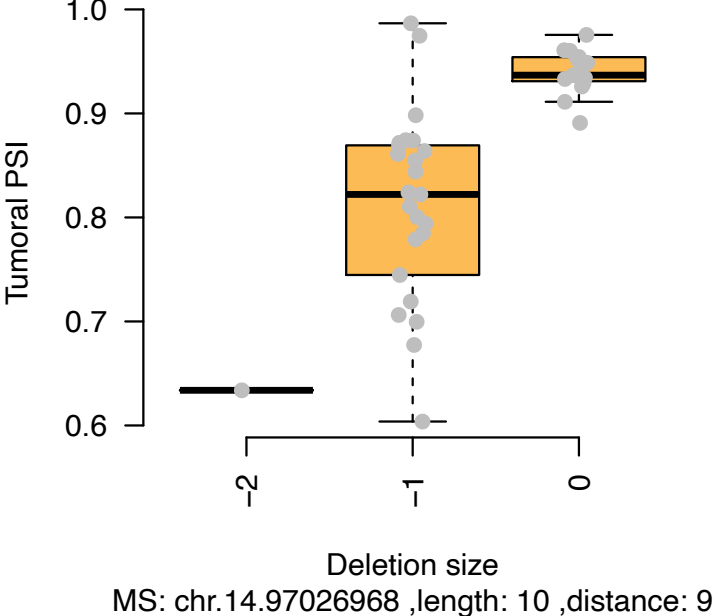
gene: DHX30 ,exon: 102 ,type: Alternative



gene: ADIPOR2 ,exon: 113 ,type: Alternative

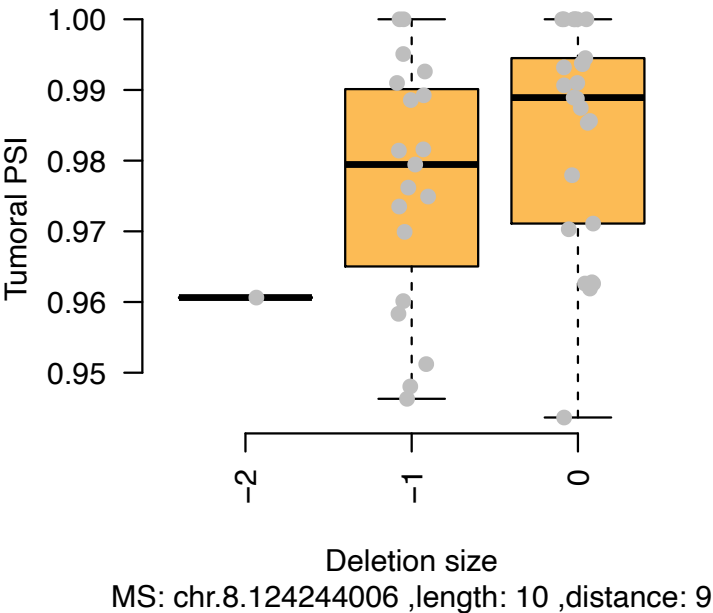


gene: PAPOLA ,exon: 70 ,type: Alternative

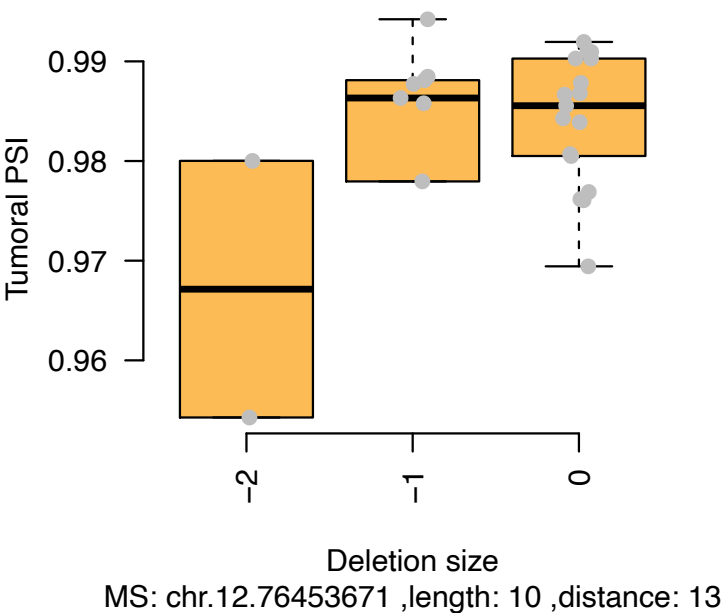


Supplementary Figure S6

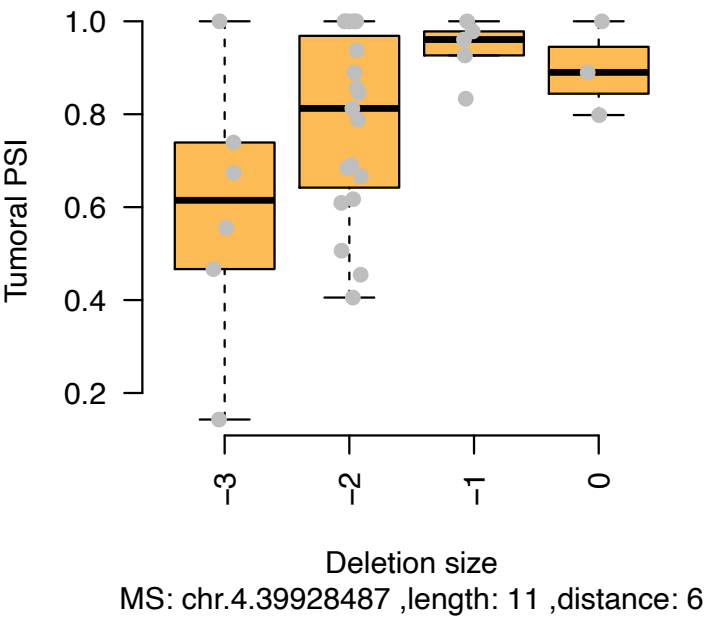
gene: C8orf76 ,exon: 91 ,type: Alternative



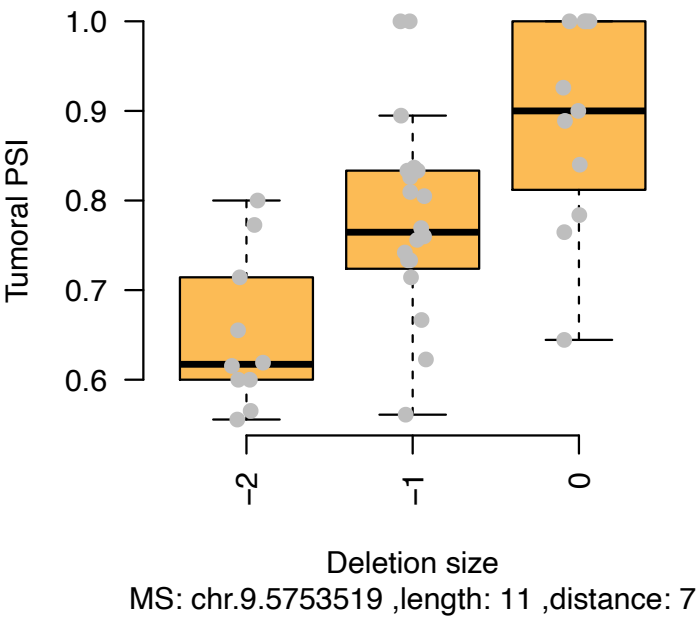
gene: NAP1L1 ,exon: 113 ,type: Alternative



gene: PDS5A ,exon: 91 ,type: Alternative

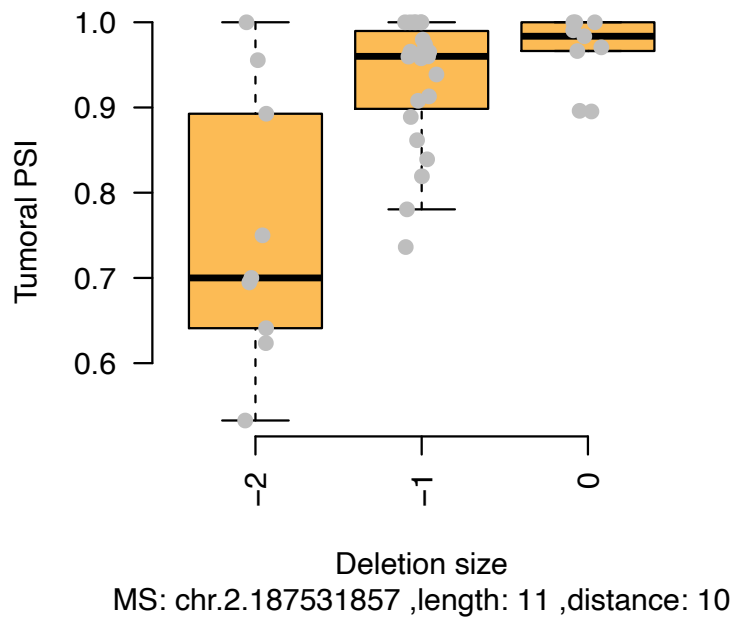


gene: KIAA1432 ,exon: 35 ,type: Alternative

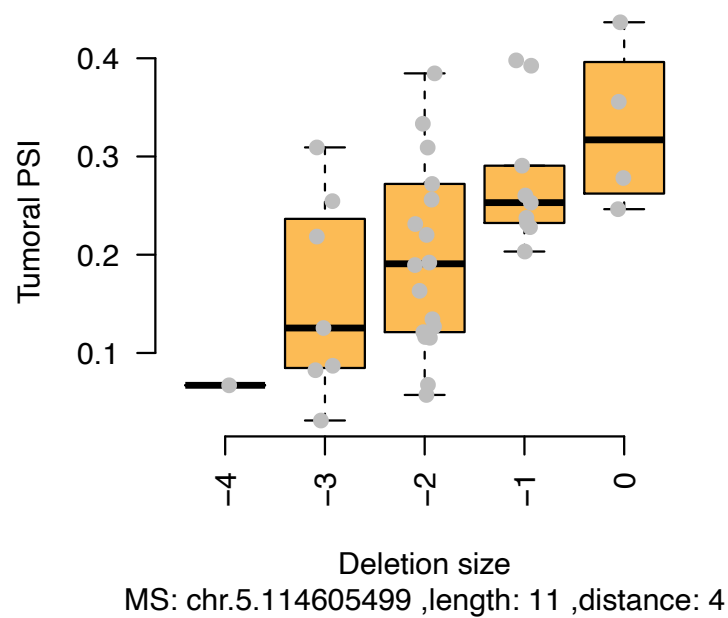


Supplementary Figure S6

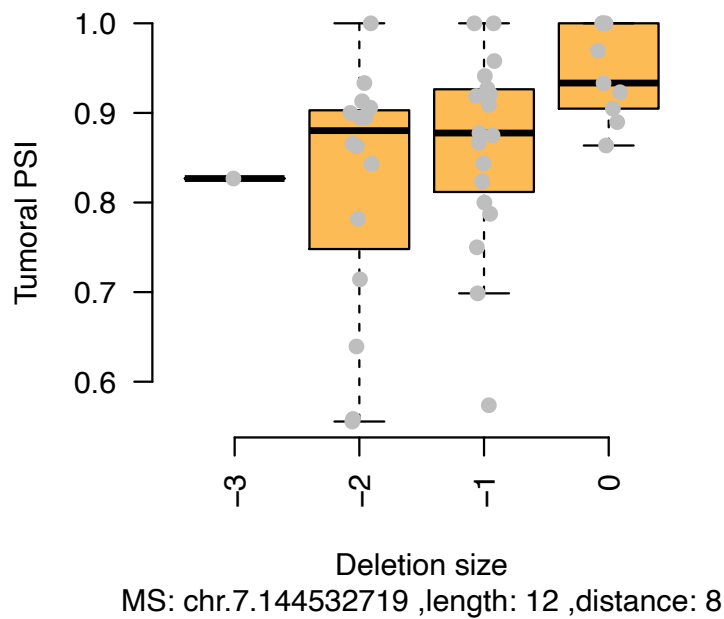
gene: ITGAV ,exon: 71 ,type: Alternative



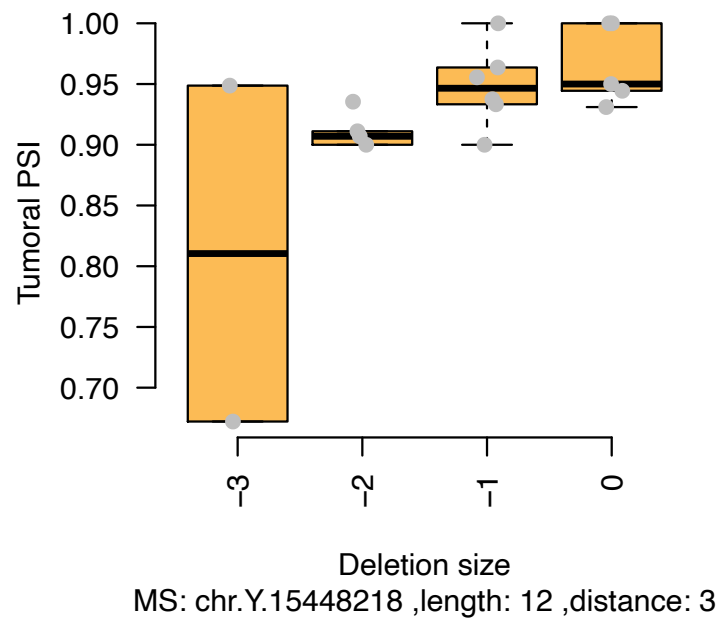
gene: CCDC112 ,exon: 135 ,type: Alternative



gene: TPK1 ,exon: 69 ,type: Alternative

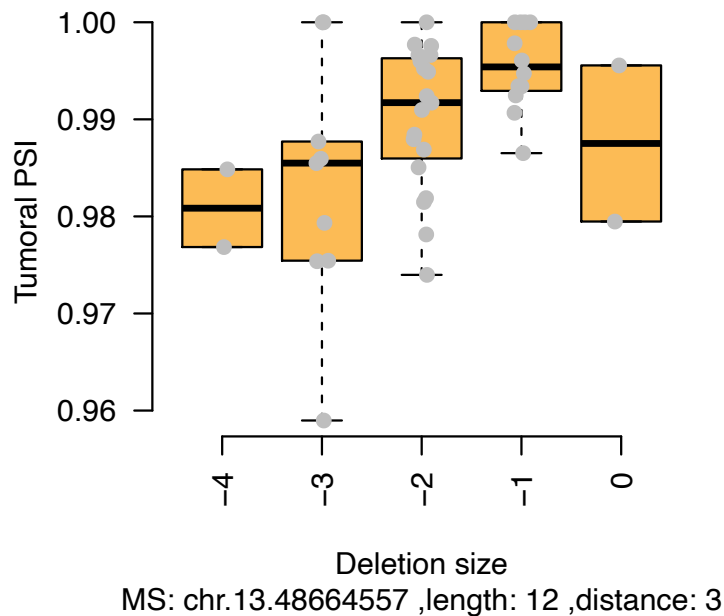


gene: UTY ,exon: 48 ,type: Alternative

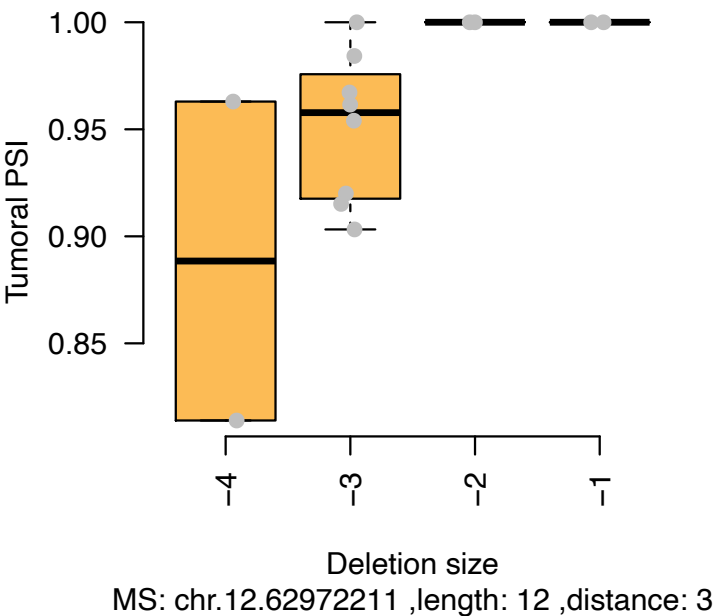


Supplementary Figure S6

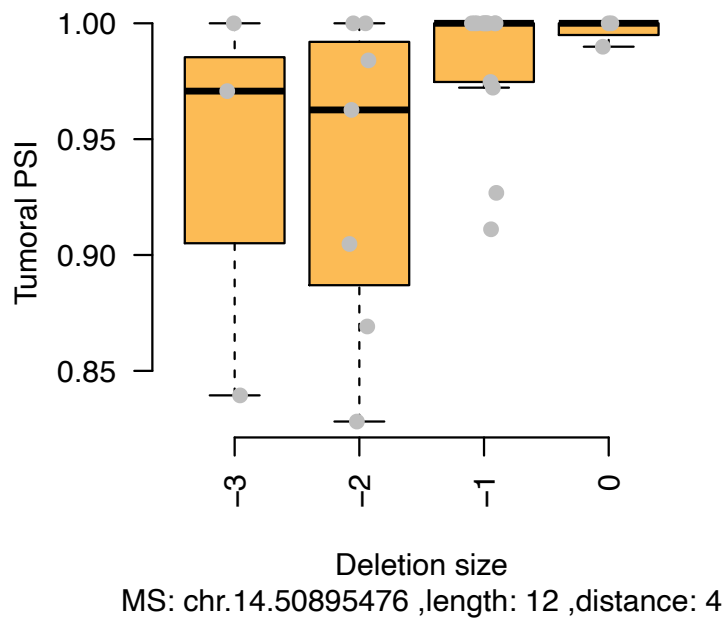
gene: MED4 ,exon: 69 ,type: Alternative



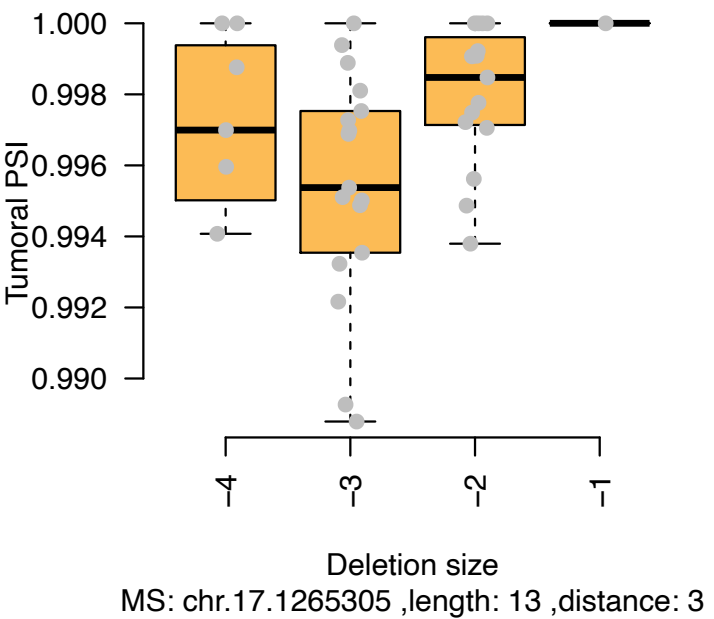
gene: MON2 ,exon: 82 ,type: Alternative



gene: MAP4K5 ,exon: 81 ,type: Alternative

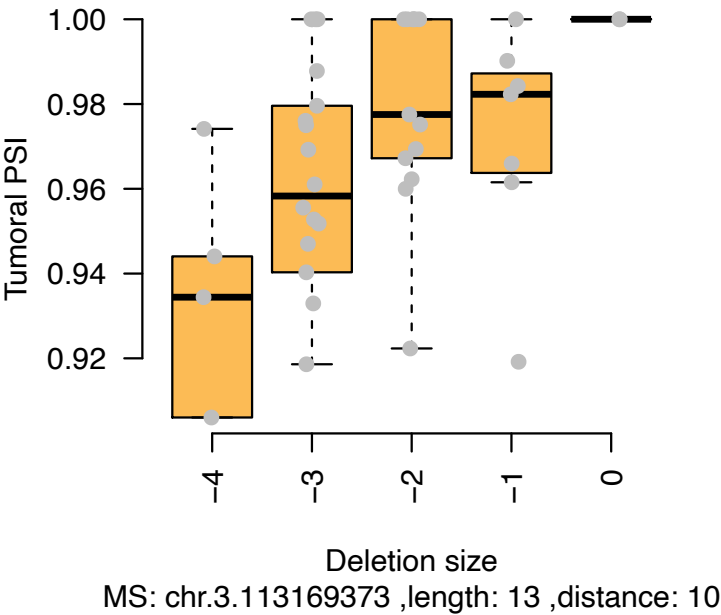


gene: YWHAE ,exon: 80 ,type: Alternative

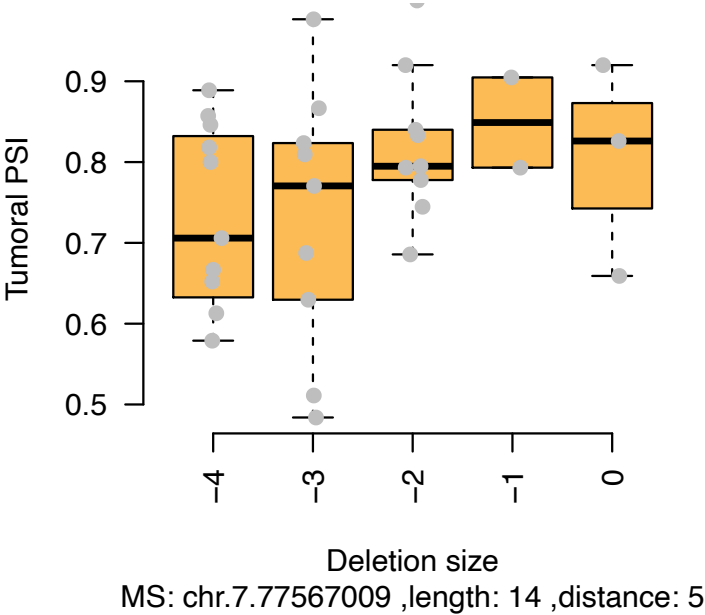


Supplementary Figure S6

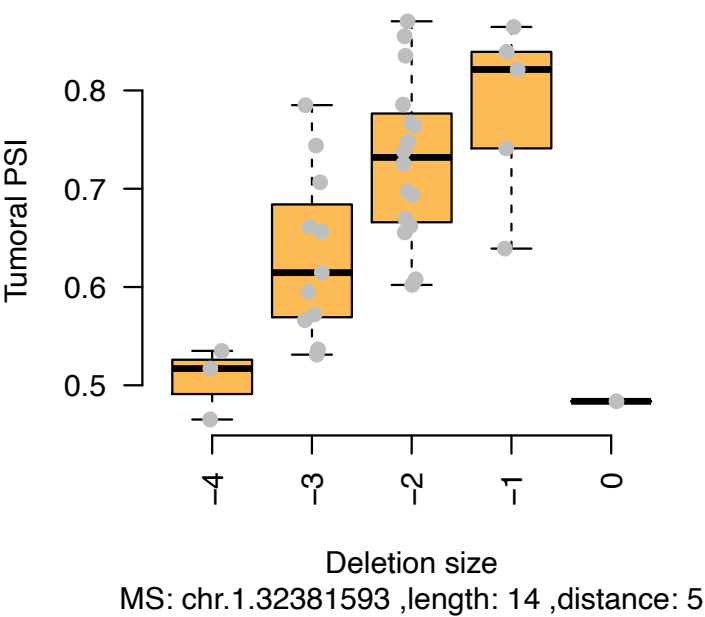
gene: **CCDC52** ,exon: 43 ,type: Alternative



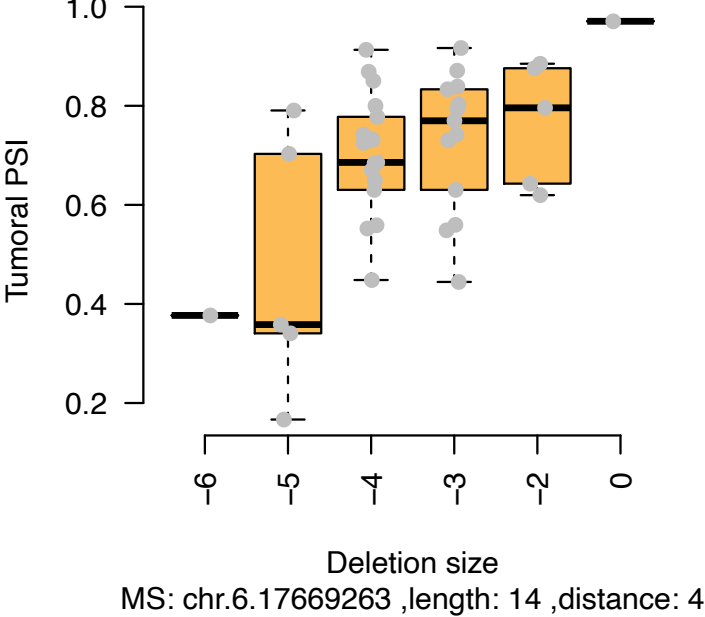
gene: **PHTF2** ,exon: 12 ,type: Alternative



gene: **PTP4A2** ,exon: 69 ,type: Alternative

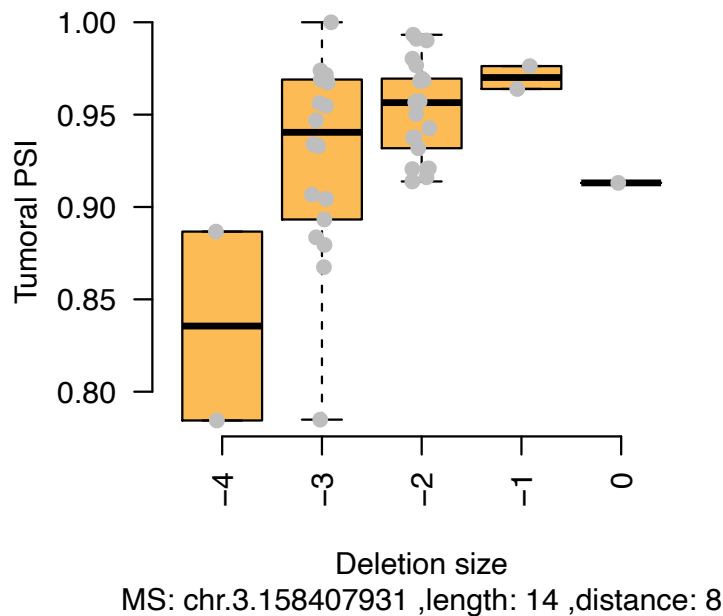


gene: **NUP153** ,exon: 135 ,type: Alternative

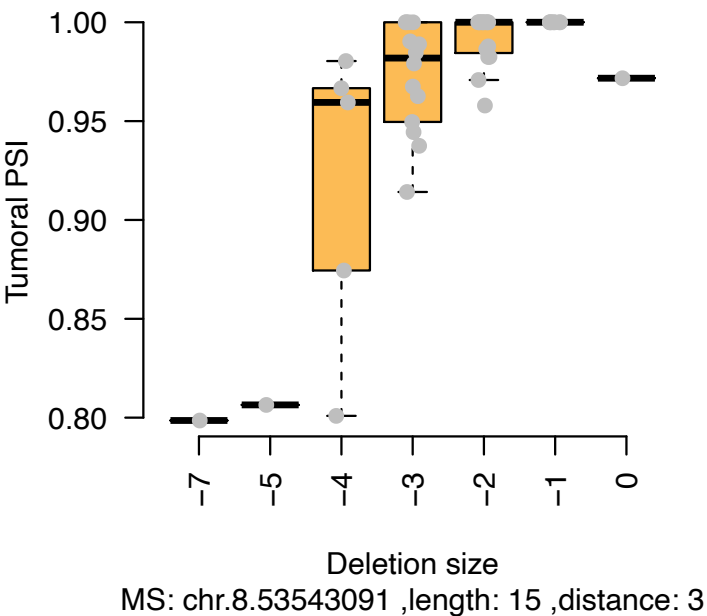


Supplementary Figure S6

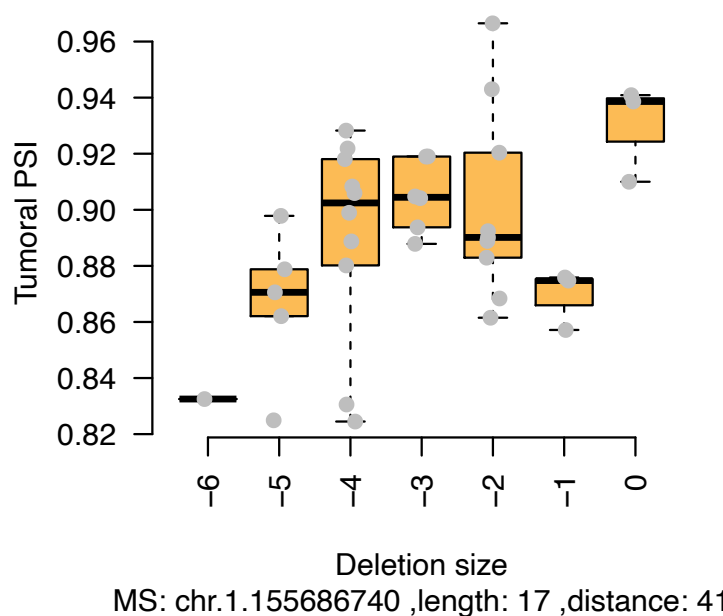
gene: GFM1 ,exon: 48 ,type: Alternative



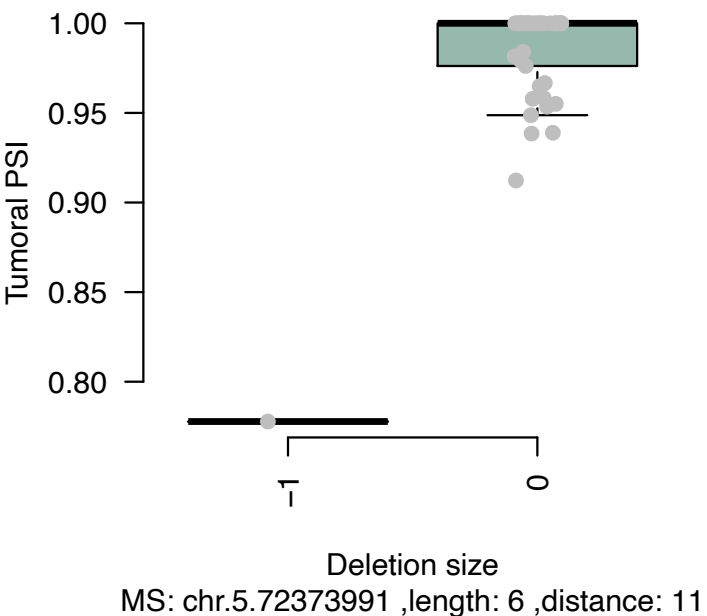
gene: RB1CC1 ,exon: 71 ,type: Alternative



gene: DAP3 ,exon: 80 ,type: Alternative

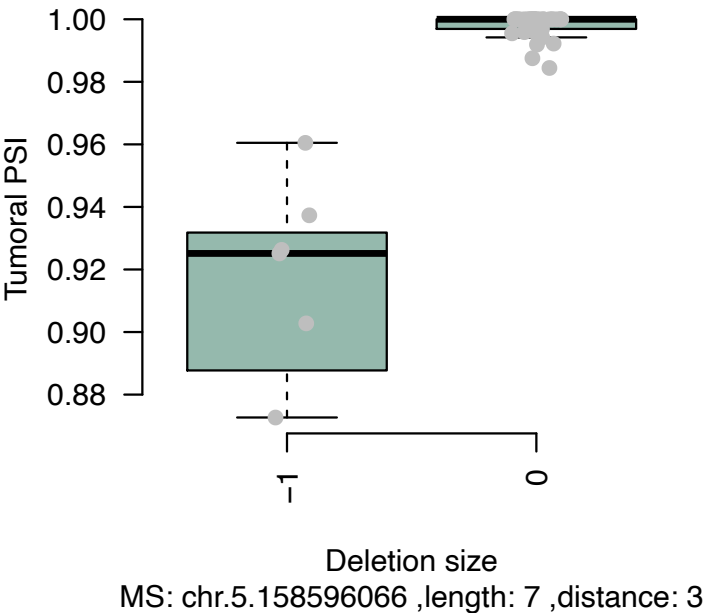


gene: FCHO2 ,exon: 71 ,type: Constitutive

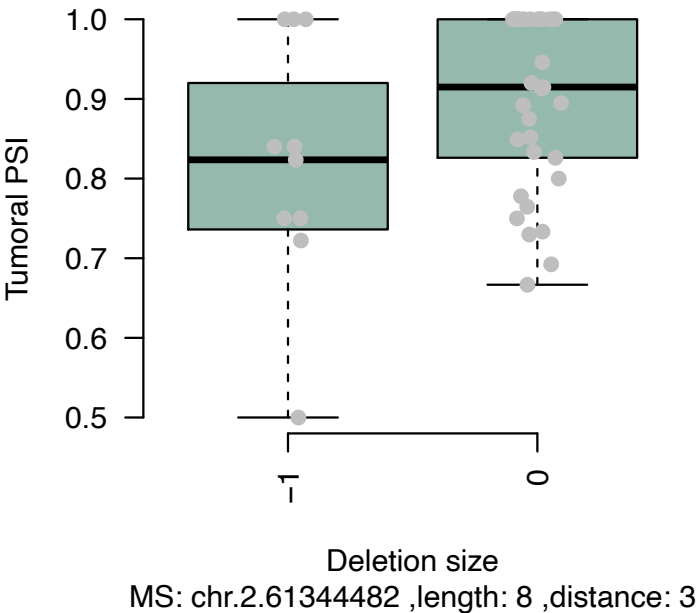


Supplementary Figure S6

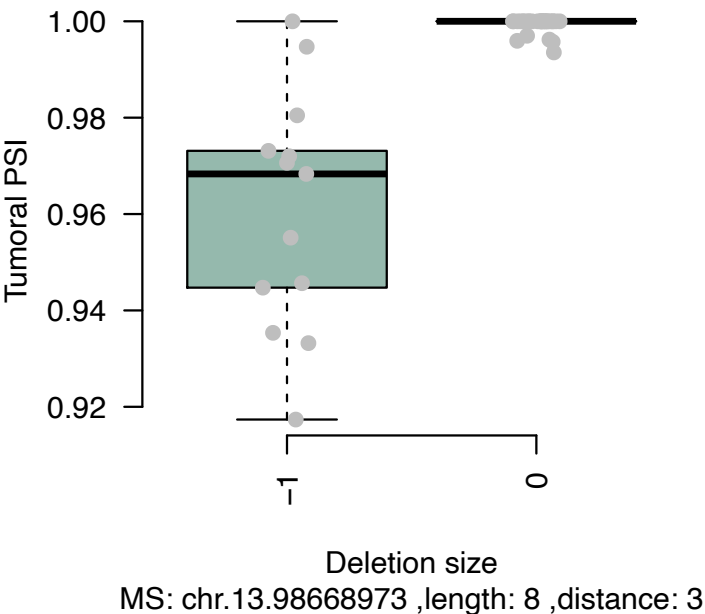
gene: RNF145 ,exon: 135 ,type: Constitutive



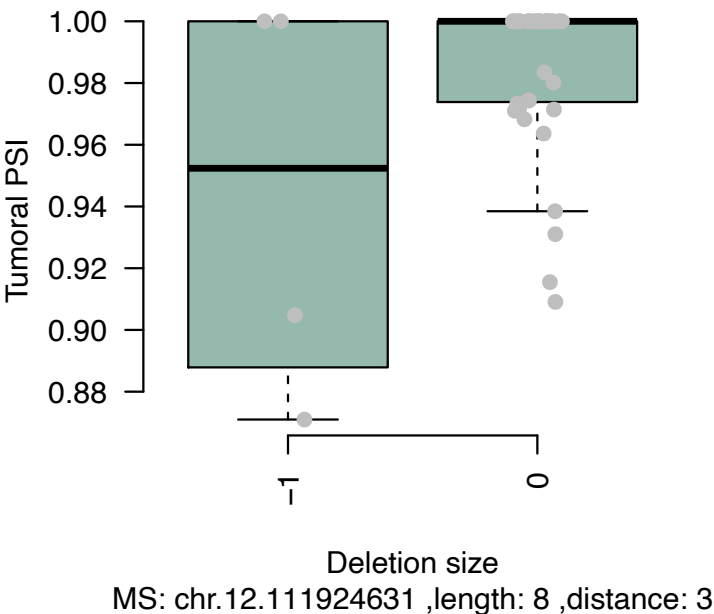
gene: KIAA1841 ,exon: 60 ,type: Constitutive



gene: IPO5 ,exon: 75 ,type: Constitutive

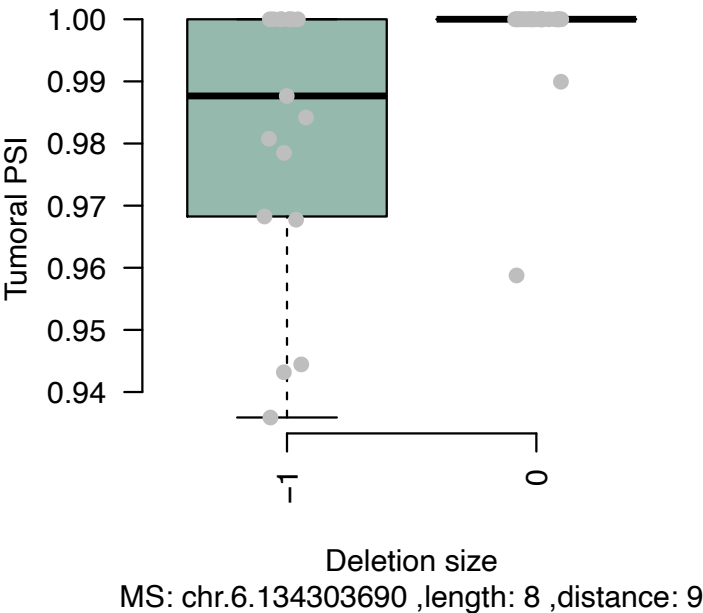


gene: ATXN2 ,exon: 48 ,type: Constitutive

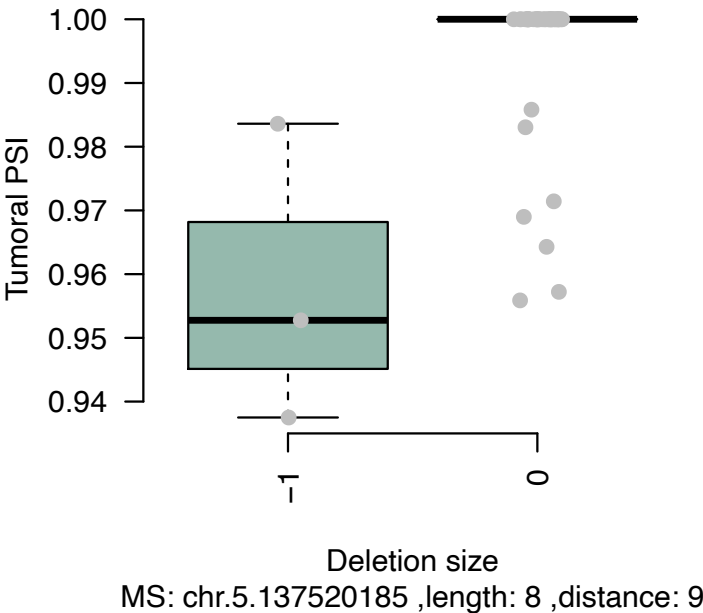


Supplementary Figure S6

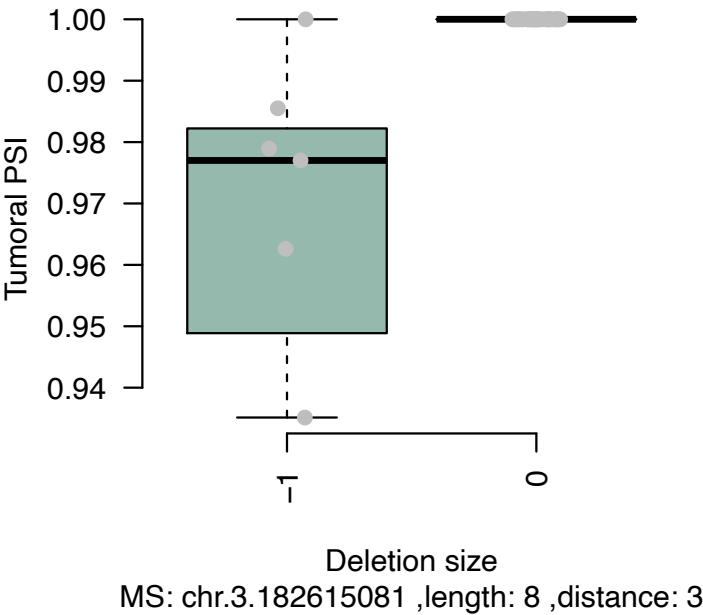
gene: **TBPL1** ,exon: 80 ,type: Constitutive



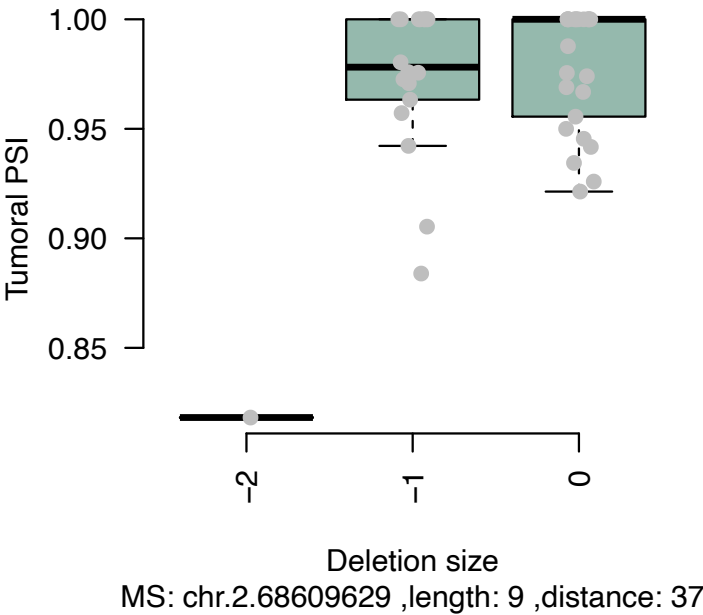
gene: **KIF20A** ,exon: 30 ,type: Constitutive



gene: **ATP11B** ,exon: 77 ,type: Constitutive

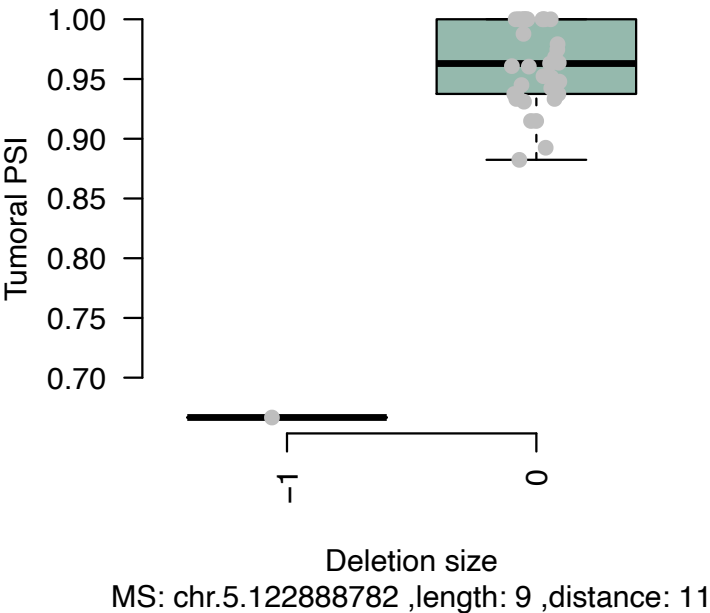


gene: **PLEK** ,exon: 91 ,type: Constitutive

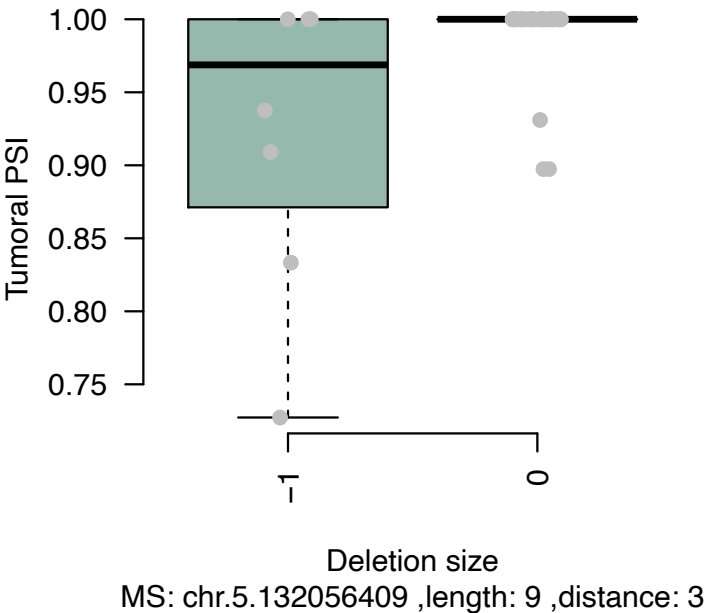


Supplementary Figure S6

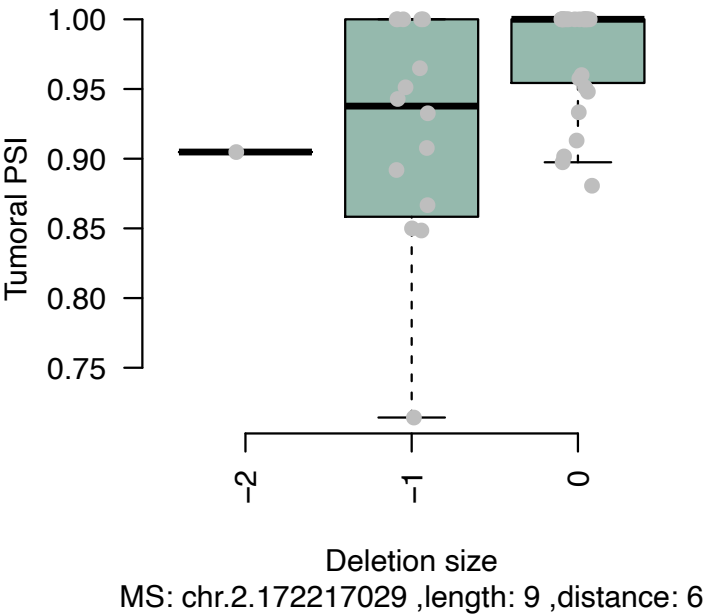
gene: CSNK1G3 ,exon: 80 ,type: Constitutive



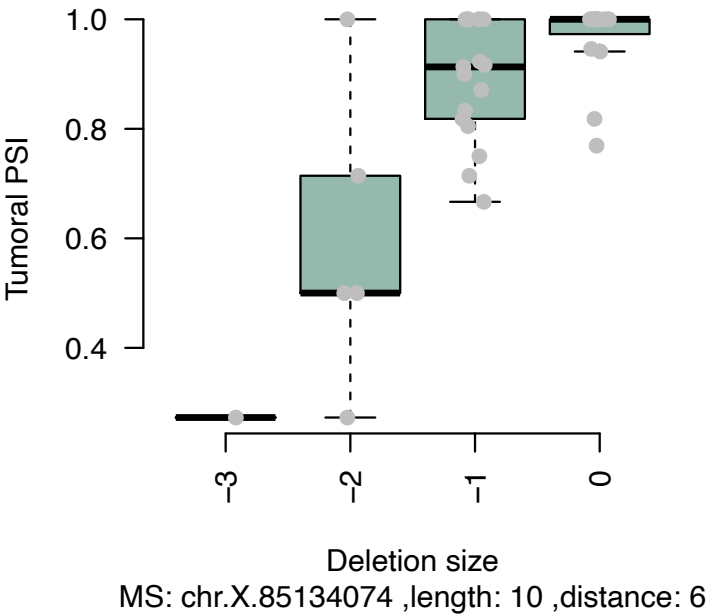
gene: KIF3A ,exon: 102 ,type: Constitutive



gene: METTL8 ,exon: 80 ,type: Constitutive

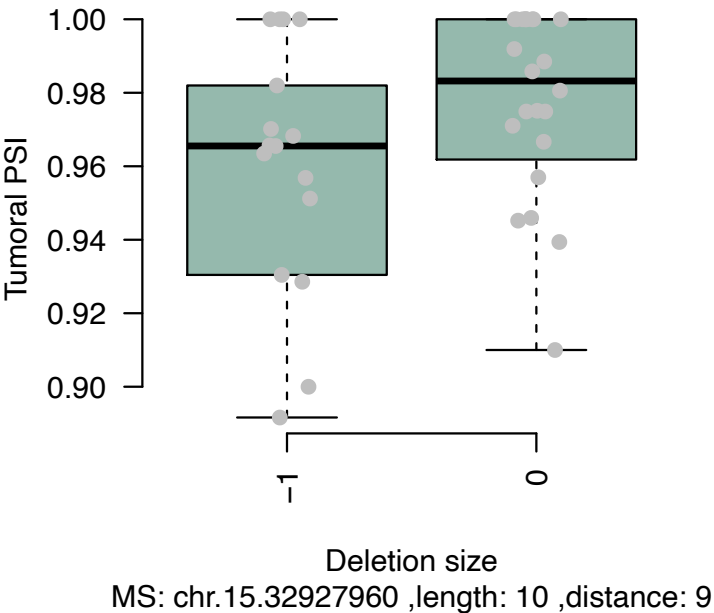


gene: CHM ,exon: 30 ,type: Constitutive

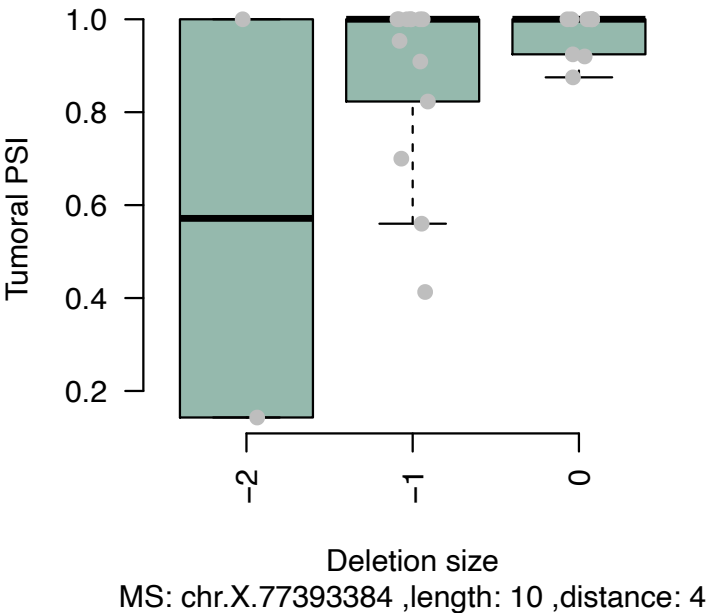


Supplementary Figure S6

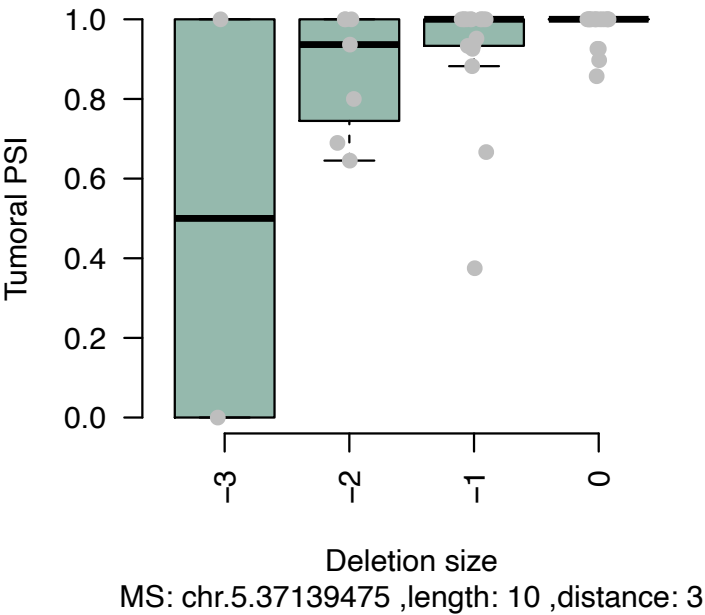
gene: ARHGAP11A ,exon: 12 ,type: Constitutive



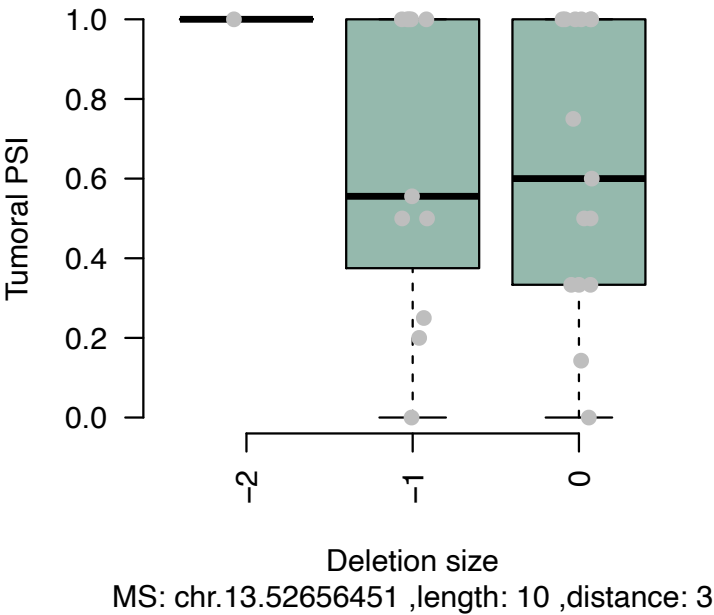
gene: TAF9B ,exon: 91 ,type: Constitutive



gene: C5orf42 ,exon: 96 ,type: Constitutive

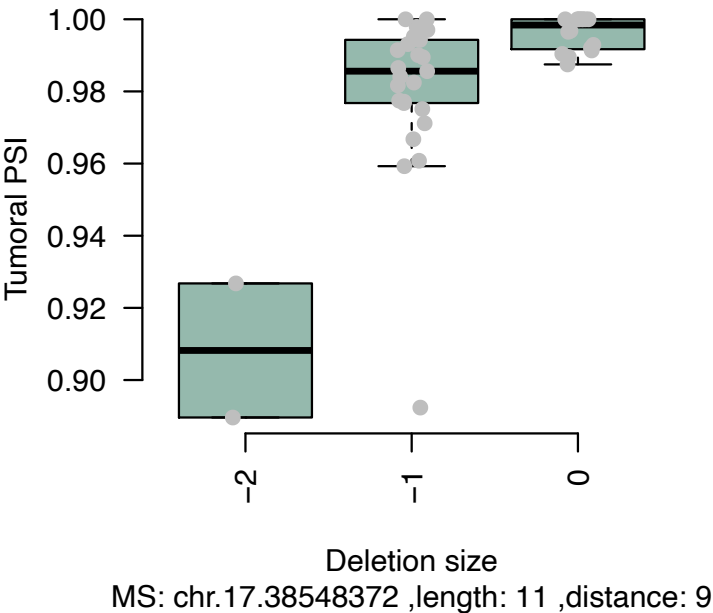


gene: NEK5 ,exon: 60 ,type: Constitutive

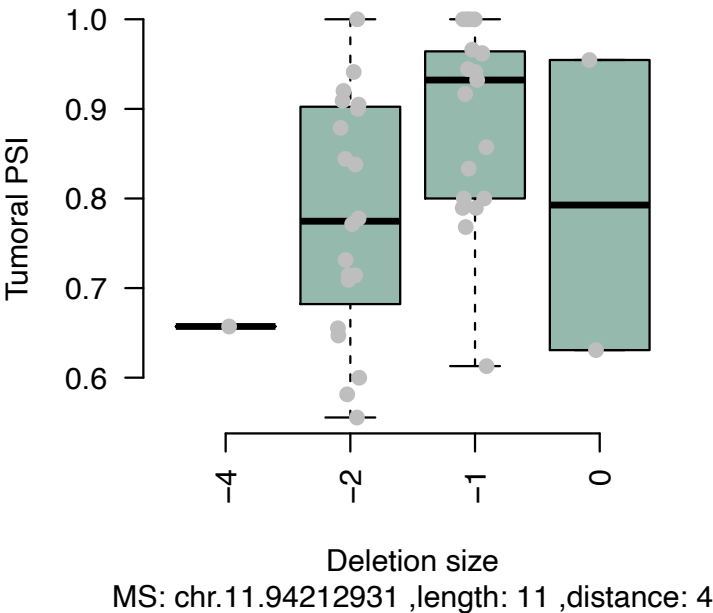


Supplementary Figure S6

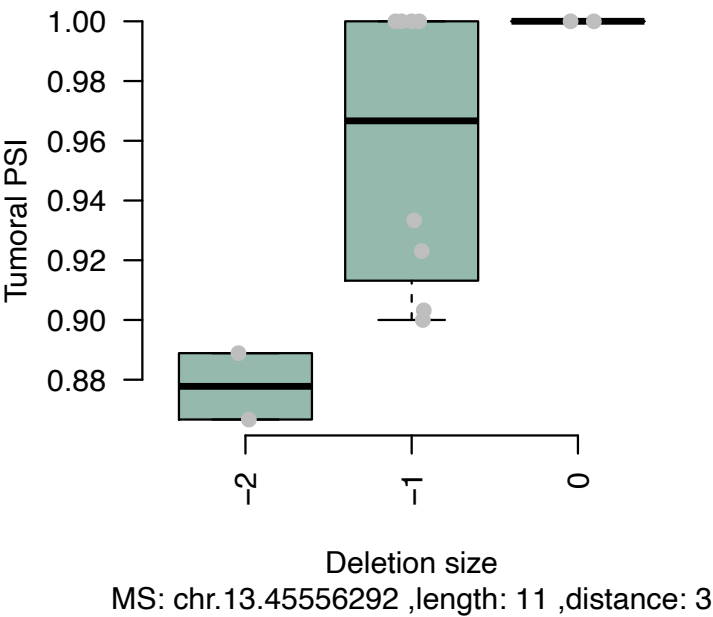
gene: TOP2A ,exon: 83 ,type: Constitutive



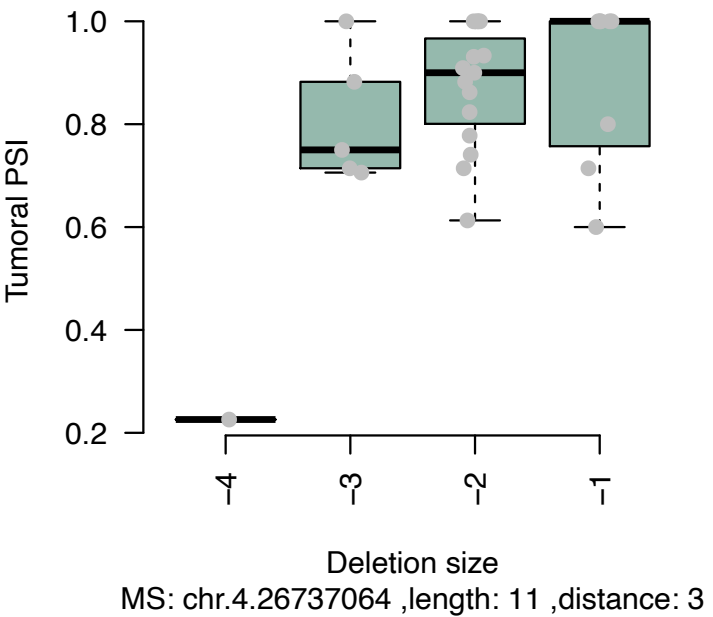
gene: MRE11A ,exon: 102 ,type: Constitutive



gene: NUFIP1 ,exon: 69 ,type: Constitutive

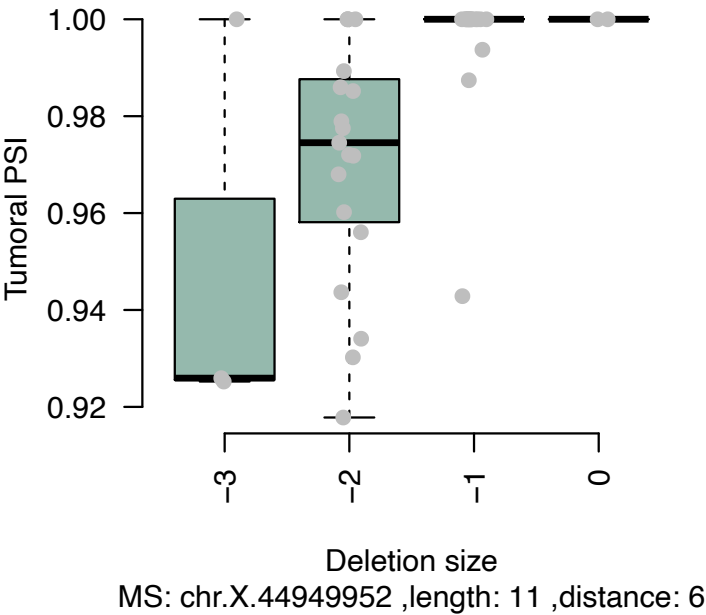


gene: TBC1D19 ,exon: 48 ,type: Constitutive

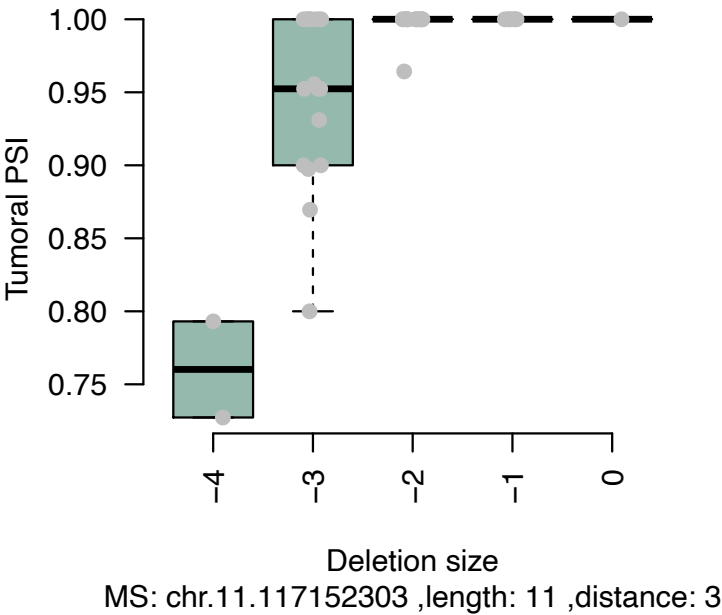


Supplementary Figure S6

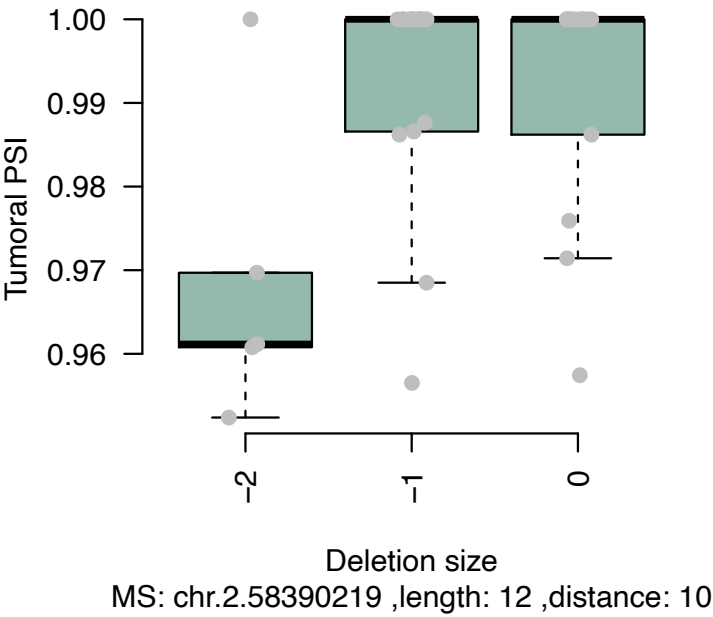
gene: KDM6A ,exon: 76 ,type: Constitutive



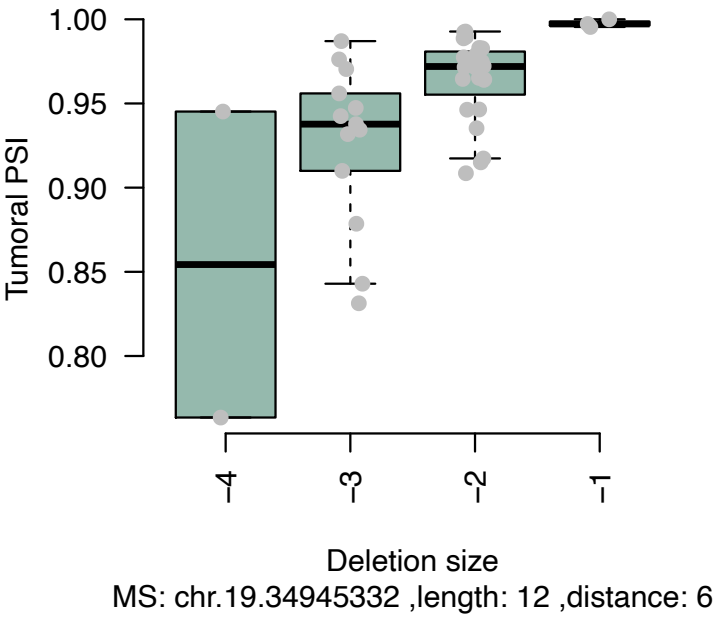
gene: RNF214 ,exon: 1 ,type: Constitutive



gene: FANCL ,exon: 1 ,type: Constitutive

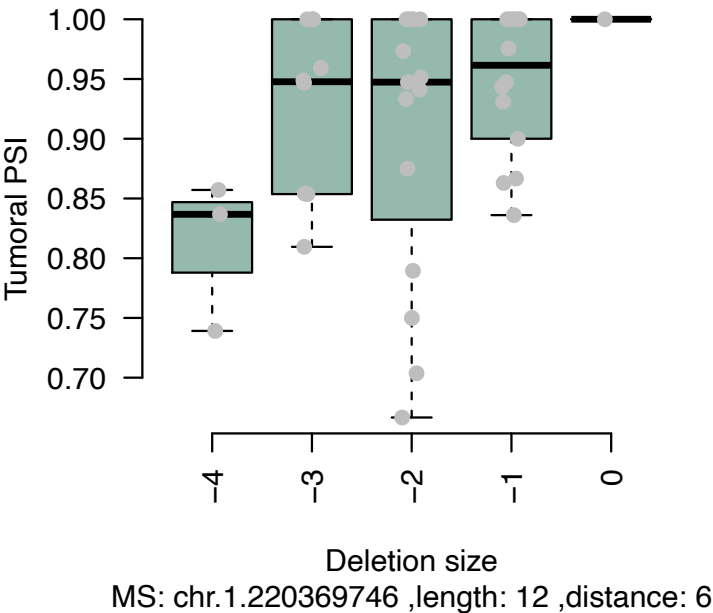


gene: UBA2 ,exon: 23 ,type: Constitutive

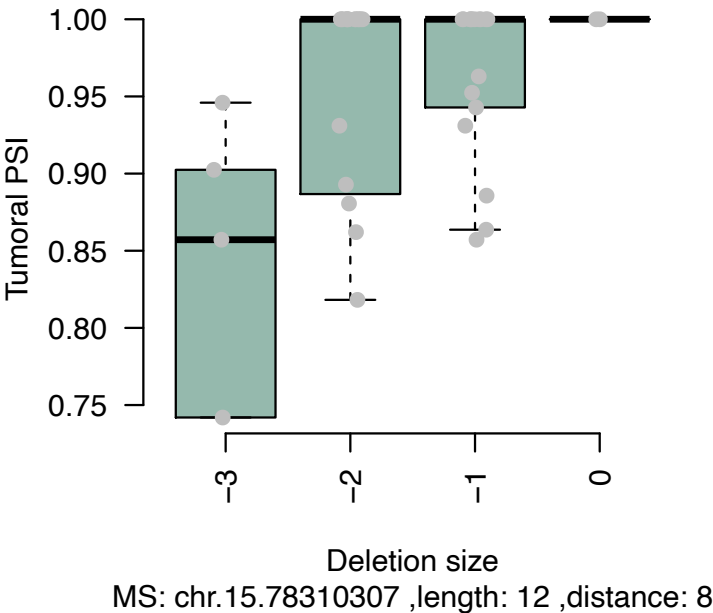


Supplementary Figure S6

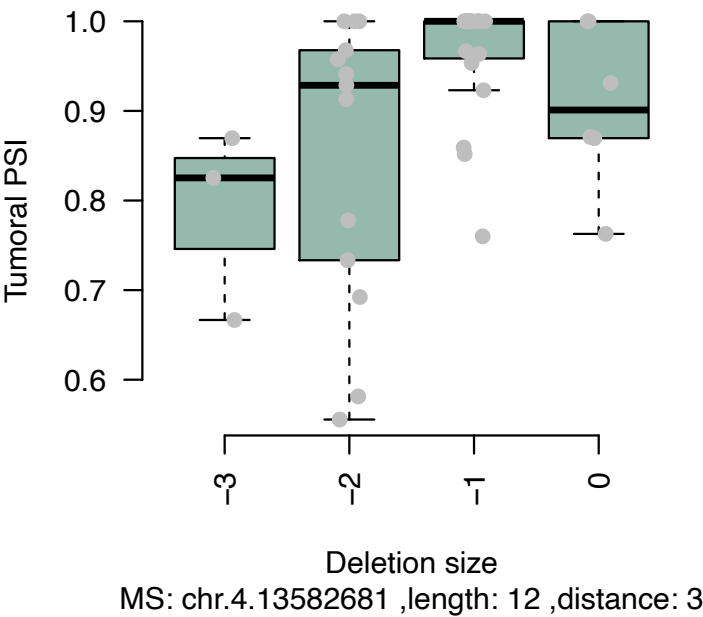
gene: RAB3GAP2 ,exon: 1 ,type: Constitutive



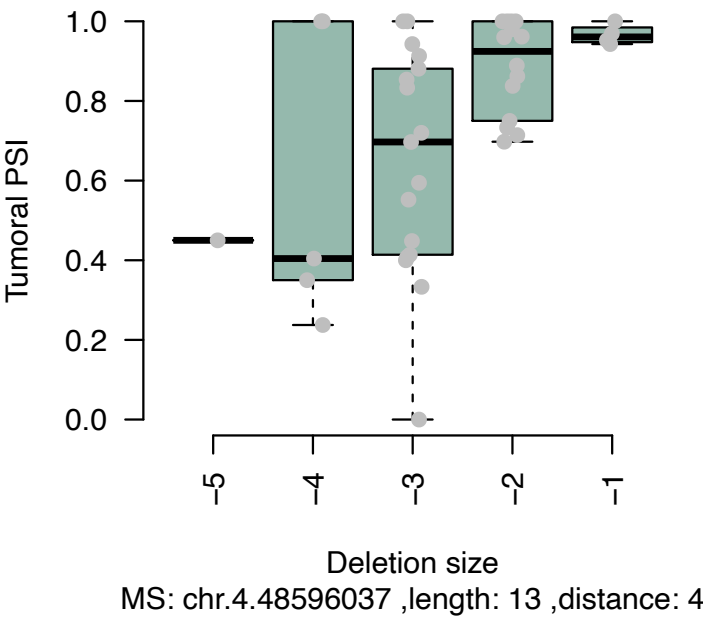
gene: TBC1D2B ,exon: 124 ,type: Constitutive



gene: BOD1L ,exon: 71 ,type: Constitutive

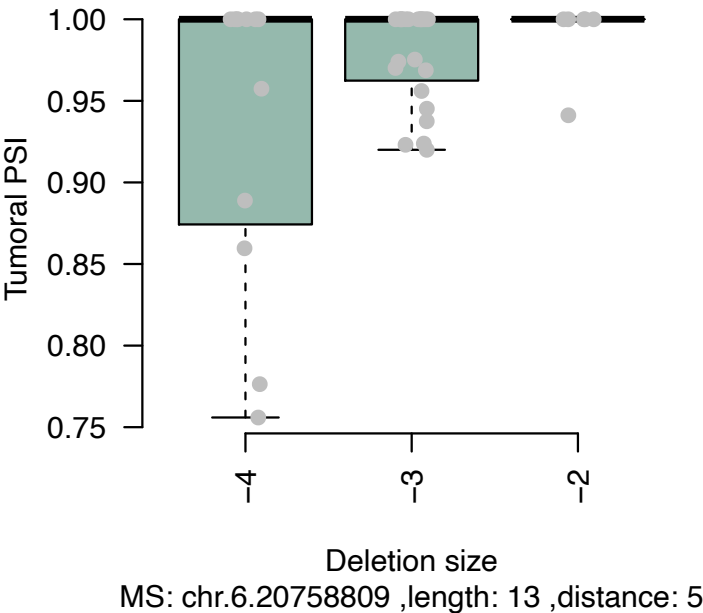


gene: FRYL ,exon: 48 ,type: Constitutive

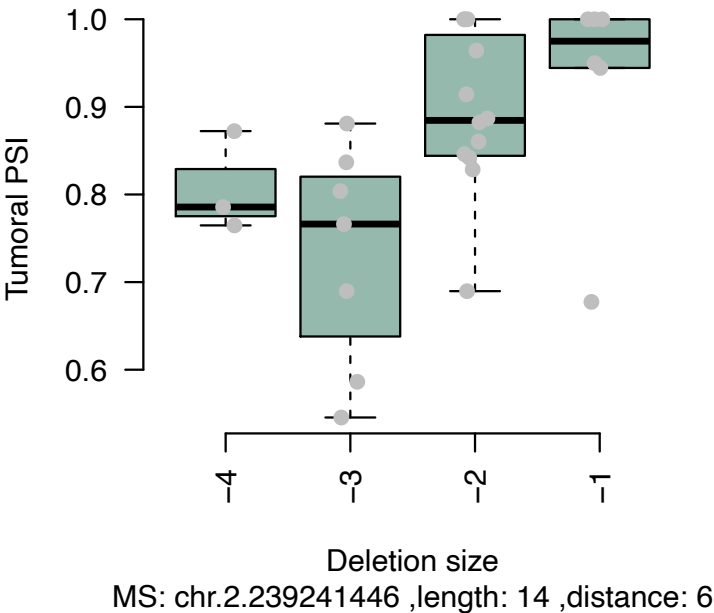


Supplementary Figure S6

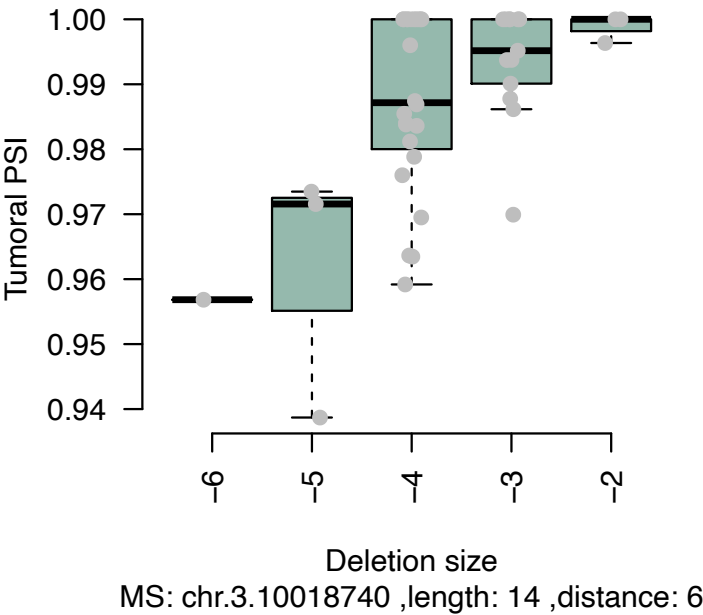
gene: CDKAL1 ,exon: 124 ,type: Constitutive



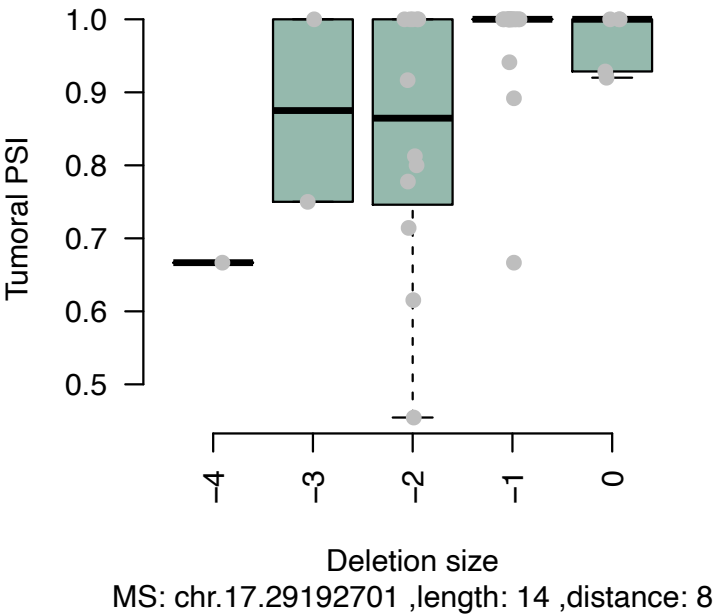
gene: TRAF3IP1 ,exon: 113 ,type: Constitutive



gene: TMEM111 ,exon: 80 ,type: Constitutive

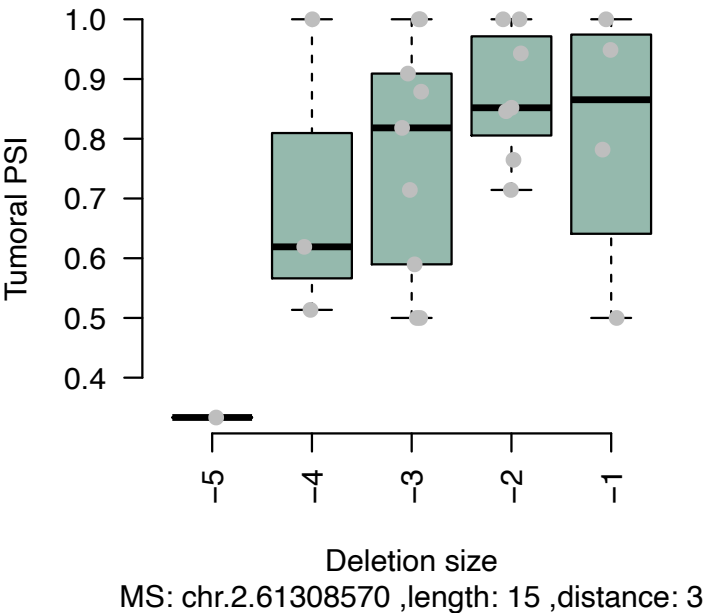


gene: ATAD5 ,exon: 12 ,type: Constitutive

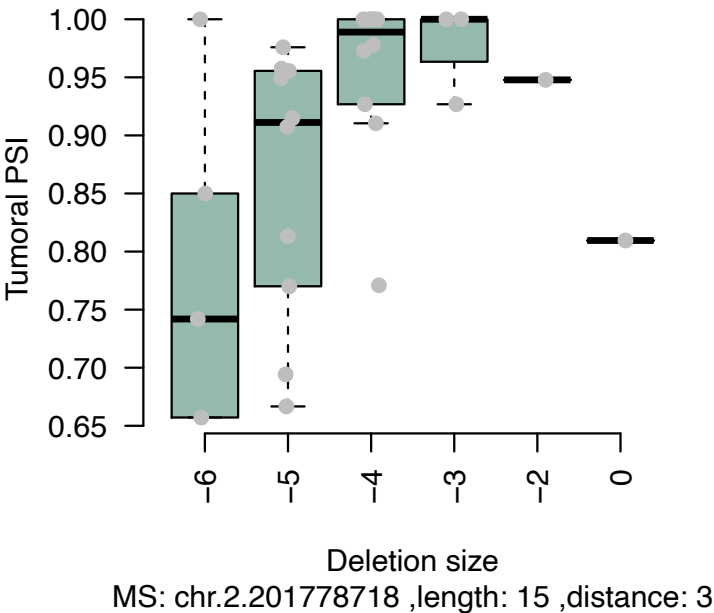


Supplementary Figure S6

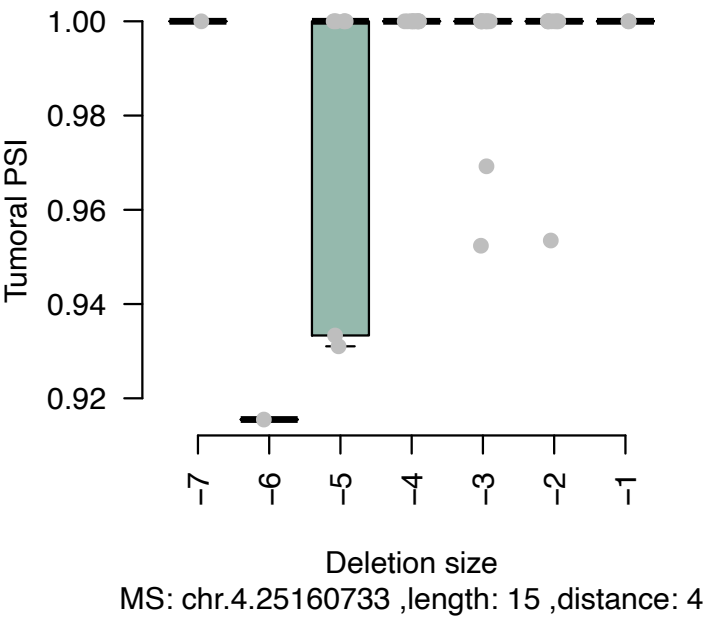
gene: KIAA1841 ,exon: 124 ,type: Constitutive



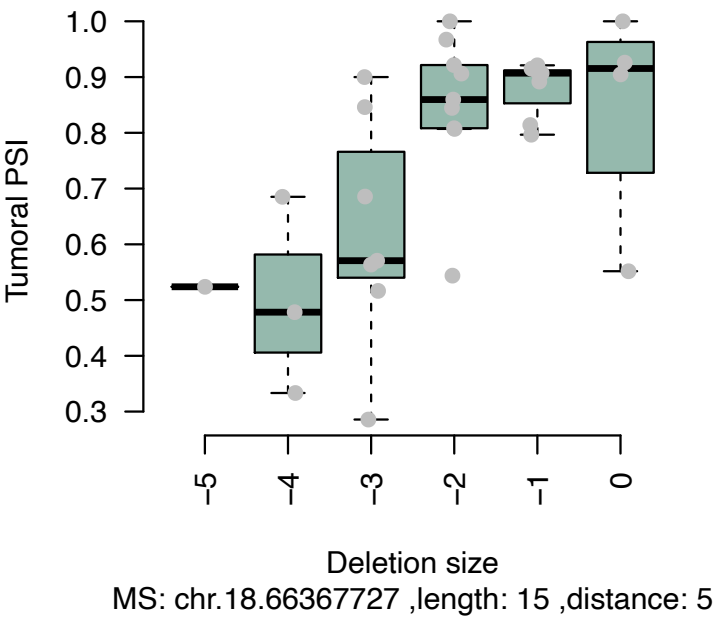
gene: ORC2L ,exon: 48 ,type: Constitutive



gene: SEPSECS ,exon: 69 ,type: Constitutive

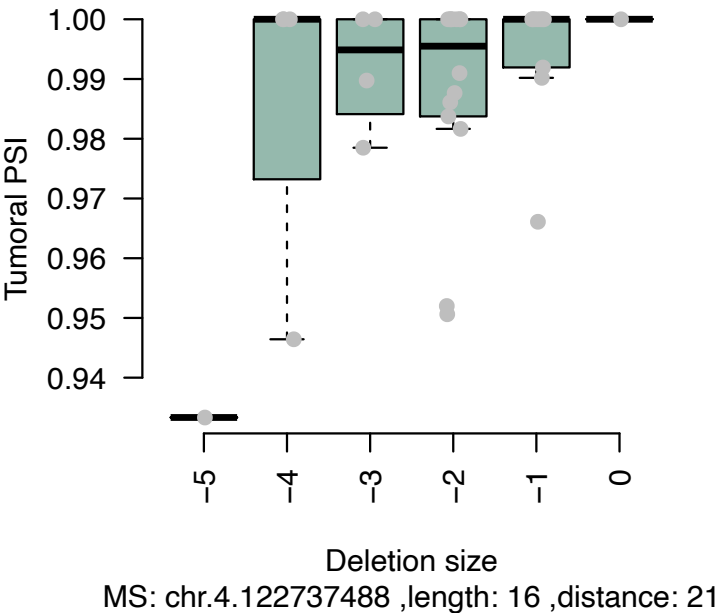


gene: TMX3 ,exon: 113 ,type: Constitutive

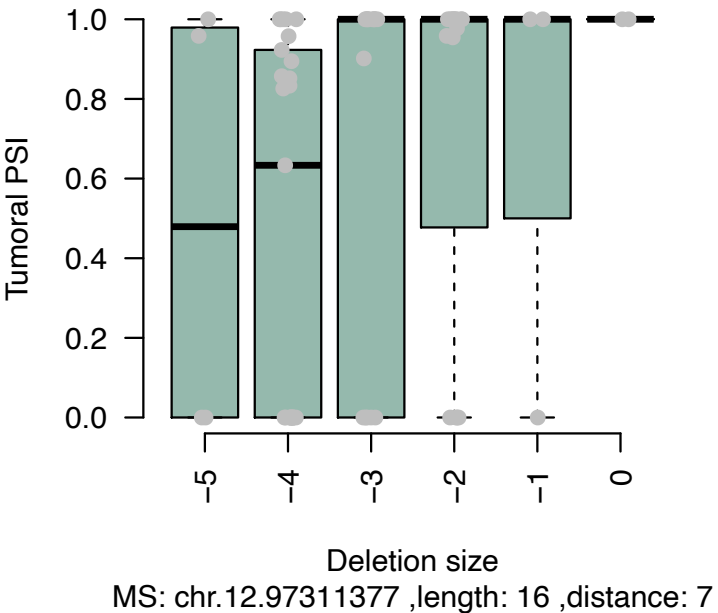


Supplementary Figure S6

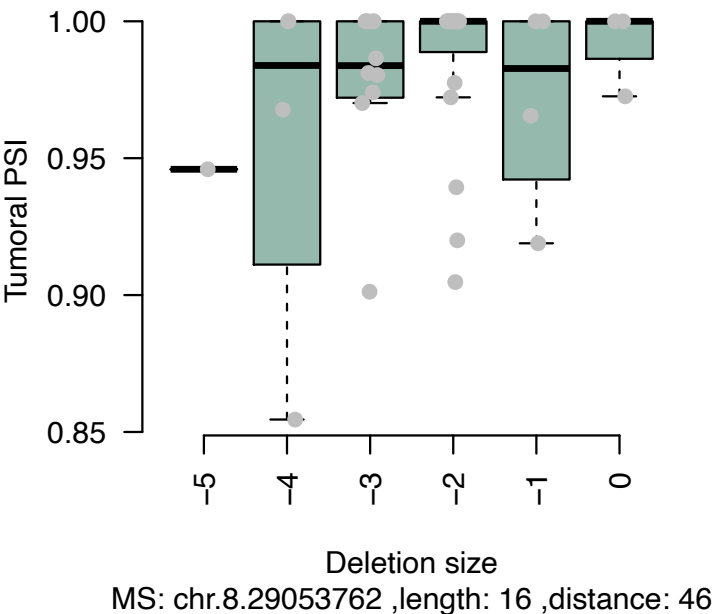
gene: EXOSC9 ,exon: 23 ,type: Constitutive



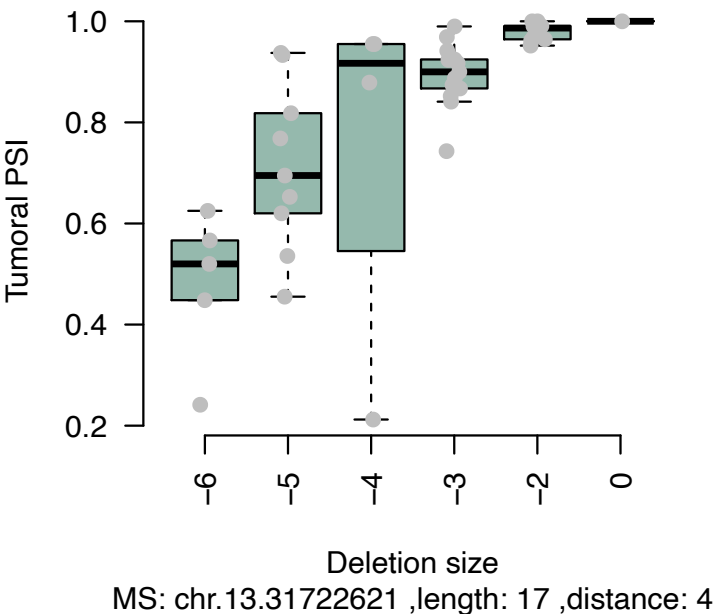
gene: NEDD1 ,exon: c(80, 102) ,type: Constitutive



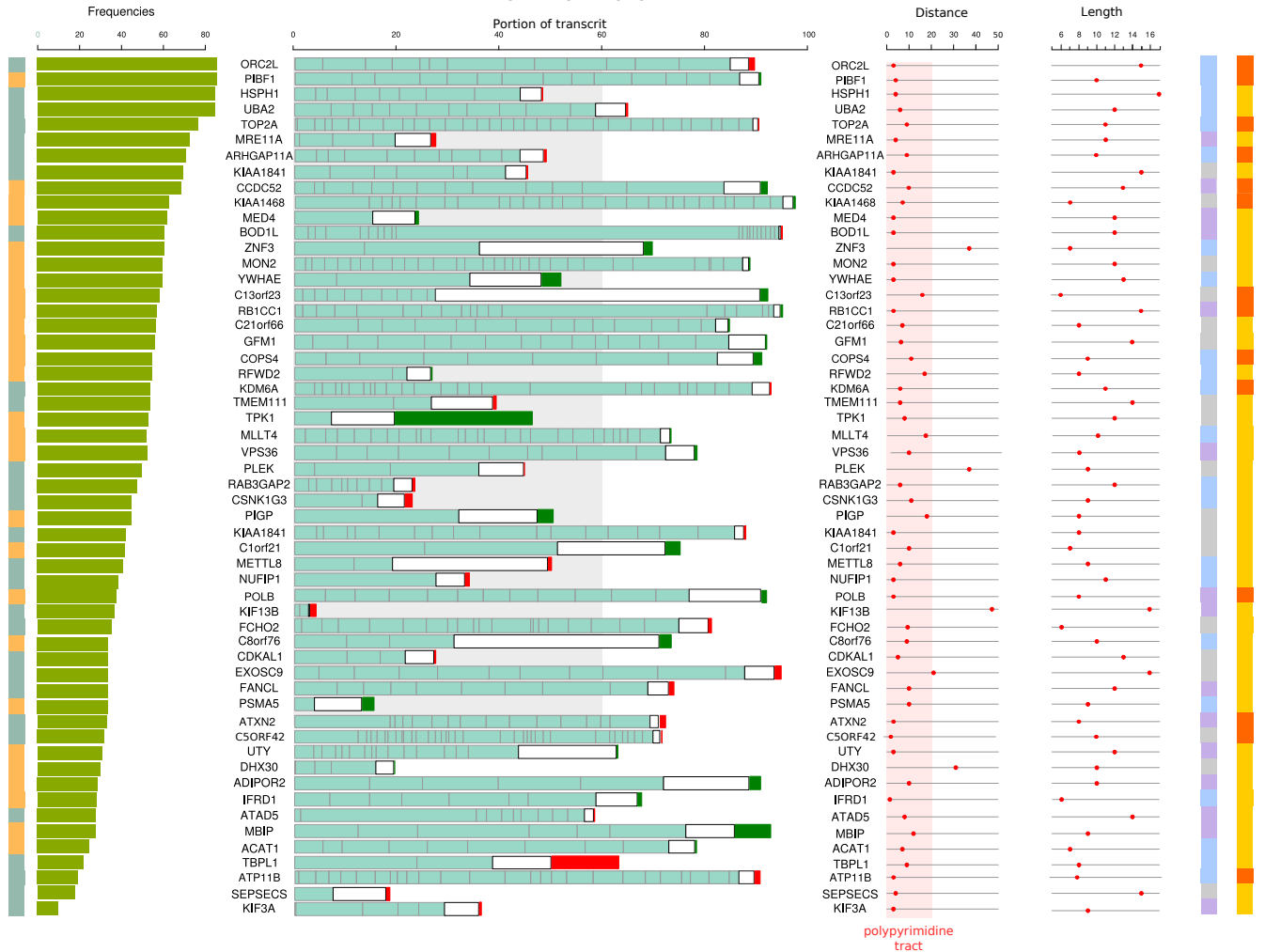
gene: KIF13B ,exon: 80 ,type: Constitutive



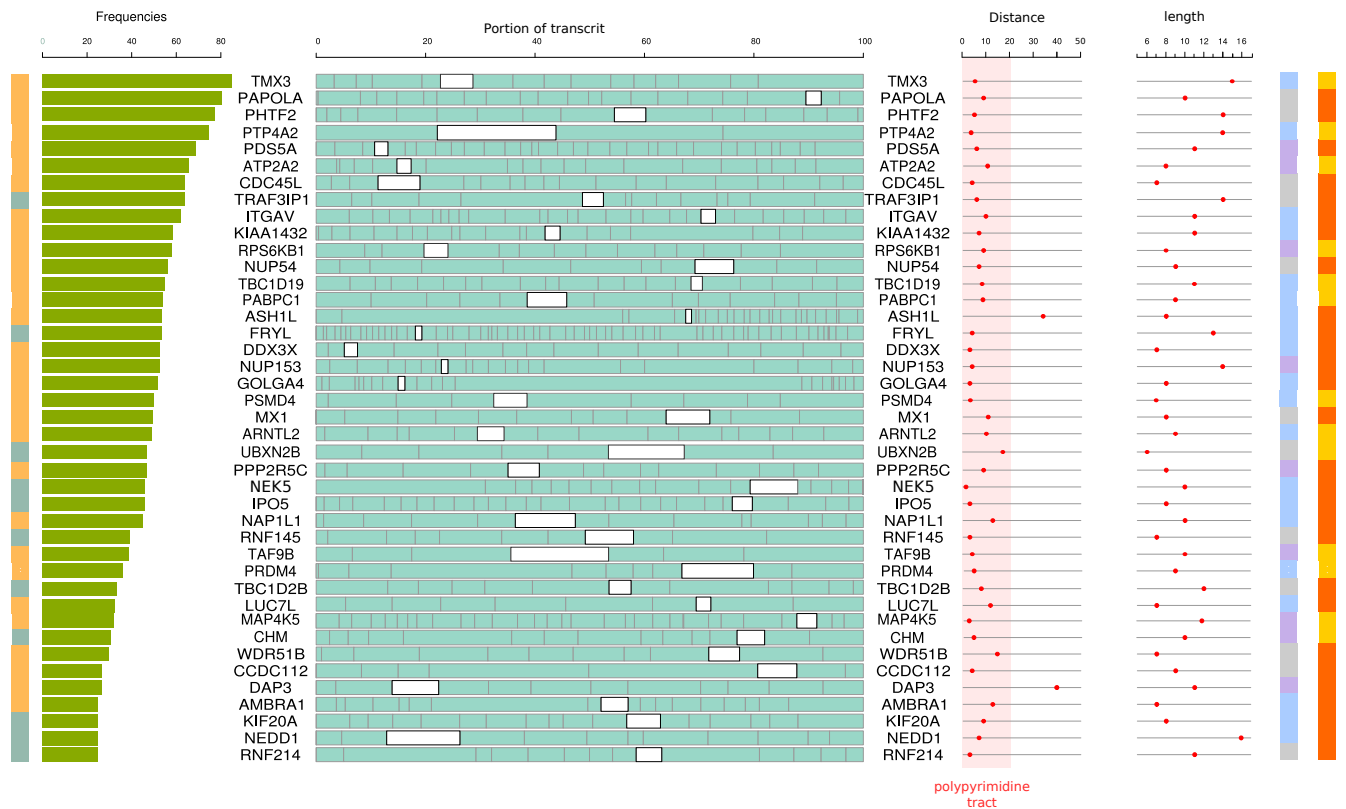
gene: HSPH1 ,exon: 146 ,type: Constitutive



FRAMESHIFT MUTATIONS



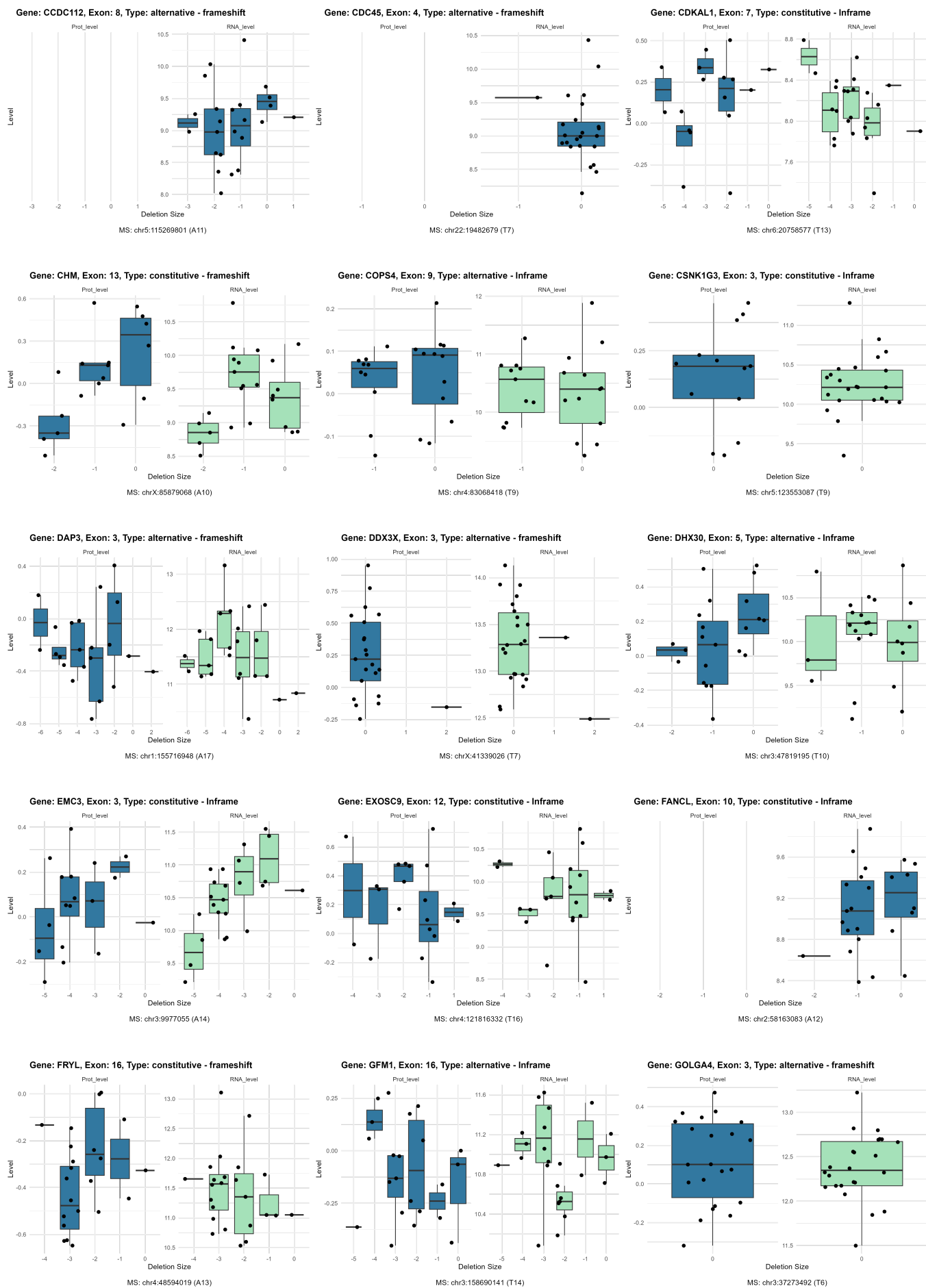
INFRAME MUTATIONS



A



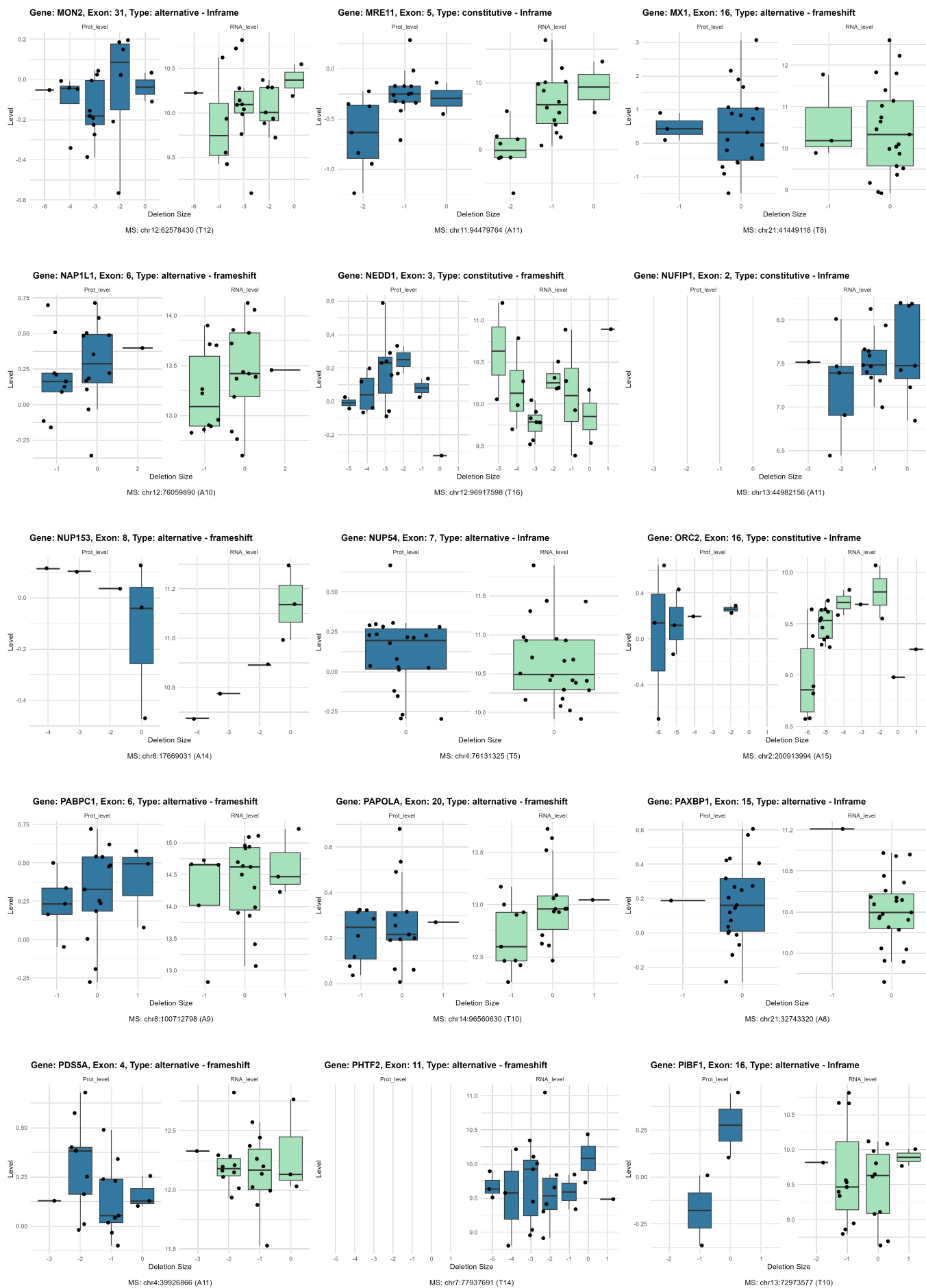
Supplementary Figure S8



Supplementary Figure S8



Supplementary Figure S8



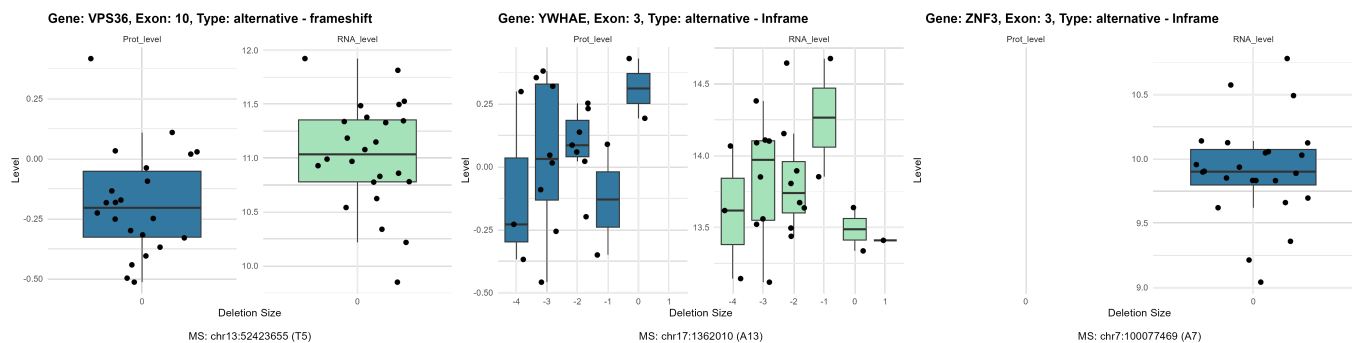
Supplementary Figure S8



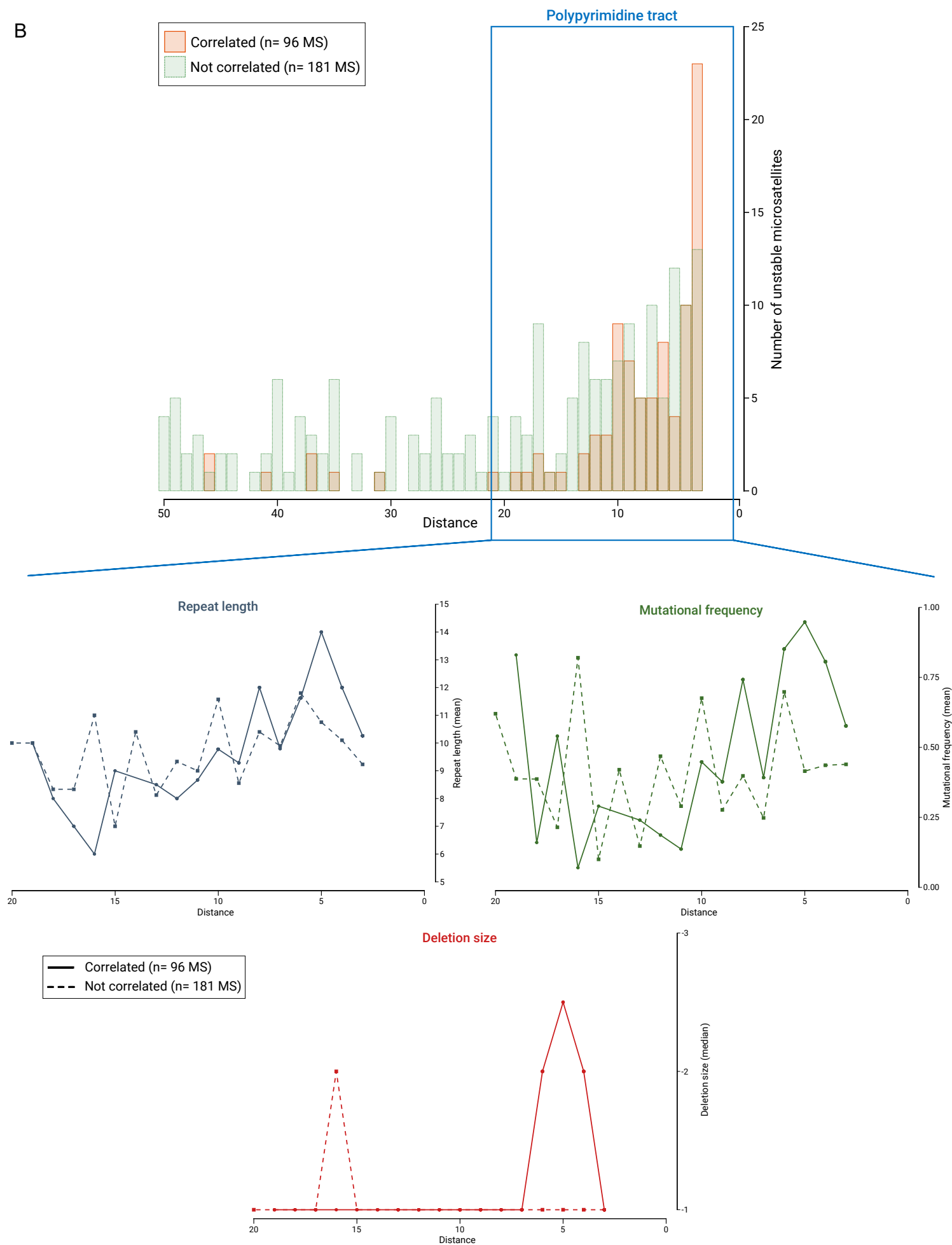
Supplementary Figure S8



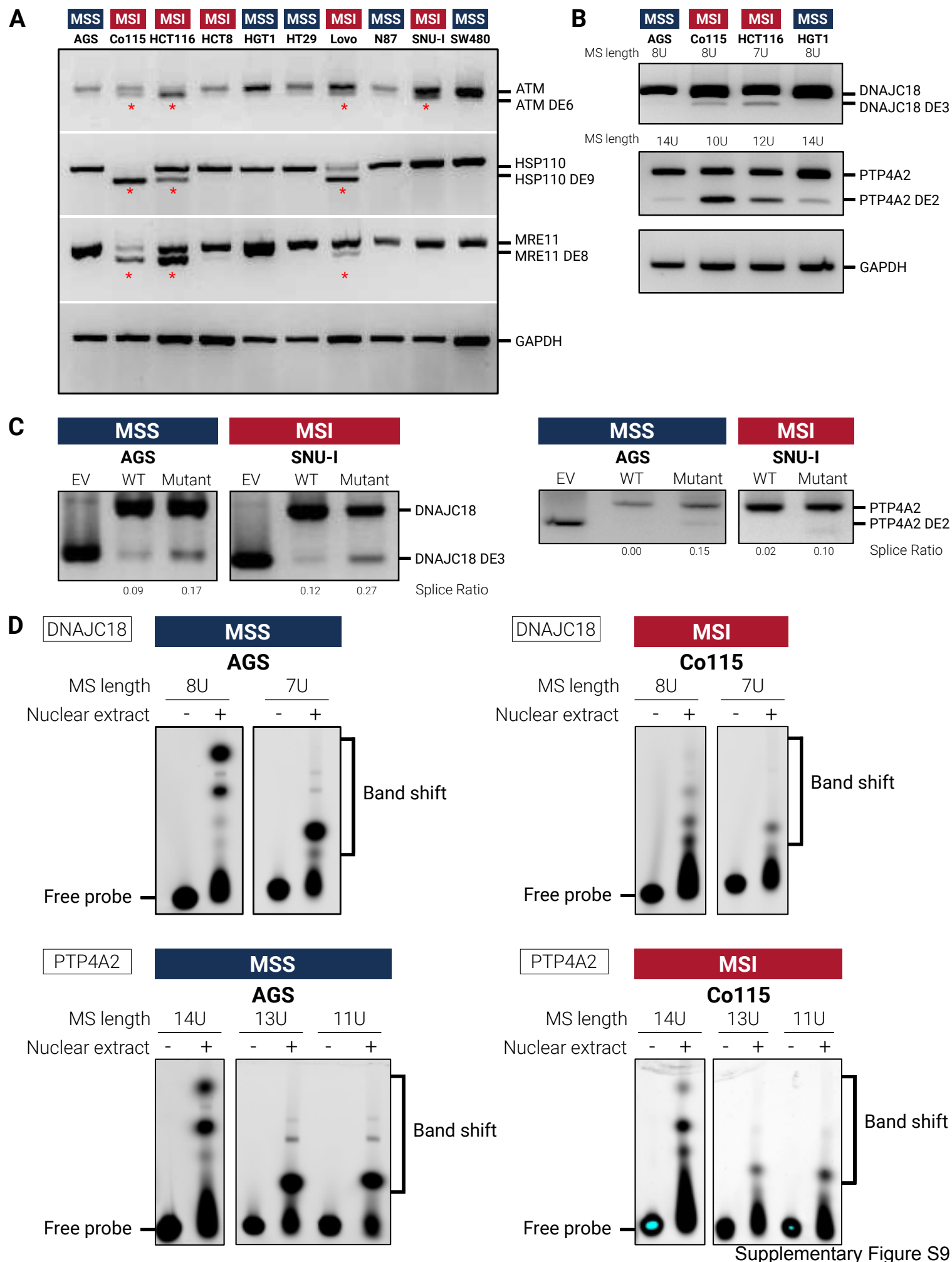
Supplementary Figure S8

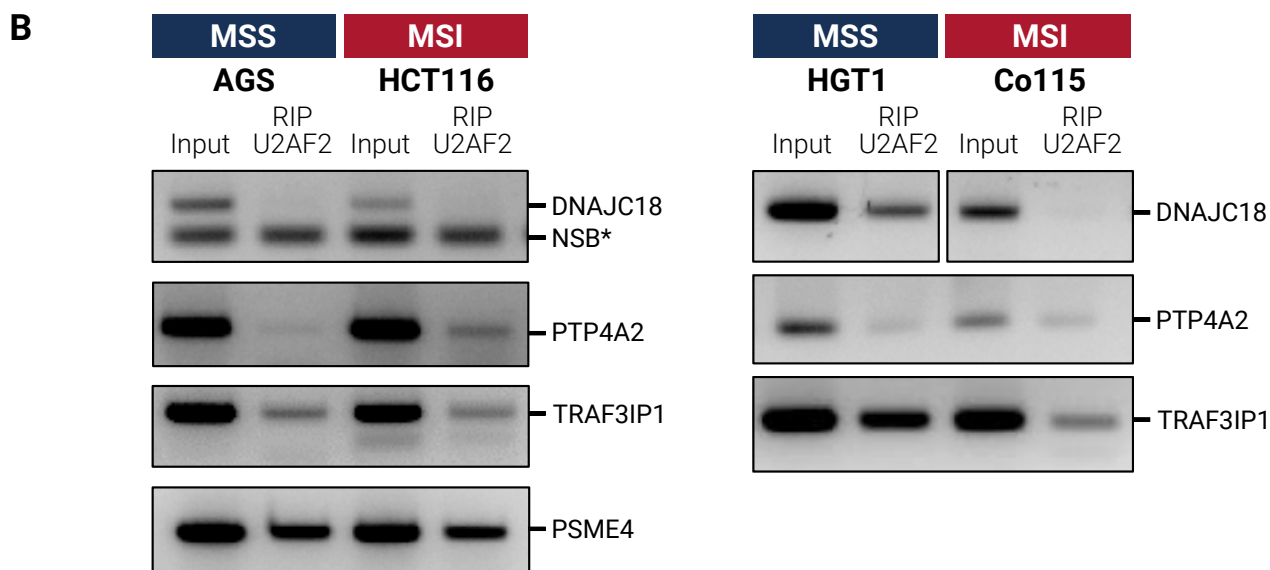
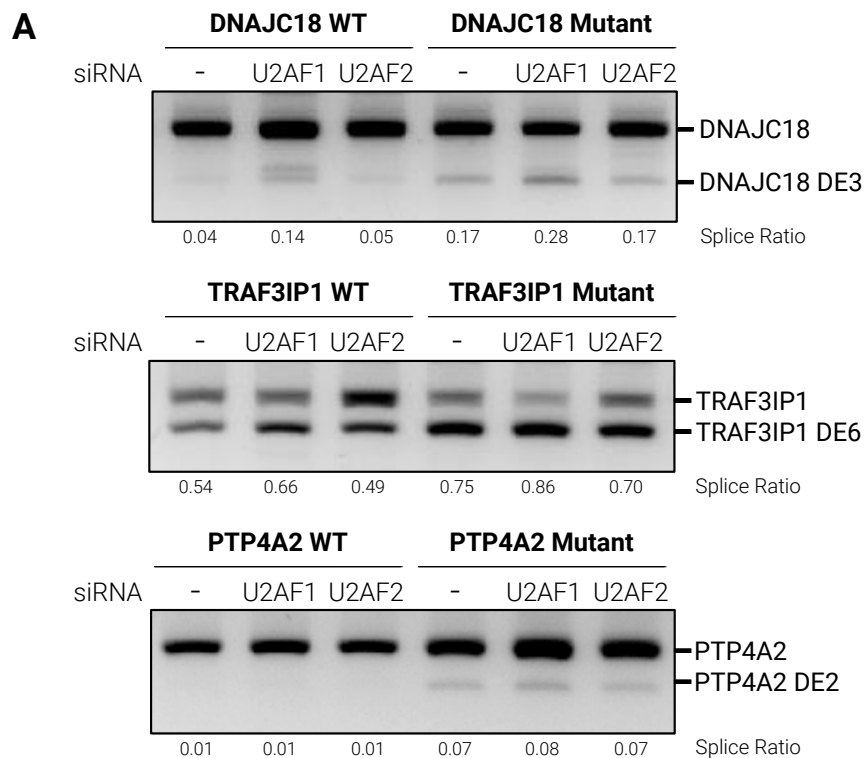


B

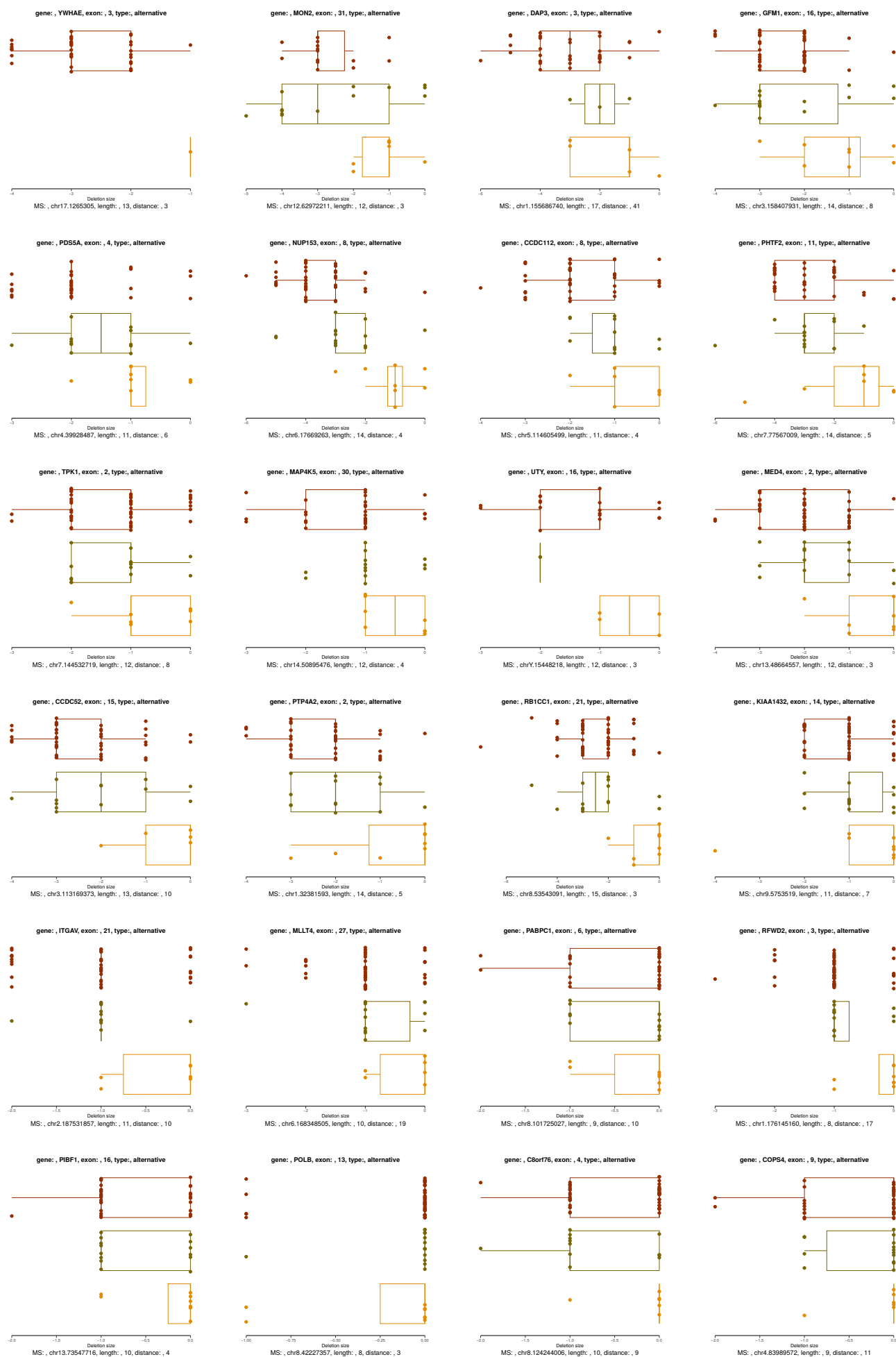


Supplementary Figure S8

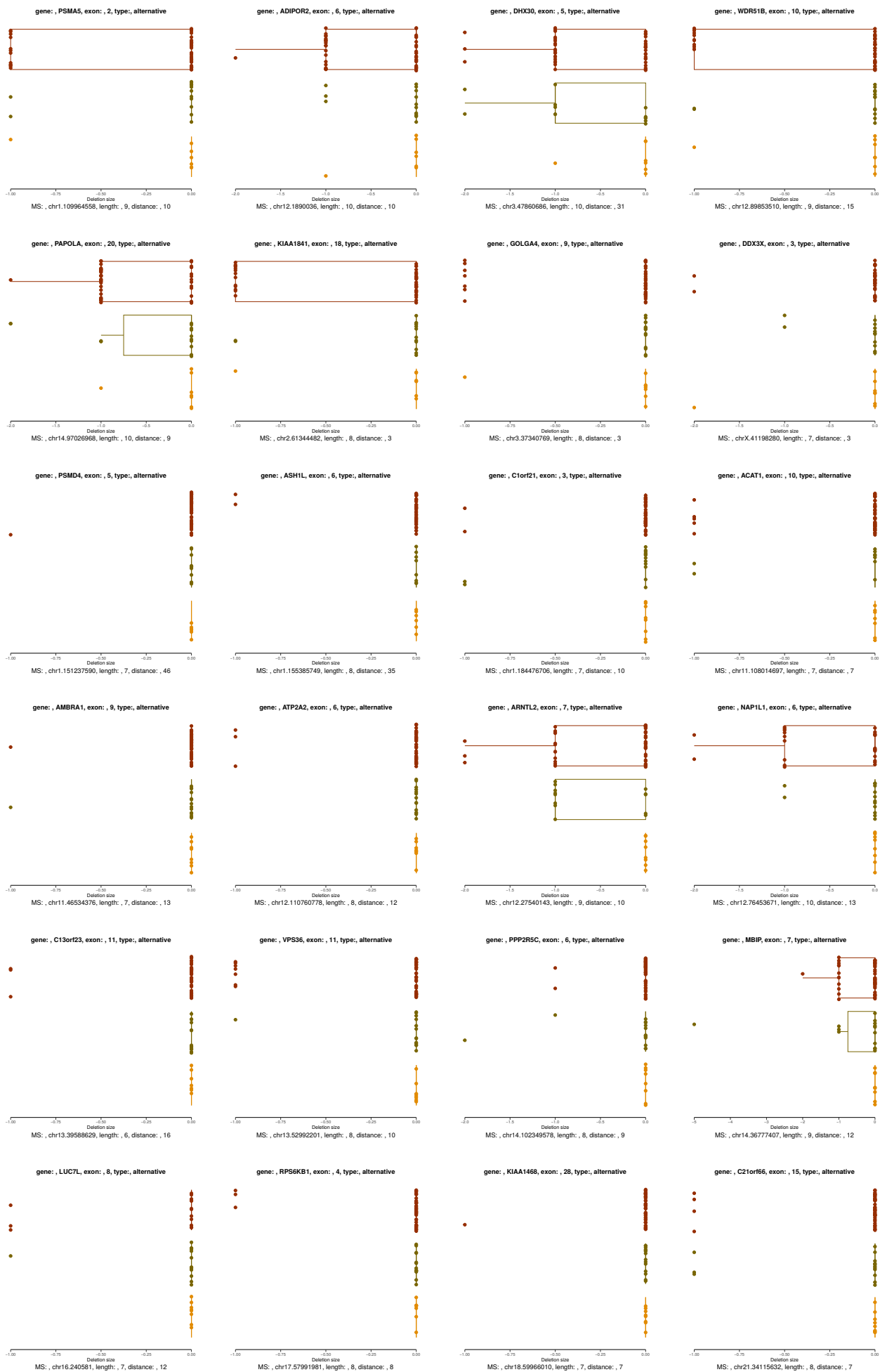




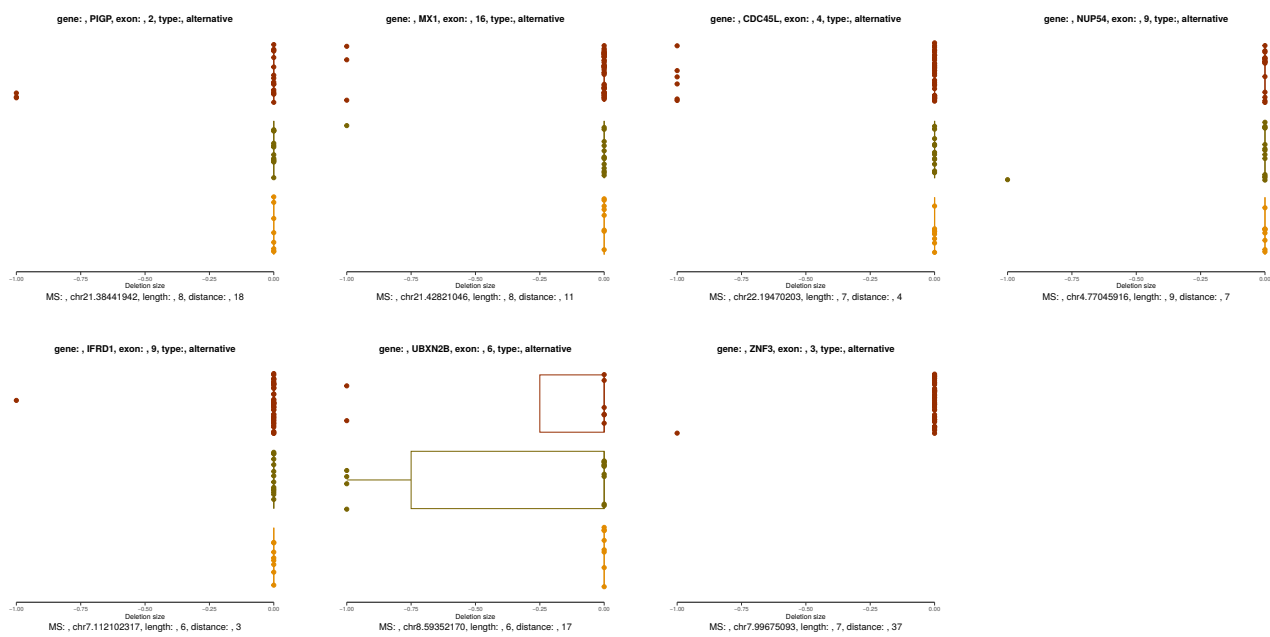
Supplementary Figure S10

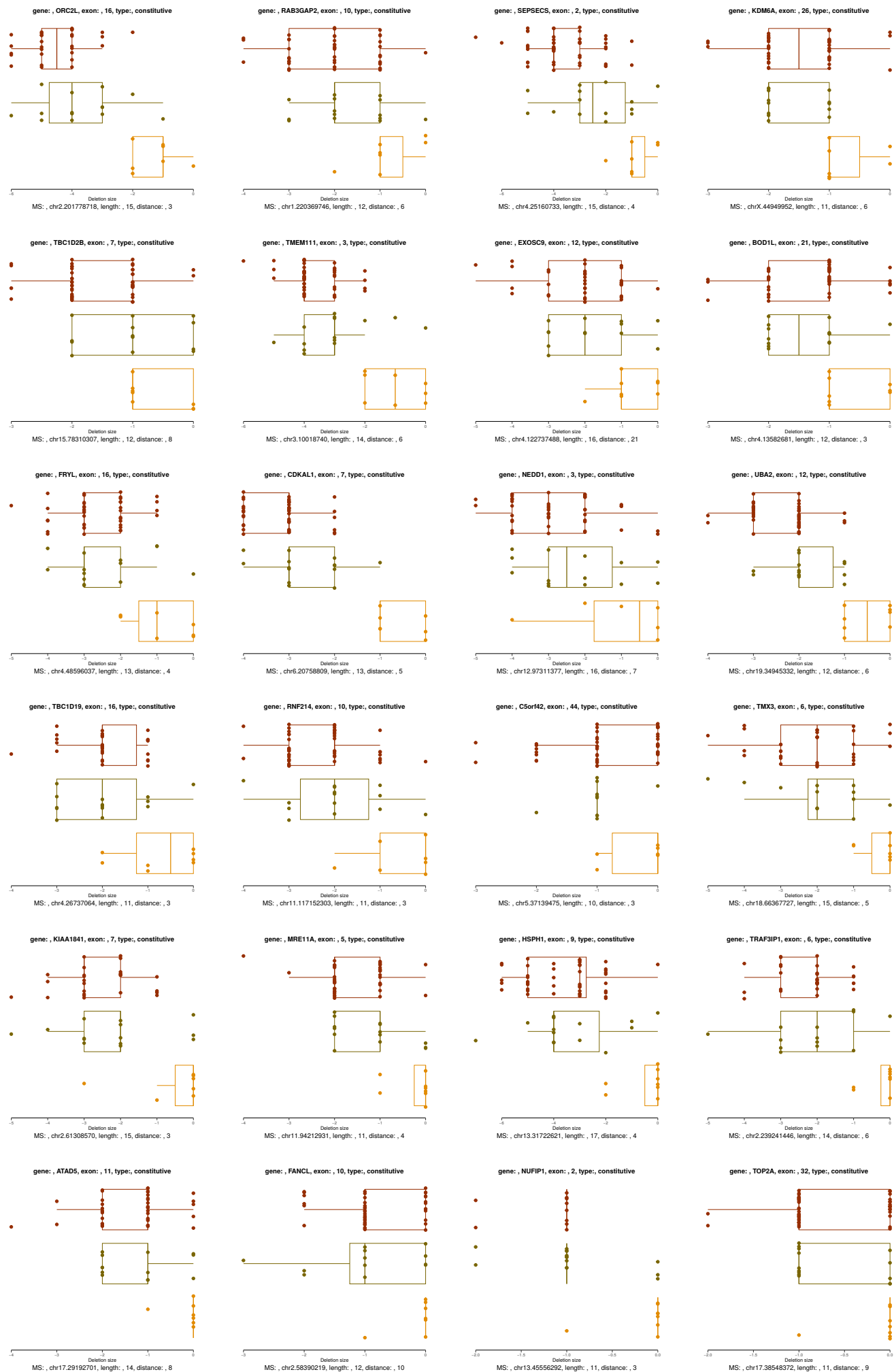


Supplementary Figure S11

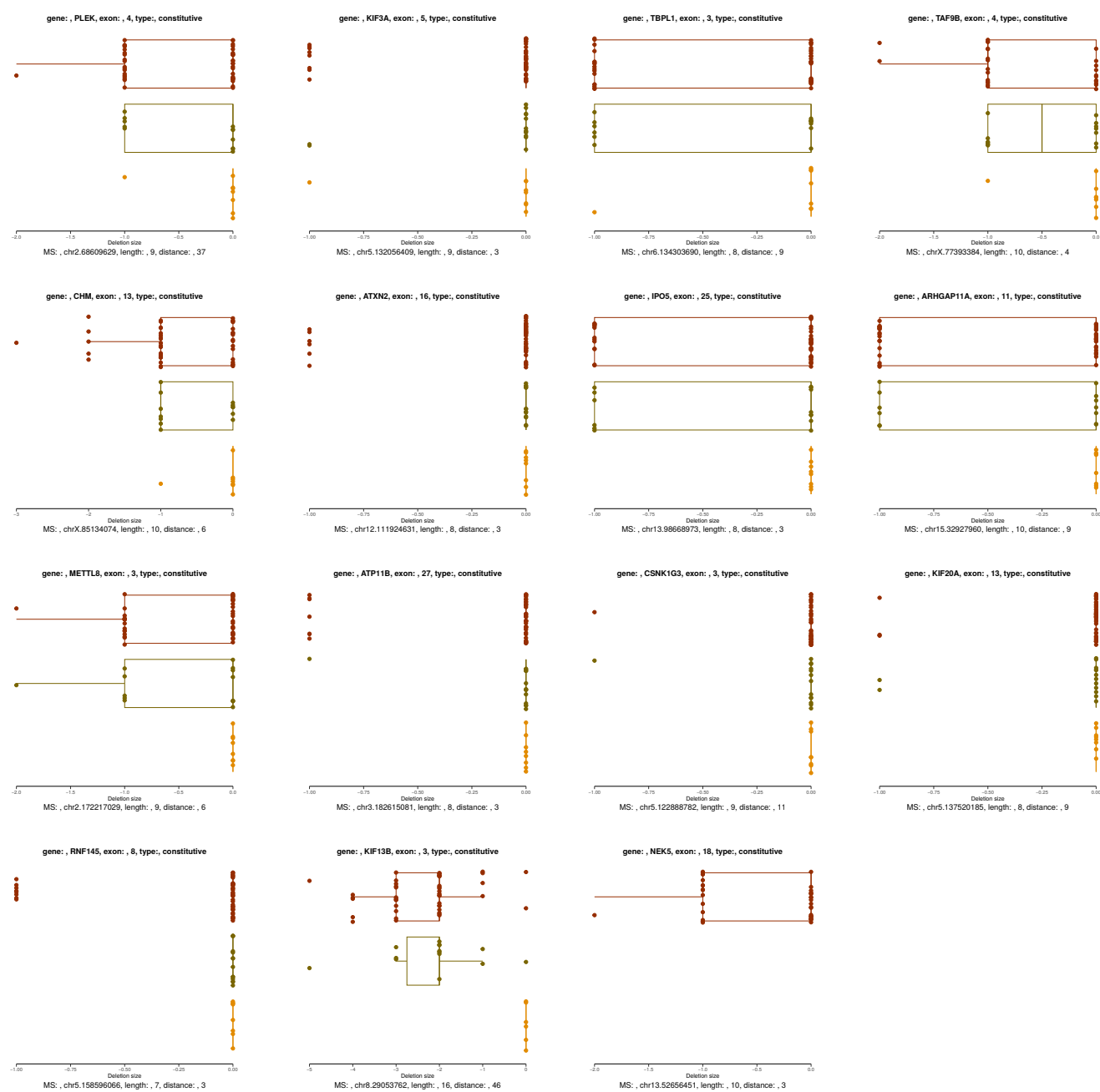


Supplementary Figure S11

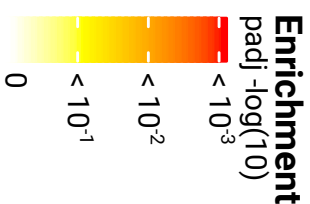
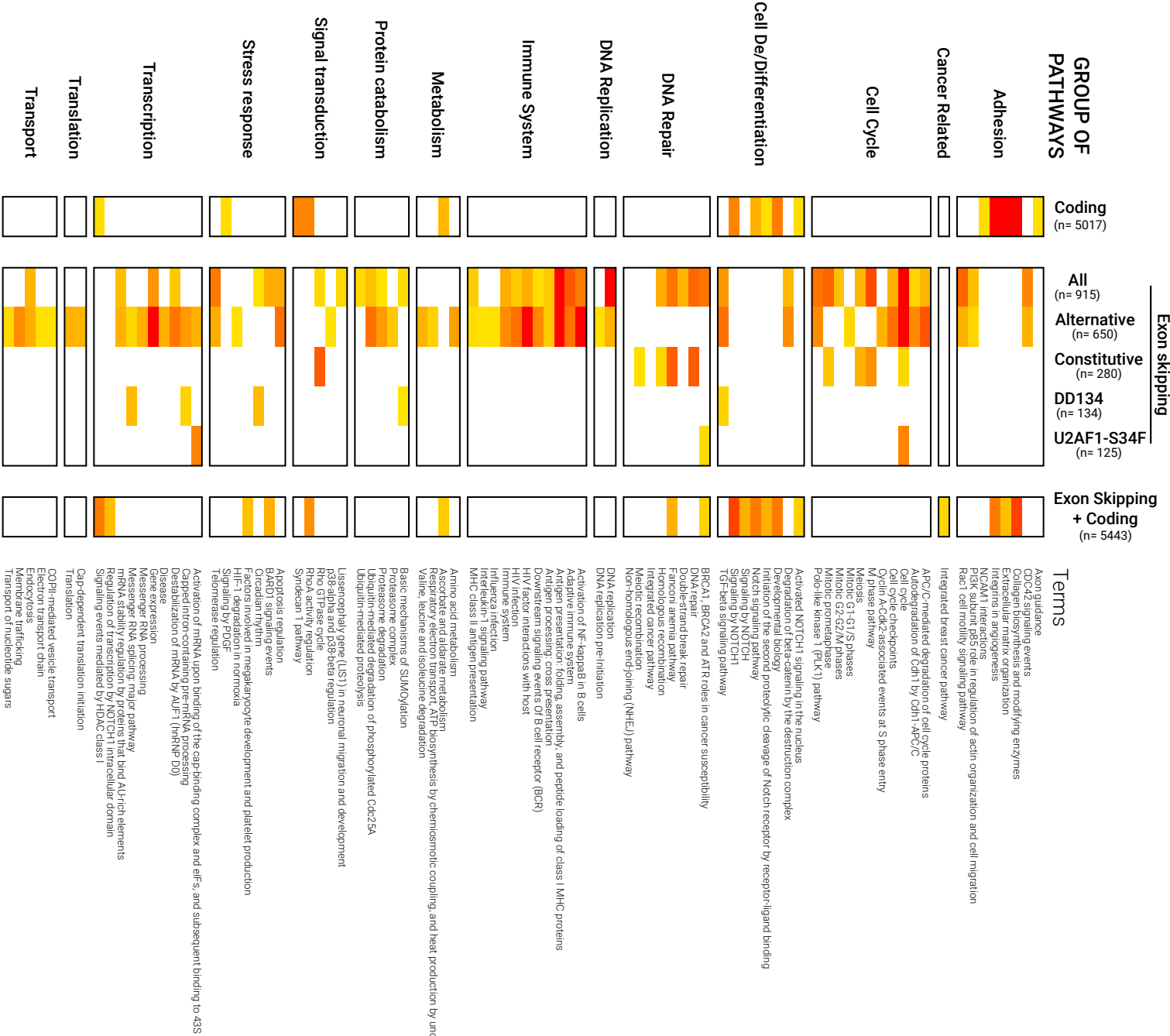




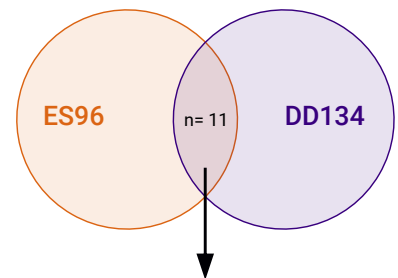
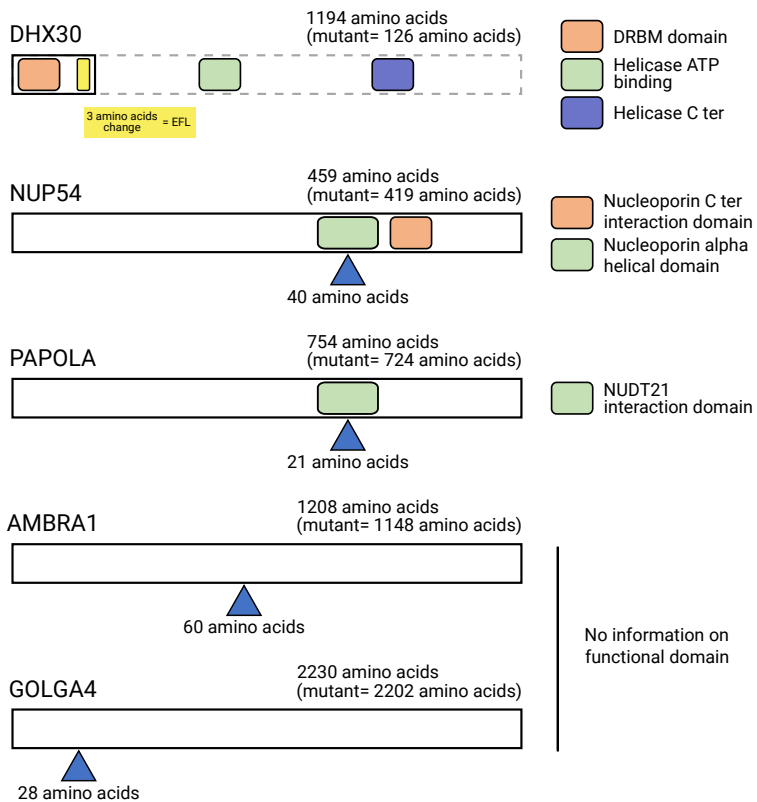
Supplementary Figure S11



Supplementary Figure S11

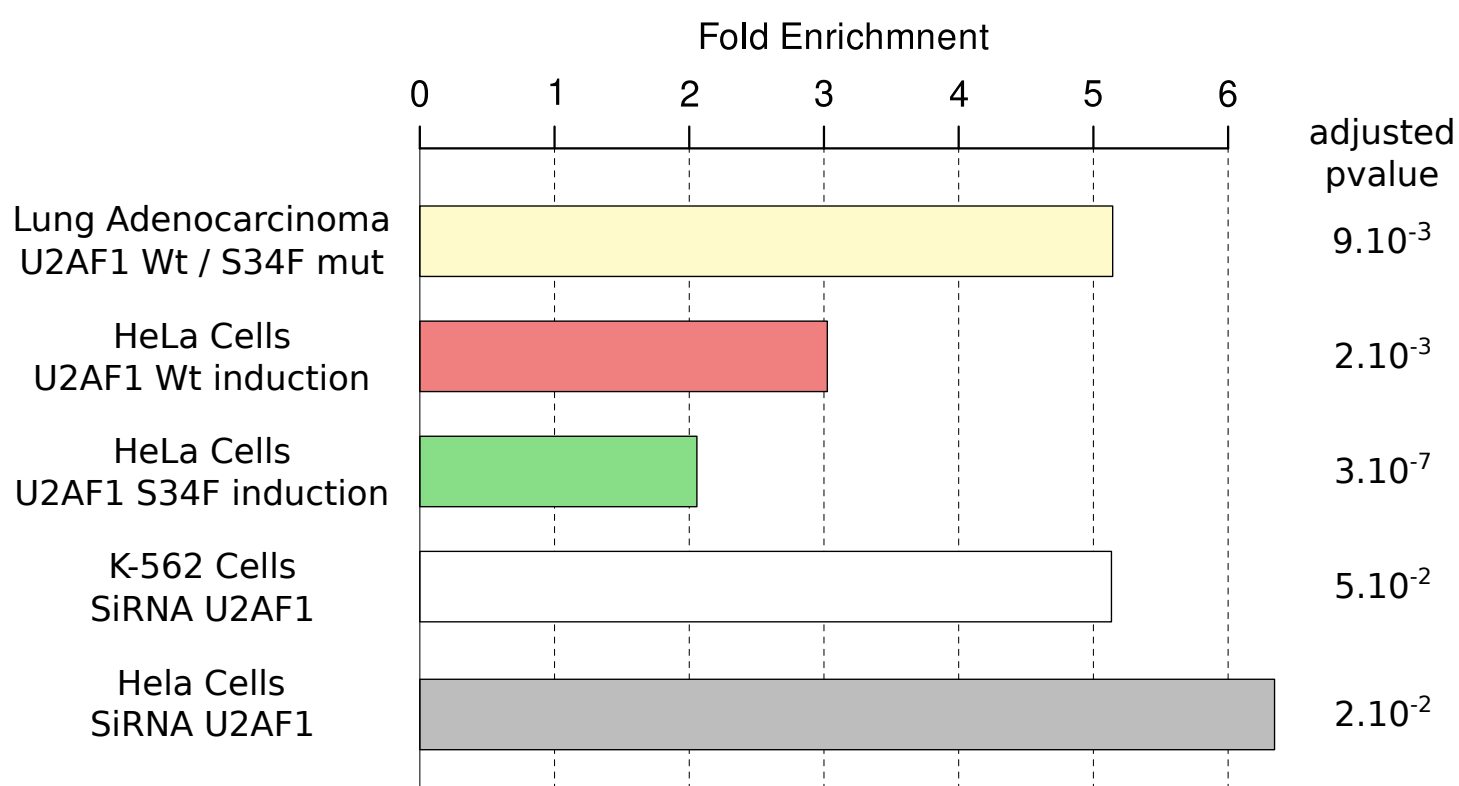


Supplementary Figure S12



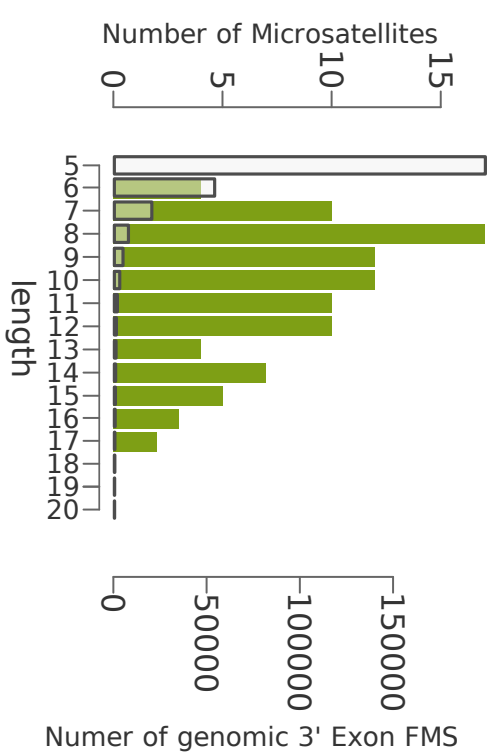
Gene	Exon	Mutation type	Exon type
DHX30	5	frameshift	alternative
NUP54	9	inframe	alternative
PAPOLA	20	inframe	alternative
AMBRA1	9	inframe	alternative
GOLGA4	9	inframe	alternative
ARHGAP11A	11	frameshift	constitutive
ATXN2	16	frameshift	constitutive
CSNK1G3	3	frameshift	constitutive
FRYL	16	inframe	constitutive
RNF145	8	inframe	constitutive
RNF214	10	inframe	constitutive

Supplementary Figure S13

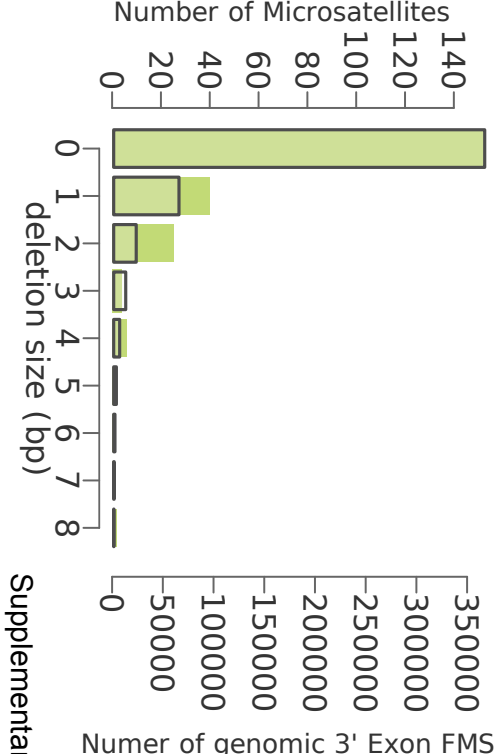
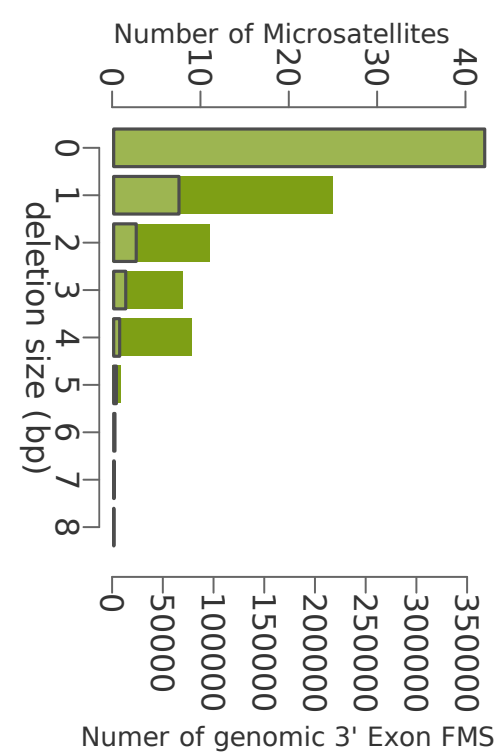
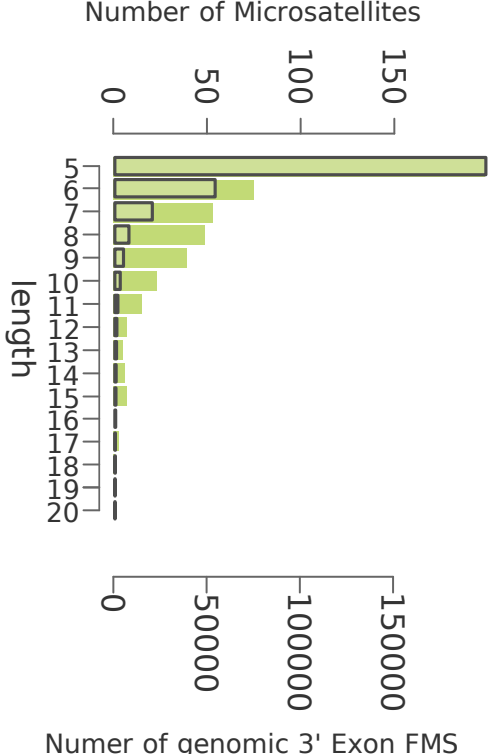


Supplementary Figure S14

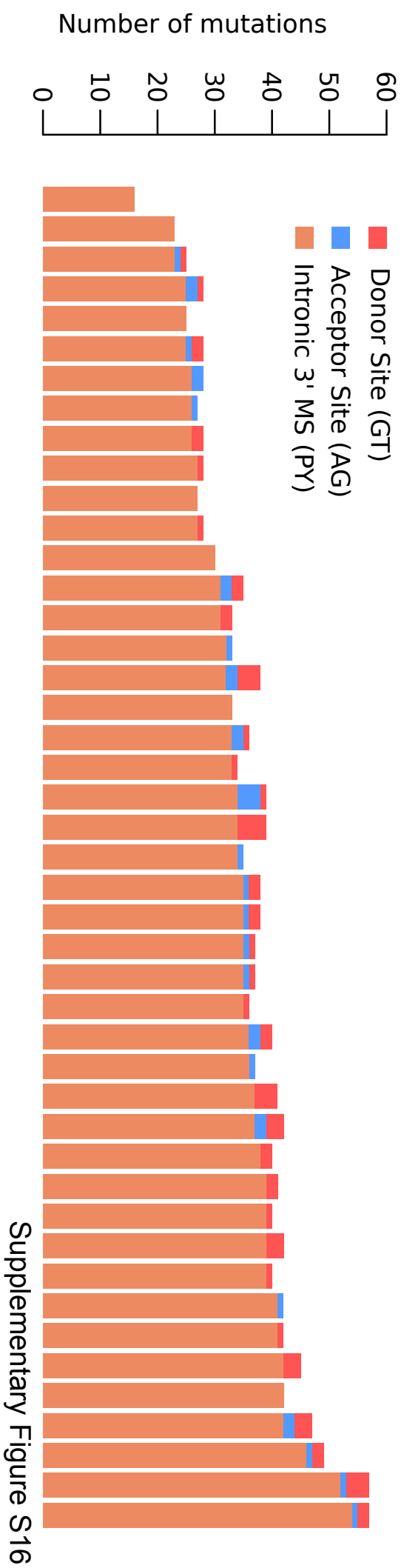
PSI correlated with deletion size of MS



PSI independent of deletion size of MS



Supplementary Figure S15



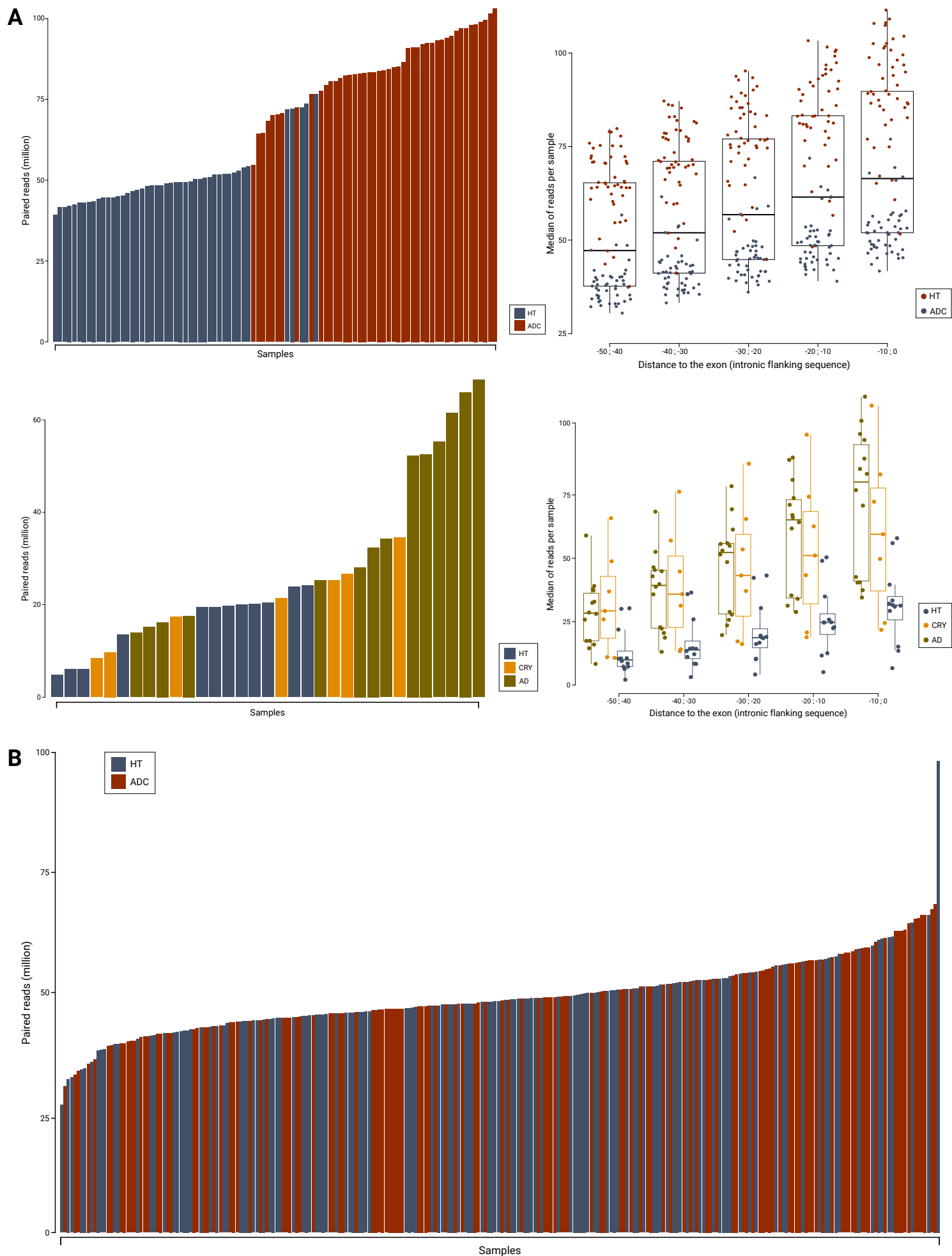


Figure S1:

(A) Distribution of mutational frequency (left) and number of unstable MS (right) by sample according to MS repeat length; samples are sorted by condition. (B) Boxplot of cumulative deletion events in MS according to region and condition. Statistics: Student's t-test, * P-value < 0.05, ** P-value < 1.10^{-3} , *** P-value < 1.10^{-4} , **** P-value < 1.10^{-5} .

Figure S2:

(A) The Percent Spliced In (PSI) value, which corresponds to the fraction of isoforms that include the exon (ranging from 0 to 1) was calculated for all samples and was used to identify the significant deregulated exon skipping events (two-sided adjusted *t*-test) between normal tissue and tumors. (B) The same approach was used to identify significant deregulated exon skipping events between MSI tumors and MSS tumors.

Figure S3:

Bar plots show the number of exon exclusion (A) or inclusion (B) events in MSI and MSS tumors (C, D) compared to normal tissue (left) or in MSI tumors compared to MSS tumors (right) (for details, see **Table S1**). All the candidate exons are flanked by at least one microsatellite in the 3' part of the previous intron (up to 50 bp in intronic sequence from the AG acceptor splicing site, 3' exon-FMS). Most deregulated exons correspond to exon skipping events in MSI tumors (surrounded by a red rectangle).

Figure S4:

(A) Distribution of exclusion reads according to inclusion reads (log2 scale) for 2 examples of constitutive exons (left) and for 2 examples of alternative exons (right) deregulated between tumors and normal tissue. (B) Two-dimensional histograms showing the number of excluded reads according to the number of included reads for all exons. The density of exons is displayed by a logarithmic color code. The upper panel represents exons classified as alternative, and the lower panel represents exons classified as constitutive. This classification is based on the interpretation of data from all normal tissues (n=133). (C) Distribution of exclusion reads according to inclusion reads (log2 scale) in all normal tissues (n=133) for 8 examples of exons classified as alternative (n=6) or constitutive (n=2).

Figure S5:

(A) The total number of cells used was the same that in the Pelka et al. study (n = 371223). (B) We used COSG library to obtain three gene markers from each cell type [66]. (C) The degree of correlation was high when using the top-level annotation provided by Pelka et al. [26]. (D) As expected, with a higher granularity using the full annotation by Pelka et al., there was more variability but conserving the cell type when compared with our annotation.

Figure S6:

Boxplots show the Percent Spliced Excluded (PSI) values according to the size of the deletion for each of the 96 microsatellites displaying a significant correlation (named ES96) (for details, see **Table S2**).

Figure S7:

The biological consequence of exon skipping leading to a frameshift (upper panel) or an inframe (lower panel) mutations. Left part represents the type of exon (determined as constitutive or alternative in normal tissue) and its mutational frequency. The center part shows the protein representation, the proportion of protein lost and size of the neo-antigenic tail. The right part shows the distance of microsatellite from 3' splice site (AG) position, and the length of the intronic 3' MS, the expected functional impact are indicated (n=96; details of all proteins are available in **Table S2**).

Figure S8:

(A) Impact of each of the 96 MSI-driven exon skipping on the overall expression of corresponding mRNAs or proteins in MSI CRC using the public CPTAC dataset (https://pdc.cancer.gov/pdc/browse/filters/pdc_study_id:PDC000109%7CPDC000116%7CPDC000117). In contrast to results obtained at the exon level (PSI Index), no impact of the 96 MSI noncoding mutations was observed both at the mRNA and protein levels, or only for a restricted number of candidate genes. This was expected since these data are limited to the analysis of a much smaller number of patients with MSI CRC, without reaching significance in a number of cases. Moreover, the majority of these exons deregulated in MSI CRC are alternative, so that their skipping may rather affect the fine tuning of specific alternative mRNA and corresponding protein isoforms in MSI tumor cells. **(B)** Upper panel, representation of the distribution of the MS at the intronic boundary (-50 to 0 of intron/exon junction) whose mutations in MSI CRC are correlated with exon skipping events (MS/ES96) or not (n= 181 MS). Lower panels further describe MS/ES96 as compared to others in term of repeat length, mutational frequency and deletion size.

Figure S9:

(A) The RT-PCR products of ATM, HSP110 and MRE11 transcripts show cDNA fragments with a skipped exon specifically in MSI cells, as it was published before [22, 23]. *: indicate genomic mutations in the PY associated with exon skipping. **(B)** Additional RT-PCR were performed on new targets identified in the exon skipping signature (DNAJC18 and PTP4A2). The RT-PCR products show cDNA fragments with a skipped exon (DNAJC18 DE3, PTP4A2 DE2), specifically in mutated MSI cell lines and increasing with deletion size. **(C)** Aberrant splicing event demonstrated (by RT-PCR) in endogenous and minigene mutants (intronic MS mutation in the PY) DNAJC18 and PTP4A2 transcripts in both MSI and MSS transfected cells. EV: empty vector; WT: wild type; Splice Ratio= ratio of the intensity of the exon-lacking cDNA fragment to the intensity of the sum of exon-containing + exon-lacking cDNA fragments. **(D)** Gel shift experiments have been conducted: nuclear protein extracts from MSI and MSS cells were incubated with DNAJC18 and PTP4A2 Wild Type or mutant labelled RNA probe. Shortened RNA probes do no longer allow the formation of large RNA/protein complexes.

Figure S10:

(A) RT-PCR products of DNAJC18, TRAF3IP1 and PTP4A2 wild type and mutant transcripts in siRNA silencing condition (none, U2AF1, U2AF2). The skipping exons are annotated by DNAJC18 DE3 or TRAF3IP1 DE6 or PTP4A2 DE2. (B) RNA immunoprecipitation analyses of U2AF2-bound PyT do not show consistent changes in U2AF2-PyT interaction according to the mutational status of the PyT. U2AF2-bound RNA was recovered and submitted to RT-PCR. cDNA products were then separated by agarose gel electrophoresis.

Figure S11:

Boxplots showing the size of the deletion for each of the 96 microsatellites displaying a significant correlation with exon skipping (ES96), according to the conditions; brown: adenocarcinoma, green: adenoma, yellow: dMMR crypts. The microsatellites are classified by exon type (alternative or constitutive) and ordered by dMMR crypts' mutational frequency (for details, see **Table S4**).

Figure S12:

Heatmap displaying pathway analysis of coding mutations, exon skipping signature (All, Alternative, Constitutive, DD134, ES96), and of the exon skipping + coding mutations; colors represent enrich pathway *Padj*, $-\log(10)$. Enrichment was realized with EnrichR - BioPlanet2019. Terms (rows, right) were sort by *Padj* (< 0.05) and regroup in groups of pathways (rows, left). Columns were sorted by enriched condition parameters.

Figure S13:

In silico analysis of the consequence of the exon skipping in 5 out of the 10 genes (DHX30, NUP54, PAPOLA, AMBRA1, GOLGA4) shared by the ES96 and DD134 signatures, whose exon lost due to MSI is an alternative exon. For DHX30, NUP54, and PAPOLA, it is highly probable that the loss of the alternative exon following MSI induces a functional change in the protein, or even a loss of function. For GOLGA4 and AMBRA genes, the lack of public information on the functional domains of these proteins prevents us from identifying or suggesting an effect of MSI-induced mutations if any.

Figure S14:

The bar plot shows the fold enrichment of the U2AF exon signature previously identified [37, 38] in our MSI exon signature (for details, see **Table S3**).

Figure S15:

The upper panel shows the distribution of microsatellites according to the repeat length, and the lower panel shows the distribution of microsatellites according to deletion size. Left, distribution of the 96 MS displaying a significant correlation between the deletion size of the intronic 3' MS and PSI. Right, distribution of the 181 MS for which the PSI value is independent of deletion size. For both panels, the background distribution of all flanking MS ($n = 335,564$) is indicated (gray).

Figure S16:

Bar plots displaying the number of mutations in the donor site (GT), and in acceptor site (AG) and the number of exons skipping for each of the 46 MSI tumors.

Figure S17:

(A) Representation of depth (left panel) and coverage at the intronic boundary (right panel) of WES in adenocarcinoma (top panel) and adenoma/crypts (lower panel) and associated with the normal mucosa tissues. (B) Depth (paired reads in million) of full-length RNA-seq in adenocarcinoma and associated with the normal mucosa tissues. HT: healthy tissue; ADC: adenocarcinoma; AD: adenoma; CRY: MMR-deficient crypts.

Table S1:

List of 3'Exon-FMS deregulated between normal tissue and MSI (Table S1A) or MSS (Table S1B) tumors or between MSI and MSS tumors (Table S1C).

Table S2:

List of the MS associated with 985 exons skipping events; the PSI values are indicated.

Table S3:

List of the 96 genes in which somatic deletions at flanking 3' non-coding MS are regions were correlated with exon skipping (ES96).

Table S4:

List of mutated genes in dMMR crypts from ES96 and coding region, as well as known coding MS (highly mutated) related to MSI CRC (independently of their MSI status, named MFMC); mutational frequency (MF) and MS deletion size (MD) are indicated throughout tumorigenesis. MF: mutational Frequency; MD: median of deletion size.

Table S5:

List of common genes between the specific genes/exons associated with intestinal stem cells and our MSI-exon signature.

Table S6:

List of publications that catalog the exon signatures of 9 RNA splicing factors and their potential overlap with our MSI-exon signatures (n=985).

Table S7:

List of depth and global coverage WES values for all samples.