
Figures and figure supplements

Recurrent disruption of tumour suppressor genes in cancer by somatic mutations in cleavage and polyadenylation signals

Yaroslav Kainov et al.

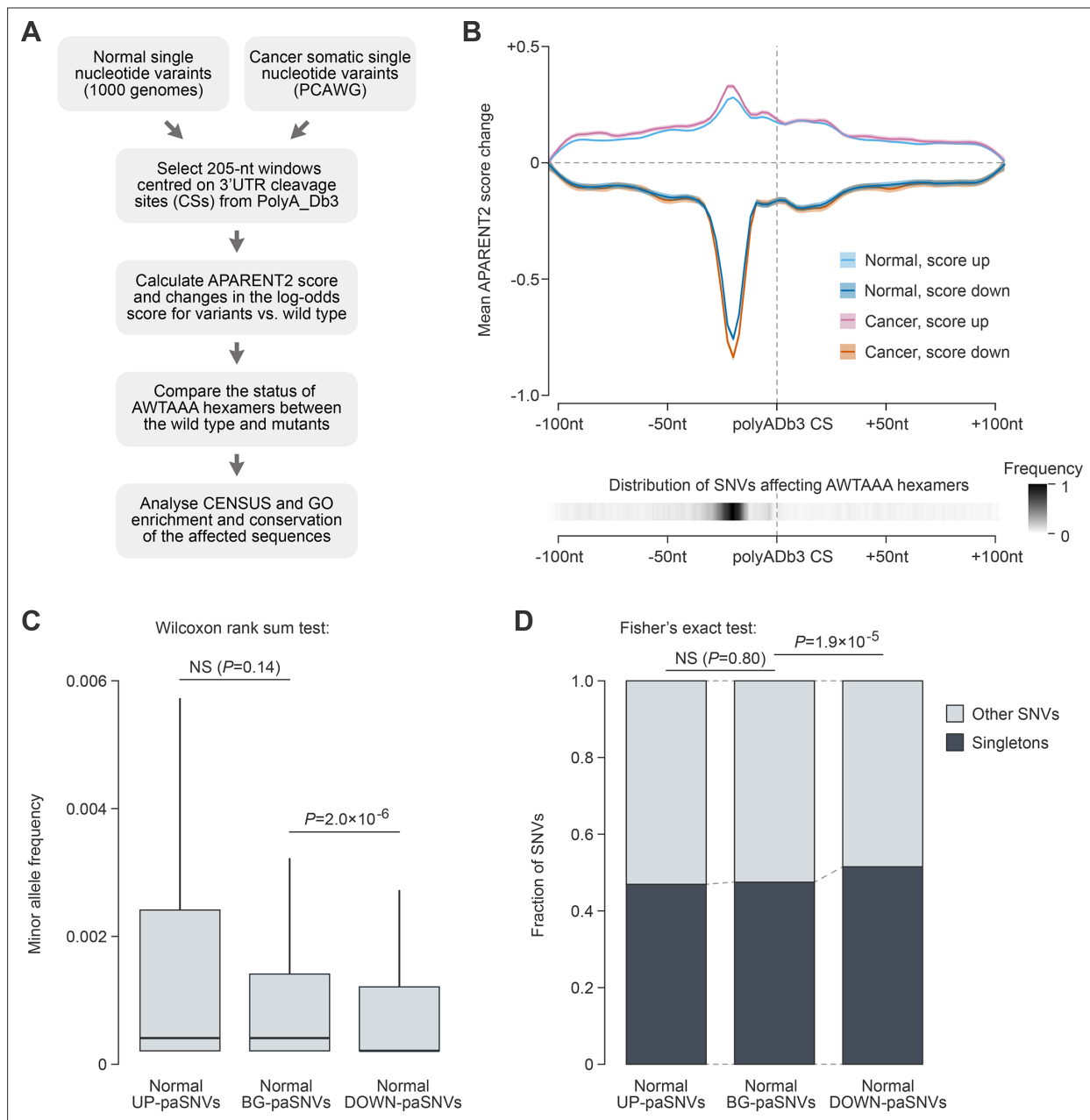


Figure 1. paSNVs disrupting cleavage/polyadenylation signals are depleted in the normal population. **(A)** Bioinformatics workflow used to analyse the effect of paSNVs on pre-mRNA cleavage and polyadenylation. **(B)** Top, effects of UP- and DOWN-paSNVs on the APARENT2 score (mean \pm SEM) as a function of their position with respect to annotated pre-mRNA cleavage sites (CSs). Bottom, combined distribution of AWTAAA-affecting paSNVs in both datasets. **(C)** Box plot showing that paSNVs disrupting polyadenylation signals are significantly less frequent compared to control groups of events in the normal population. **(D)** paSNVs disrupting polyadenylation signals are enriched for singletons, consistent with purifying selection against such events in the normal population.

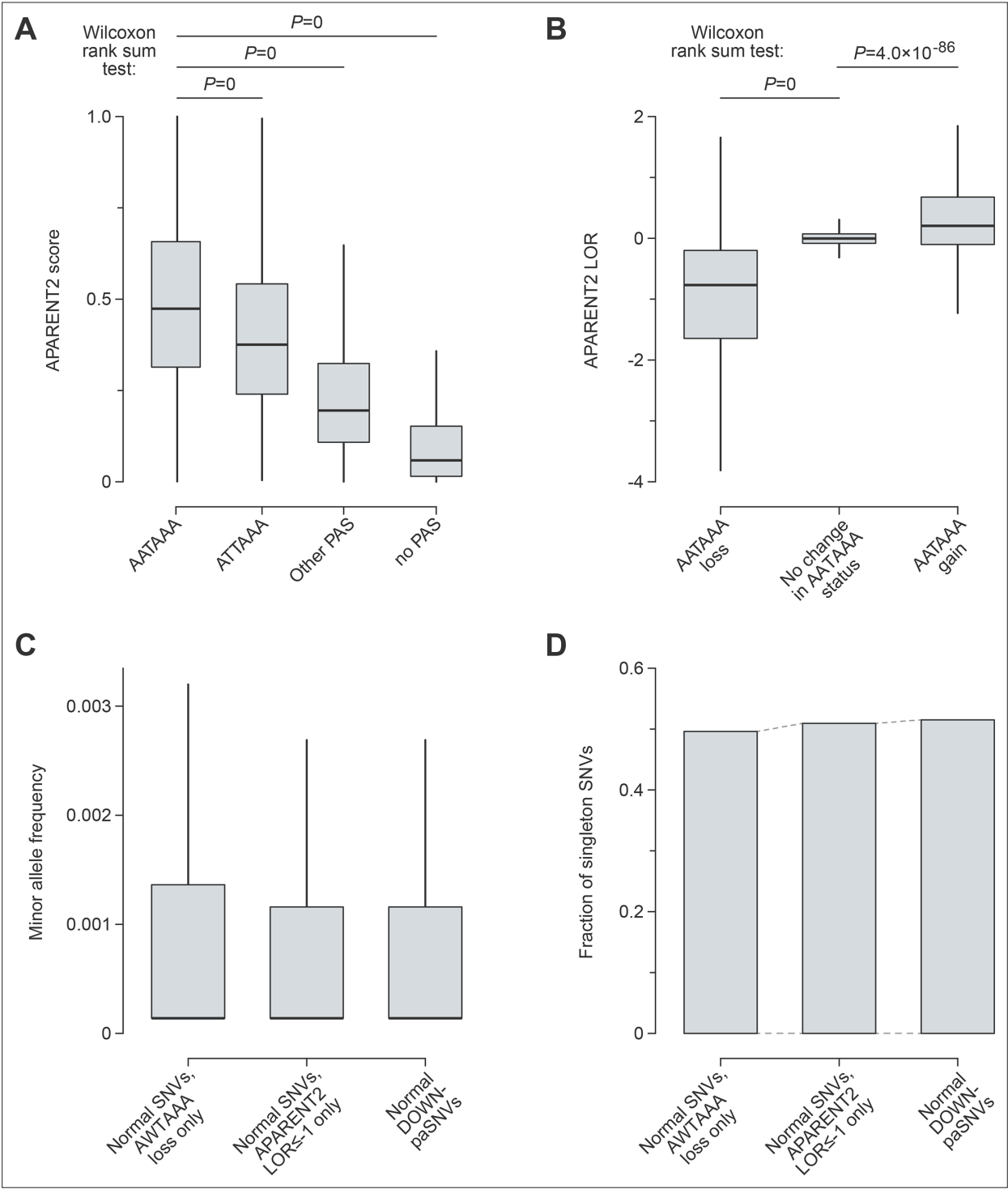


Figure 1—figure supplement 1. Distribution of cleavage/polyadenylation signal-disrupting mutations in the normal population (1000 genomes dataset). (A–B) As expected, (A) AWTAAA-containing paSNVs tend to be associated with a relatively high APARENT2 score and (B) the loss or gain of AATAAA typically reduced or increased the score, respectively. (C) Box plot comparison of normal-population allele frequencies of cleavage/polyadenylation signal-disrupting mutations defined by considering only AWTAAA gain/loss, only APARENT2 score changes, or both (DOWN-paSNVs). (D) Bar plot comparison of normal-population fractions of singletons for cleavage/polyadenylation signal-disrupting mutations defined by considering only AWTAAA gain/loss, only APARENT2 score changes, or both (DOWN-paSNVs).

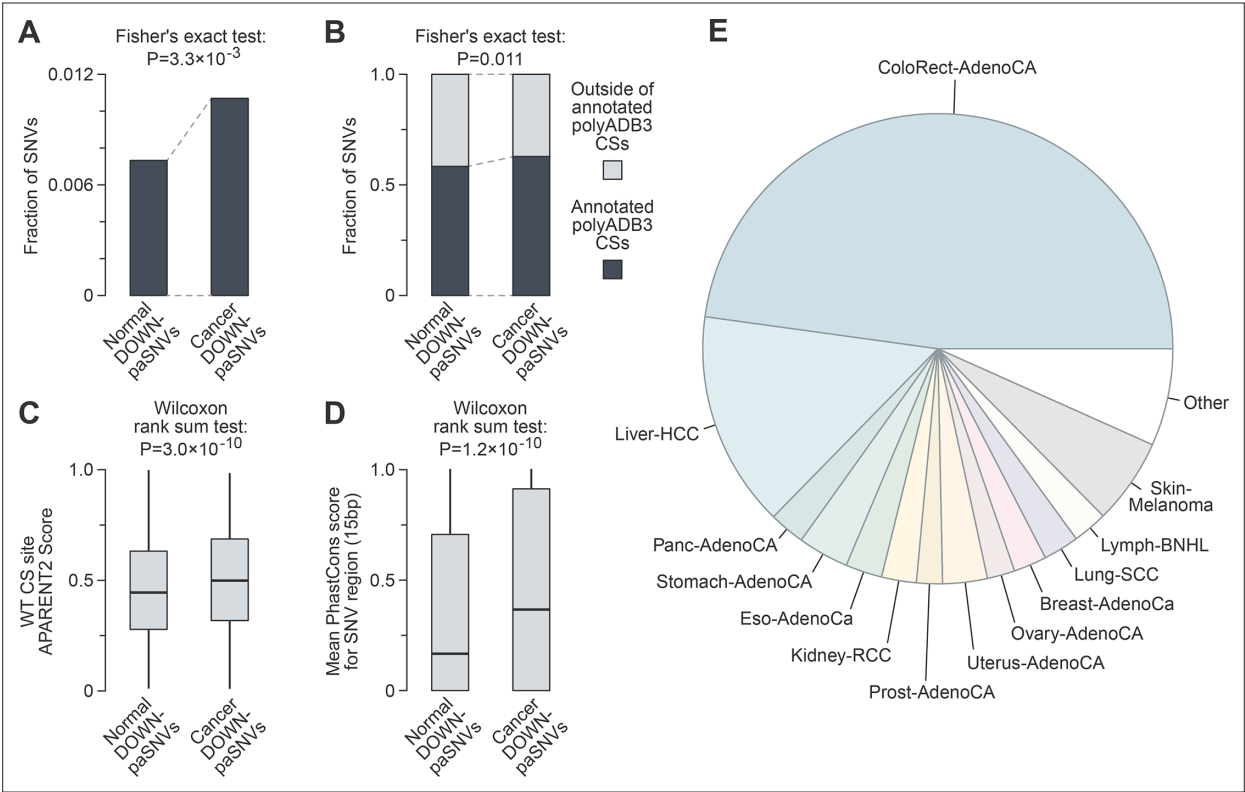


Figure 2. Cancer somatic mutations tend to disrupt functional cleavage/polyadenylation signals. **(A)** Bar plot showing enrichment of paSNVs disrupting polyadenylation signals among cancer somatic mutations. **(B)** Bar plot showing enrichment of single-nucleotide variants (SNVs) affecting AWTA AAA sequences in 3'UTRs close to annotated cleavage sites (CSs) among cancer somatic mutations. **(C)** Box plot showing that somatic mutations disrupt stronger cleavage/polyadenylation signals in cancer. **(D)** paSNVs disrupting polyadenylation signals occur in more evolutionary conserved regions in cancer (mean PhastCons score in 15-nt window centred at SNVs). **(E)** Distribution of DOWN-paSNVs across cancer types in the Pan-Cancer Analysis of Whole Genomes (PCAWG) project.

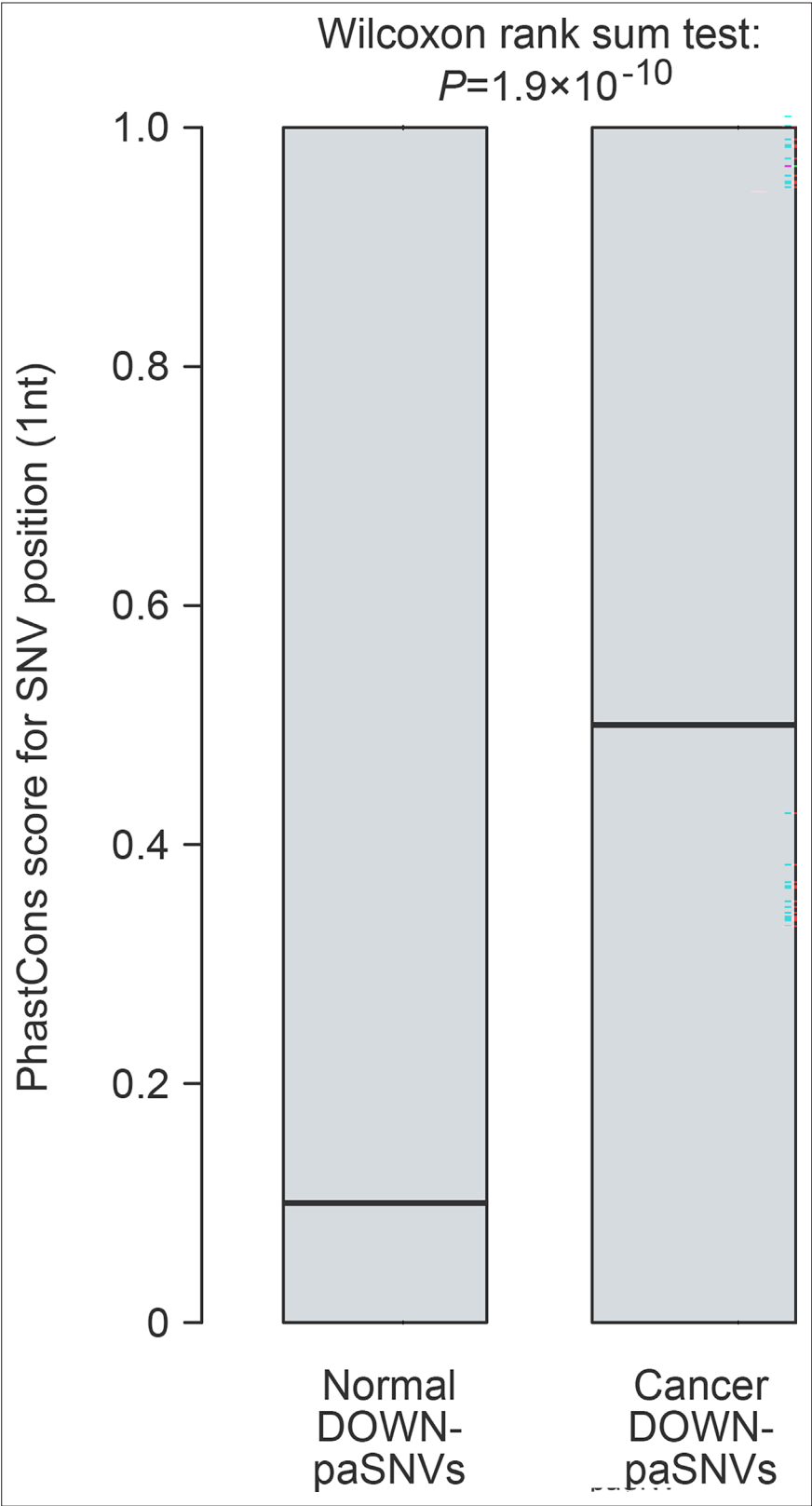


Figure 2—figure supplement 1. Cancer somatic DOWN-paSNVs often occur in evolutionarily conserved regions. The plot is generated similarly to **Figure 2D** except the conservation was calculated for the exact single-nucleotide variant (SNV) position.

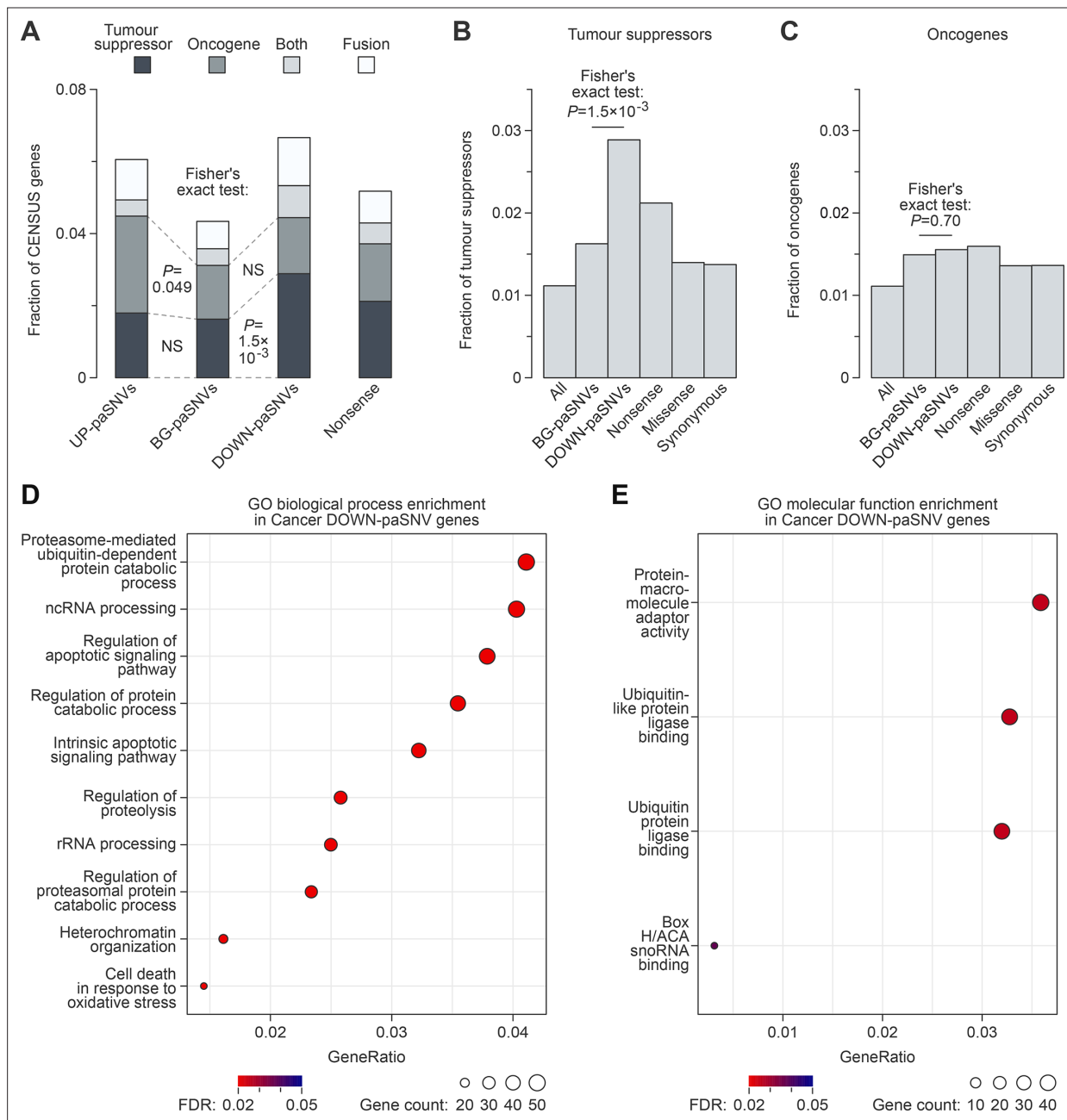


Figure 3. Somatic cancer mutations often disrupt cleavage/polyadenylation signals in tumour suppressor genes. **(A)** Stacked bar plot showing enrichment of single-nucleotide variants (SNVs) disrupting polyadenylation signals (DOWN-paSNVs) in tumour suppressors in cancer. **(B–C)** Overrepresentation of **(B)** tumour suppressors but not **(C)** oncogenes among genes with cancer somatic DOWN-paSNVs, as compared to genes with cancer somatic BG-paSNVs. Fractions of tumour suppressors and oncogenes are also shown for all genes and genes containing cancer somatic nonsense (premature stop codons), missense (altered amino acid residues) and synonymous (synonymous codons) mutations. Note that the enrichment of tumour suppressors is stronger for DOWN-paSNVs compared to nonsense mutations. **(C)** Top 10 GO Biological Process terms significantly enriched in genes with cancer somatic DOWN-paSNVs. Note the enrichment of apoptosis- and cell death-related functions. **(D)** GO Molecular Function terms significantly enriched in genes with cancer somatic DOWN-paSNVs.

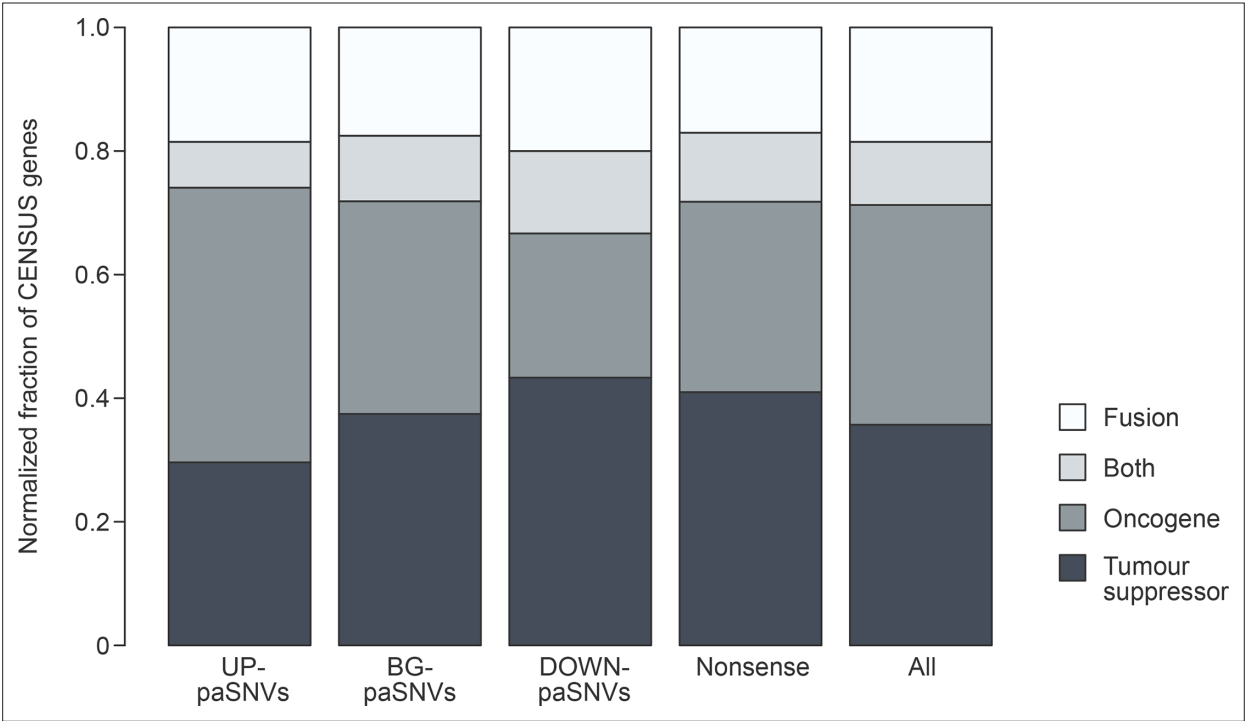


Figure 3—figure supplement 1. Cancer somatic DOWN-paSNVs often reside in genes with tumour-suppressive functions. Normalized stacked bar plot showing enrichment of single-nucleotide variants (SNVs) disrupting polyadenylation signals (DOWN-paSNVs) in tumour suppressors for Censur genes only. Note that nonsense mutations show a similar to DOWN-paSNVs enrichment in tumour suppressors, but not oncogenes. Conversely, UP-paSNVs are enriched in oncogenes but not tumour suppressors.

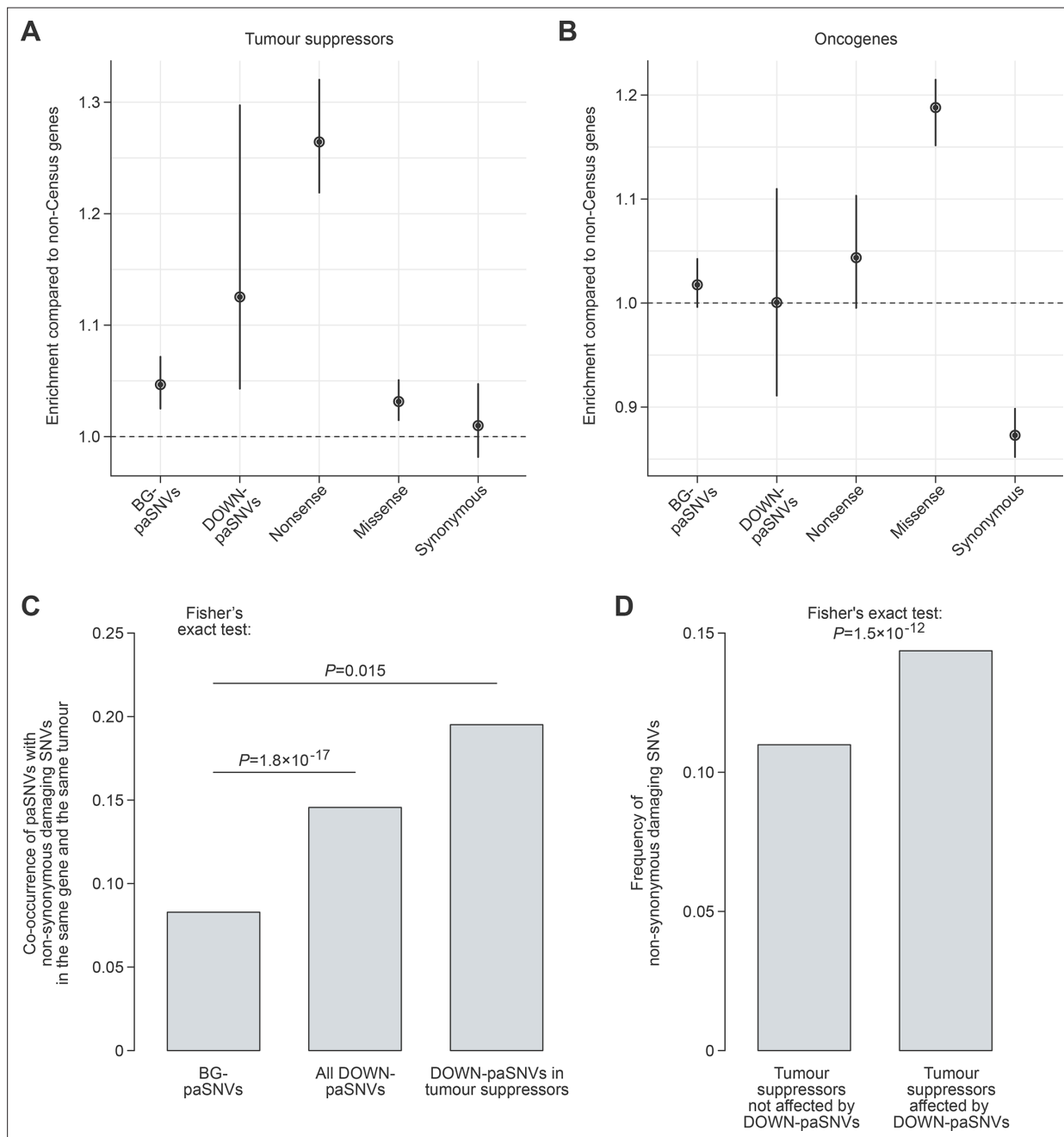


Figure 4. Disruption of cleavage/polyadenylation signals in tumour suppressors, along with other damaging mutations, may facilitate cancer progression. **(A–B)** Enrichment of different groups of cancer somatic single-nucleotide variants (SNVs) in **(A)** tumour suppressors and **(B)** oncogenes calculated using DigDriver relative to genes not listed in Cancer Census (non-Census) and presented with 95% confidence intervals. Note that DOWN-paSNVs and nonsense mutations are enriched in tumour suppressors but not in oncogenes. In contrast, oncogenes are often affected by missense mutations, as expected. **(C)** Cancer somatic DOWN-paSNVs co-occur in the same tumour with non-synonymous damaging SNVs, a group of somatic mutations defined in *Calabrese et al., 2020*, more often than BG-paSNVs. Note that the co-occurrence is particularly high for tumour suppressors. **(D)** The overall frequency of non-synonymous damaging SNVs is significantly higher in the DOWN-paSNV-containing group compared to the DOWN-paSNV-lacking group of tumour suppressor genes.

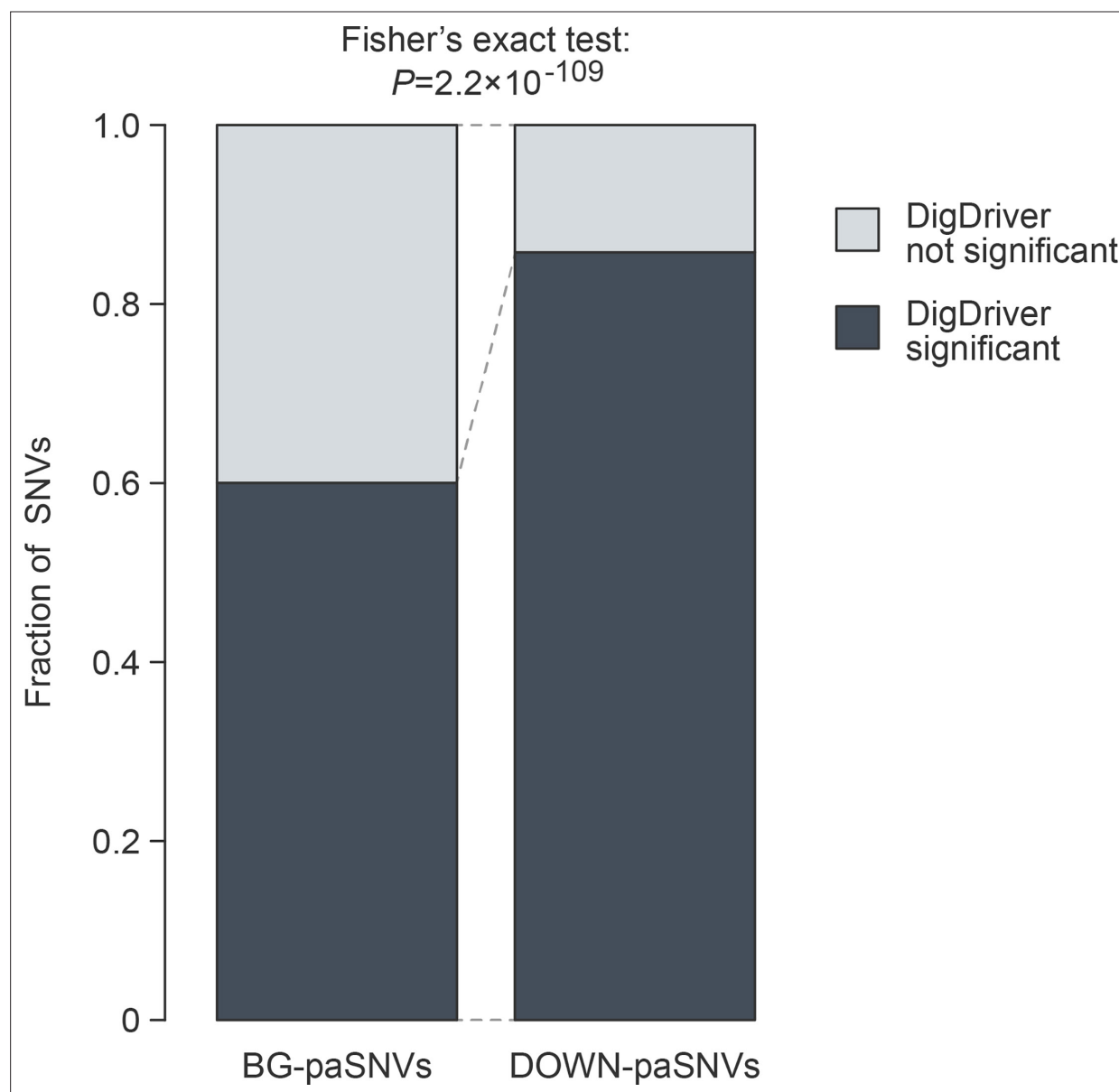


Figure 4—figure supplement 1. Cancer somatic DOWN-paSNVs are enriched for statistically significant DigDriver events (BH-adjusted $p<0.01$), suggesting that they may be under positive selection in cancer.

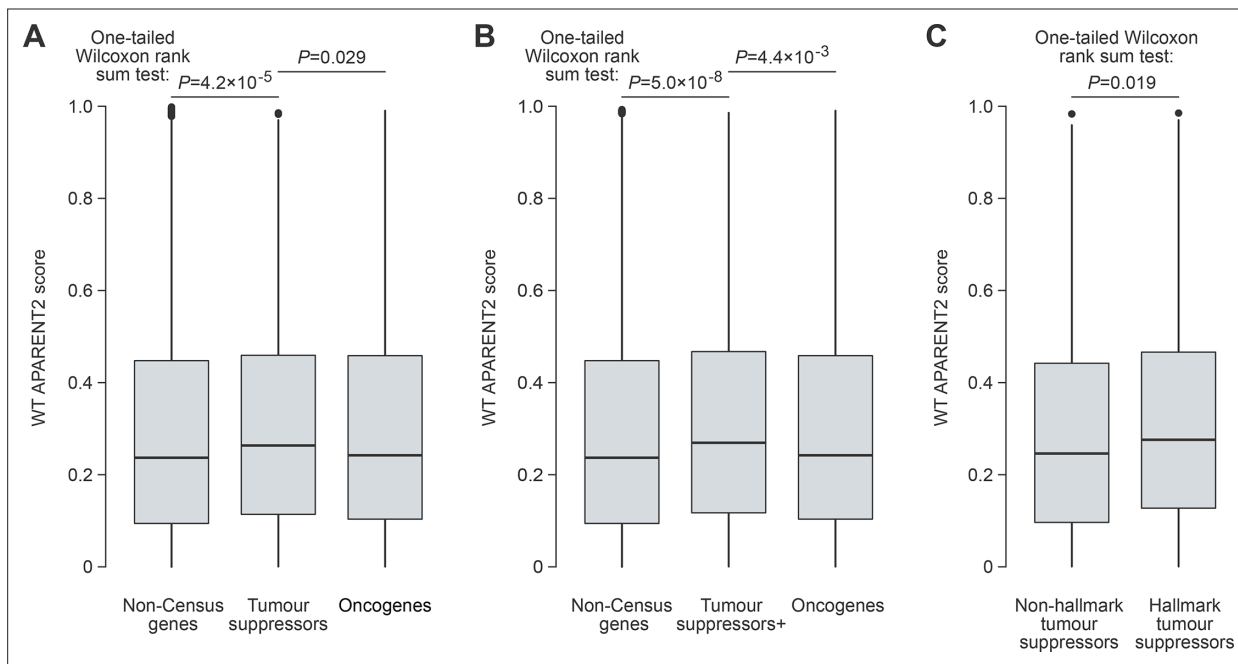


Figure 4—figure supplement 2. Wild-type tumour suppressor genes tend to have efficient cleavage/polyadenylation signals. **(A)** Box plot showing that wild-type tumour suppressors have stronger cleavage/polyadenylation signals than oncogenes and non-Census genes. **(B)** All Census genes classifiable as tumour suppressors ('Tumour suppressors+,' see Materials and methods) have stronger cleavage/polyadenylation signals compared to oncogenes and non-Census genes. **(C)** Tumour suppressors associated with 'hallmarks of cancer' have stronger cleavage/polyadenylation signals than 'non-hallmark' tumour suppressor genes.

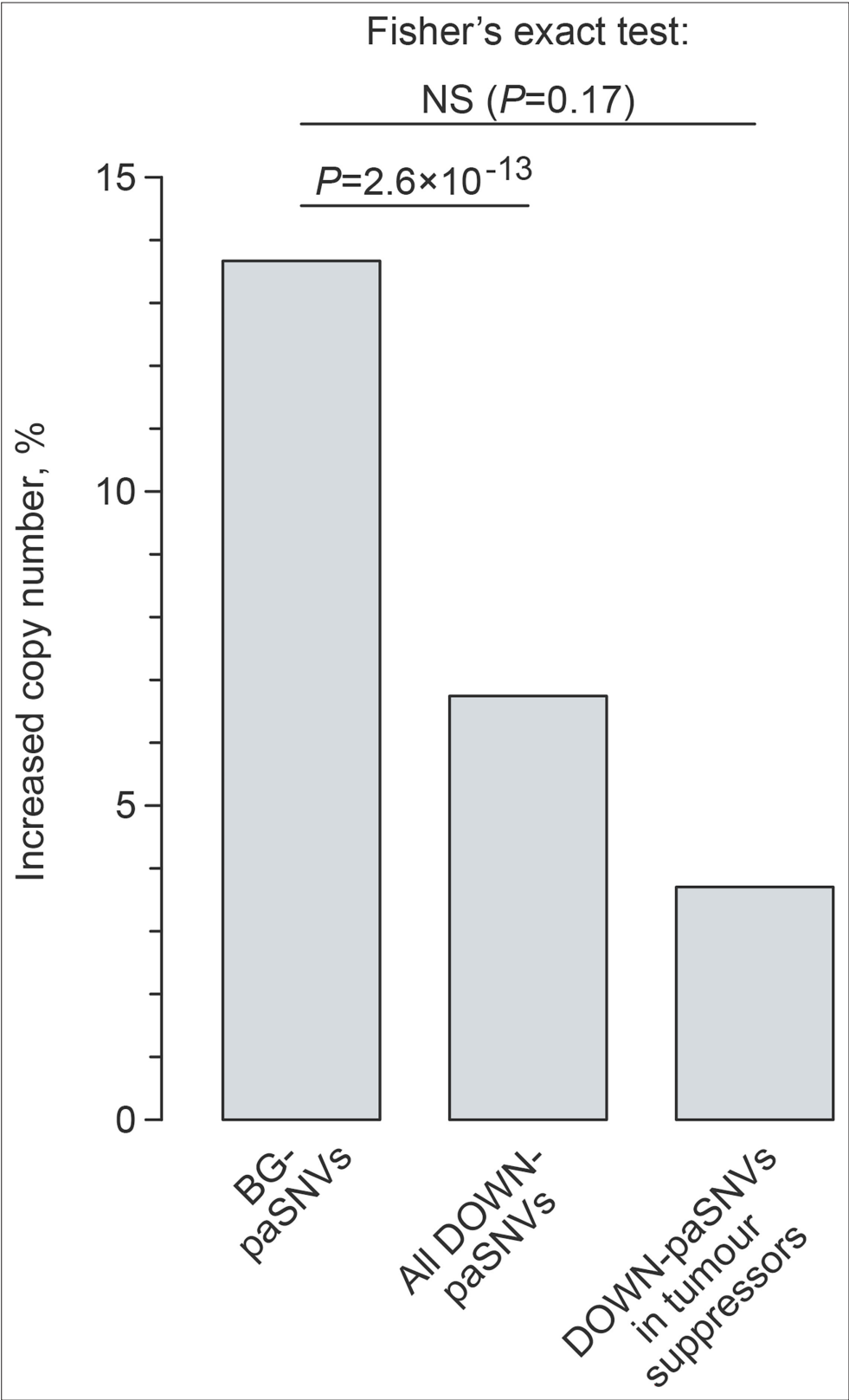


Figure 4—figure supplement 3. Reduced tendency for copy number increases in genes with DOWN-paSNVs. The graph illustrates the fraction of genes with copy number increases >3 in at least one cancer sample. Notably, copy number increases are significantly less frequent in cancer genes affected by DOWN-paSNVs compared to their BG-paSNV counterparts. The lack of statistical significance in the comparison between tumour suppressors

Figure 4—figure supplement 3 continued on next page

Figure 4—figure supplement 3 continued

with DOWN-paSNVs and BG-paSNV genes is probably due to the limited number of copy number variation (CNV) events in the former group.

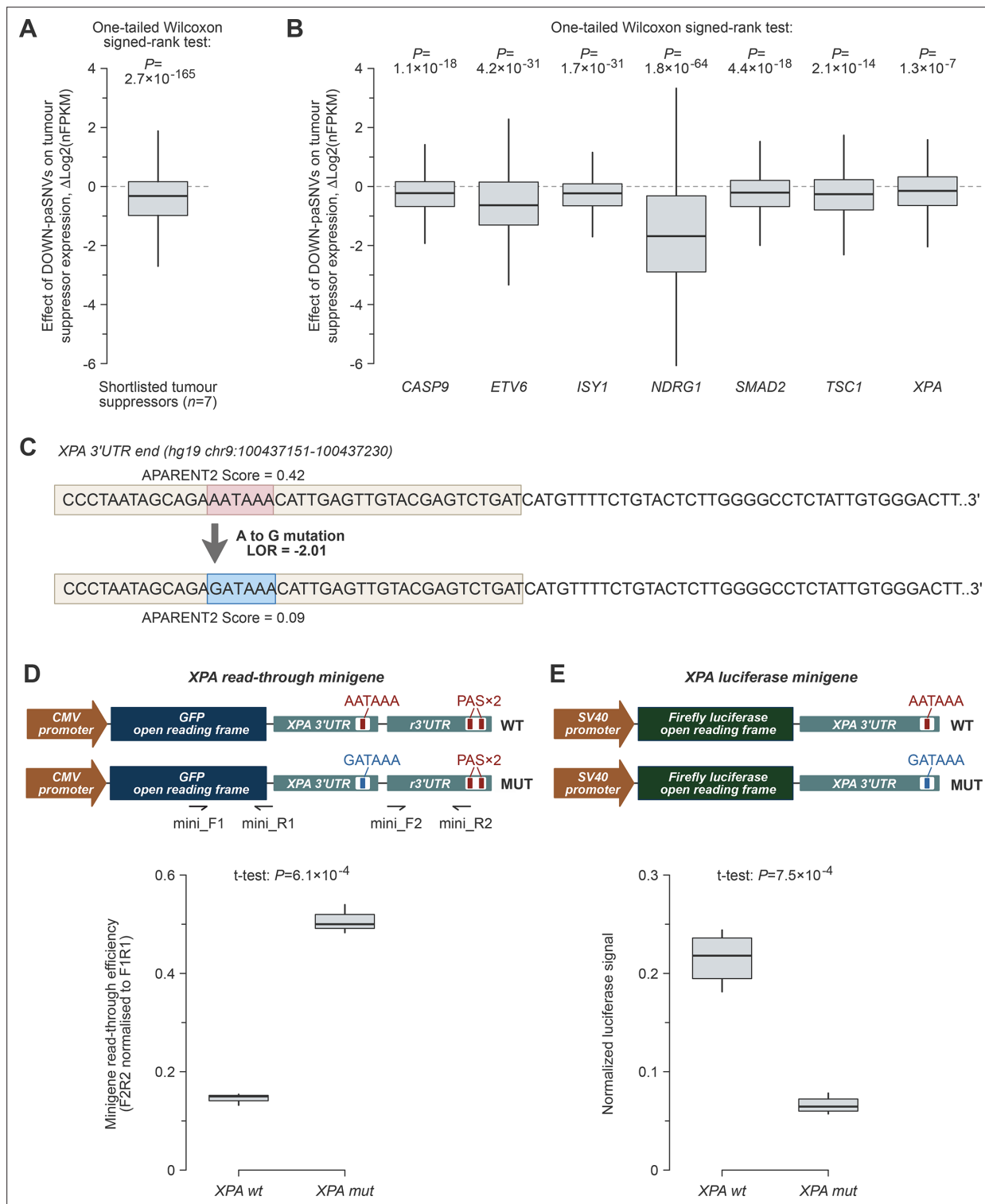


Figure 5. Somatic cancer DOWN-paSNVs are sufficient to downregulate tumour suppressor genes. **(A, B)** Gene-specific expression differences between DOWN-paSNV-containing and wild-type samples (ΔLog_2 of copy number variation-normalized FPKM values; see Materials and Methods) reveal a consistently negative effect of DOWN-paSNV on tumour suppressor mRNA abundance in colorectal cancers. Box plots are shown for **(A)** an aggregated set of qualifying tumour suppressors and **(B)** individual genes from this set. Outliers are omitted for clarity. **(C)** Wild-type and mutated sequences of the XPA tumour suppressor gene cleavage/polyadenylation signal. The polyadenylation signal (PAS) hexamer is enclosed within a box. **(D)** Top, XPA cleavage site read-through minigenes and corresponding primers used for RT-qPCR analyses. Bottom, RT-qPCR data showing stronger read-through (weaker polyadenylation) in the mutant minigene. **(E)** Top, luciferase expression minigenes. Bottom, luciferase assay reveals that the cancer-specific PAS mutation dampens the expression of the reporter gene.