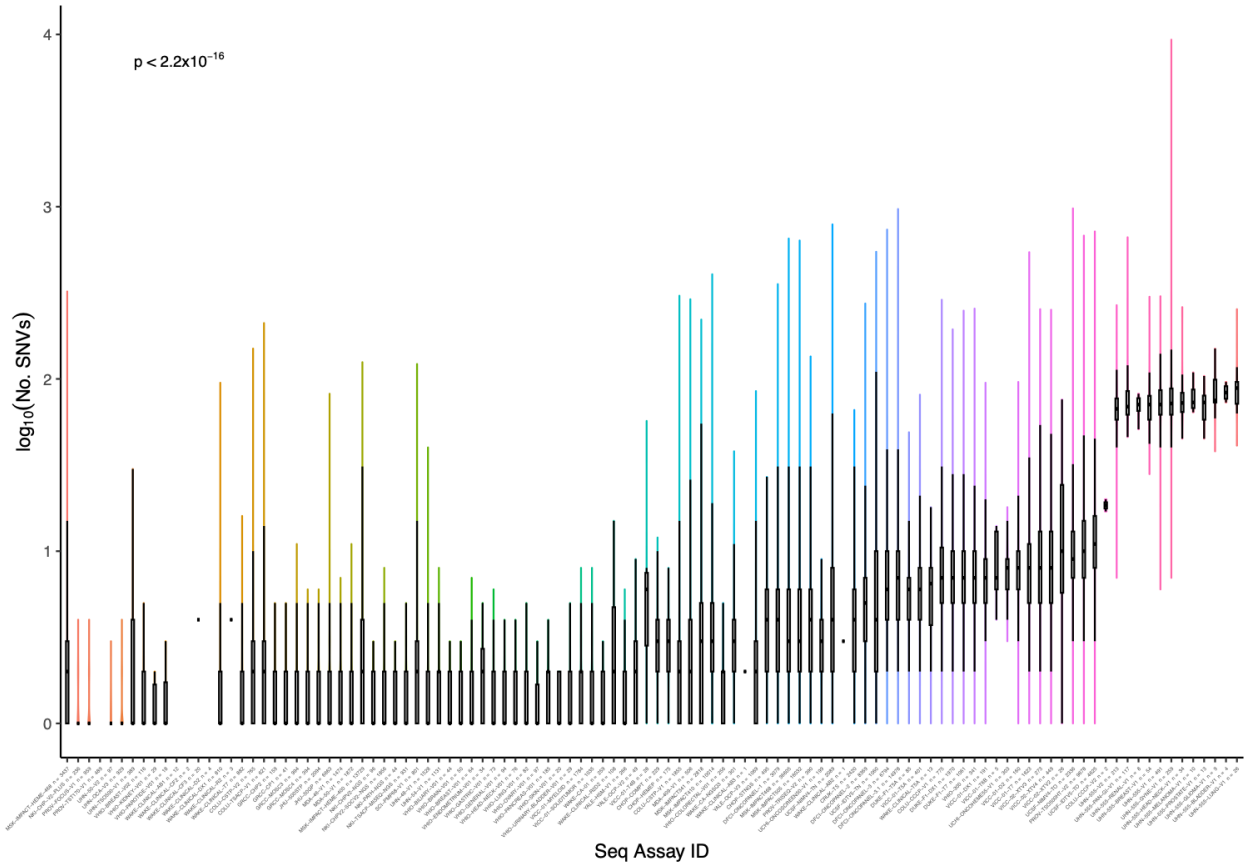

Figure S1. Log 10 of the number of SNVs each sequencing assay included in the GENIE dataset. The median (dot), 67% interval (Box plot edges) 95% (black lines) and the full range (colored lines) are shown.

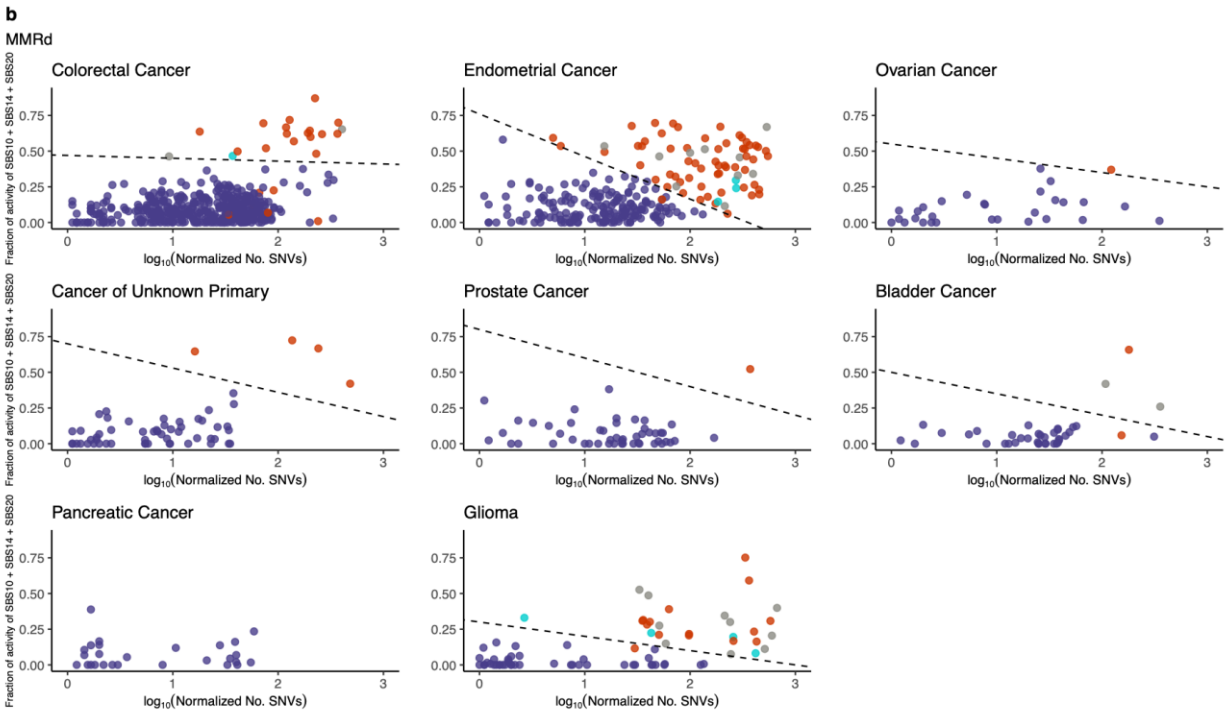

Figure S2. The clustering of the PPD status of the MMRd tumors separated per cancer type

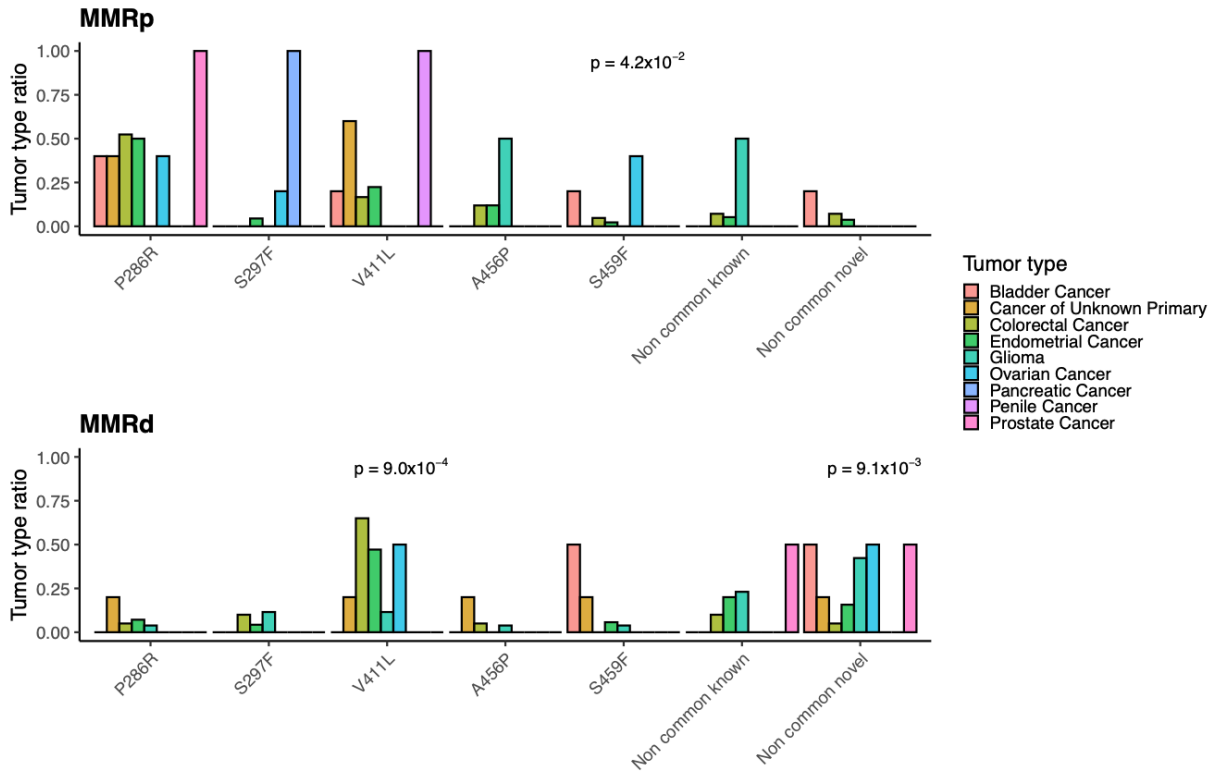

Figure S3. The fraction of tumors each variant, or group of variants has in each tumor type, is a fraction from all the tumors that have these variants.

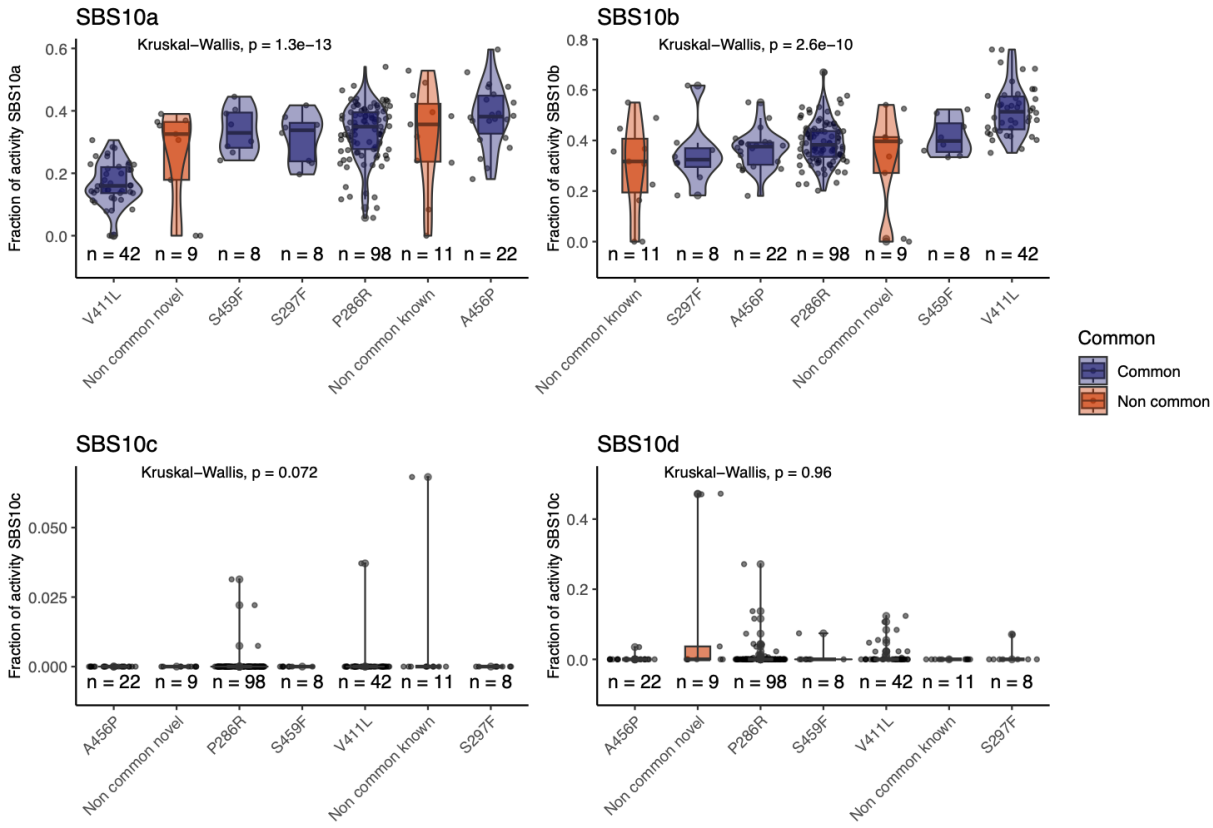

Figure S4. Log_{10} of the activity of each PPD signature by itself. The colors represent common mutations and non-common mutations. The x-axis is ordered according to each group median.

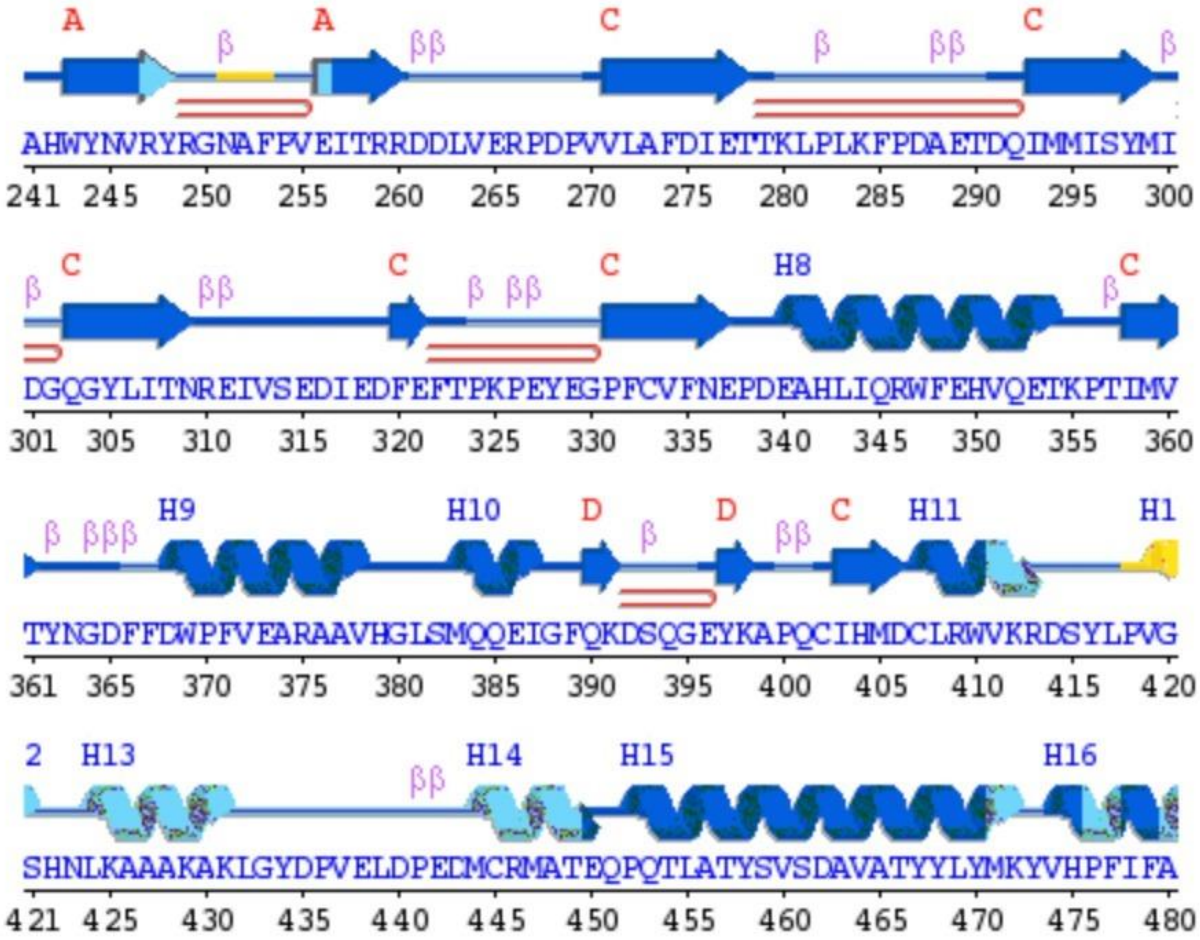

Figure S5. Secondary structure of POLE ExoD. As was generated by EMBL-EBI.

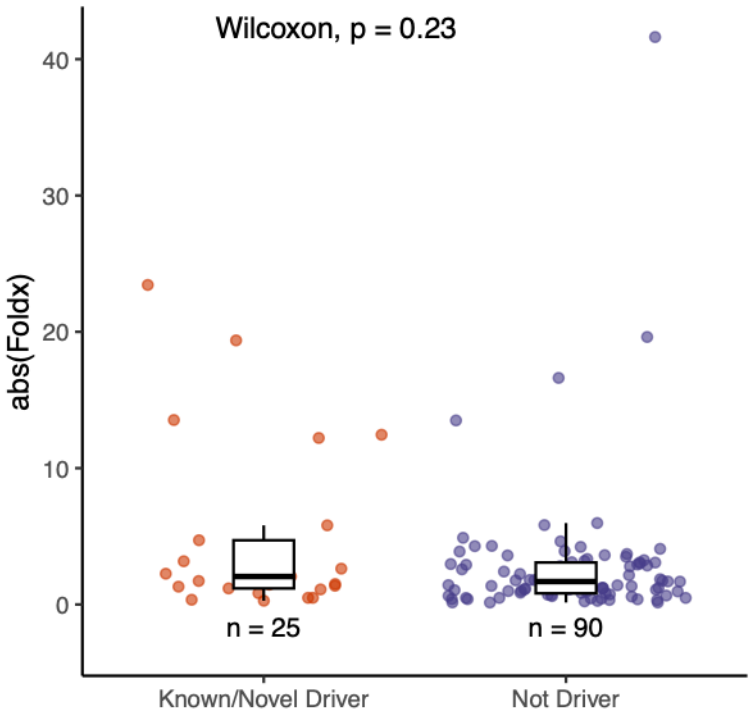

Figure S6. The absolute value of the Gibbes free energy changes due to each variant as predicted by the FoldX tool. The driver variants are compared towards non-drivers variants that were found in the GENIE database, that were up to 10AA far from the driver variants.

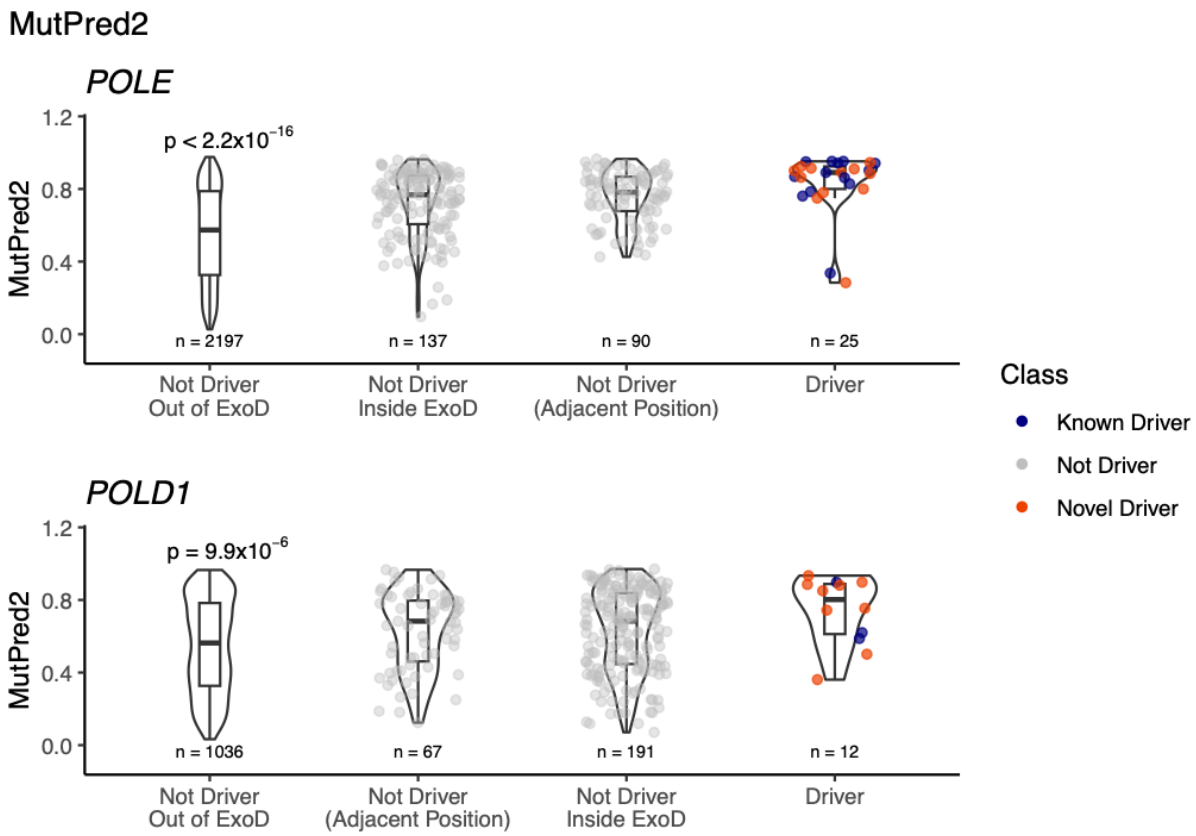

Figure S7. MutPred2 pathogenicity predictions for POLE and POLD1. Variants were grouped as: (i)PPD drivers, (ii)variants that are adjacent to PPD drivers, (iii) ExoD variants, and (iv) out of ExoD variants.

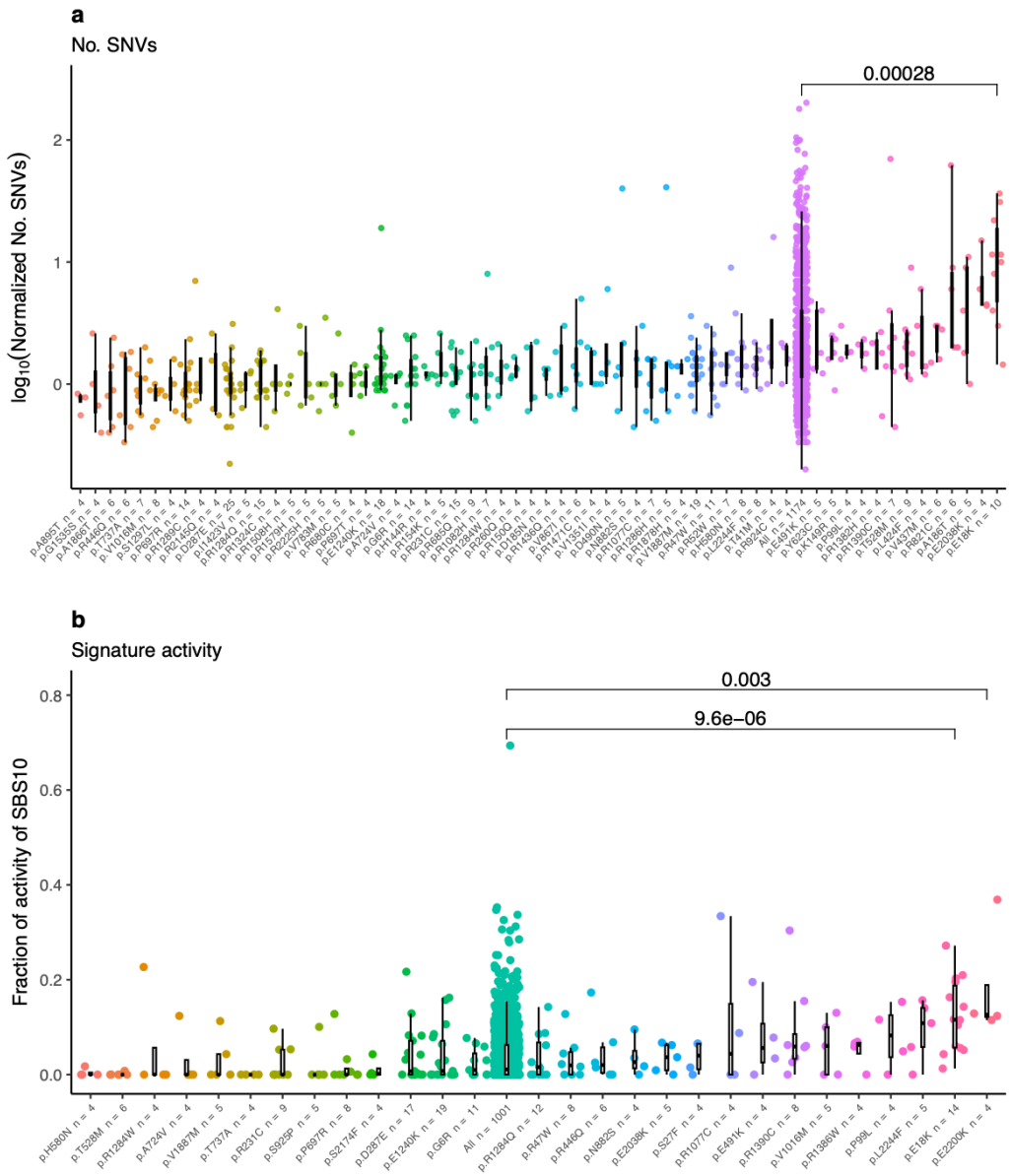

Figure S8 Non-driver hotspot mutations. a. Log_{10} of the normalized number of SNVs, for each sample that has a POLE hotspot mutation compared against all the non-driver which are non-hotpsot mutations. Tumor types with elevated mutation rate (Melanoma, Lung and Breast are excluded) B. The fraction of SBS10 activity for each sample that has a hot-spot mutation is compared against all the non-driver which are non-hotpsot mutations. Here all the tumor types are included. In both a&b Significant differences (q<0.1 Mann-Withey test, after BH FDR correction are presented).

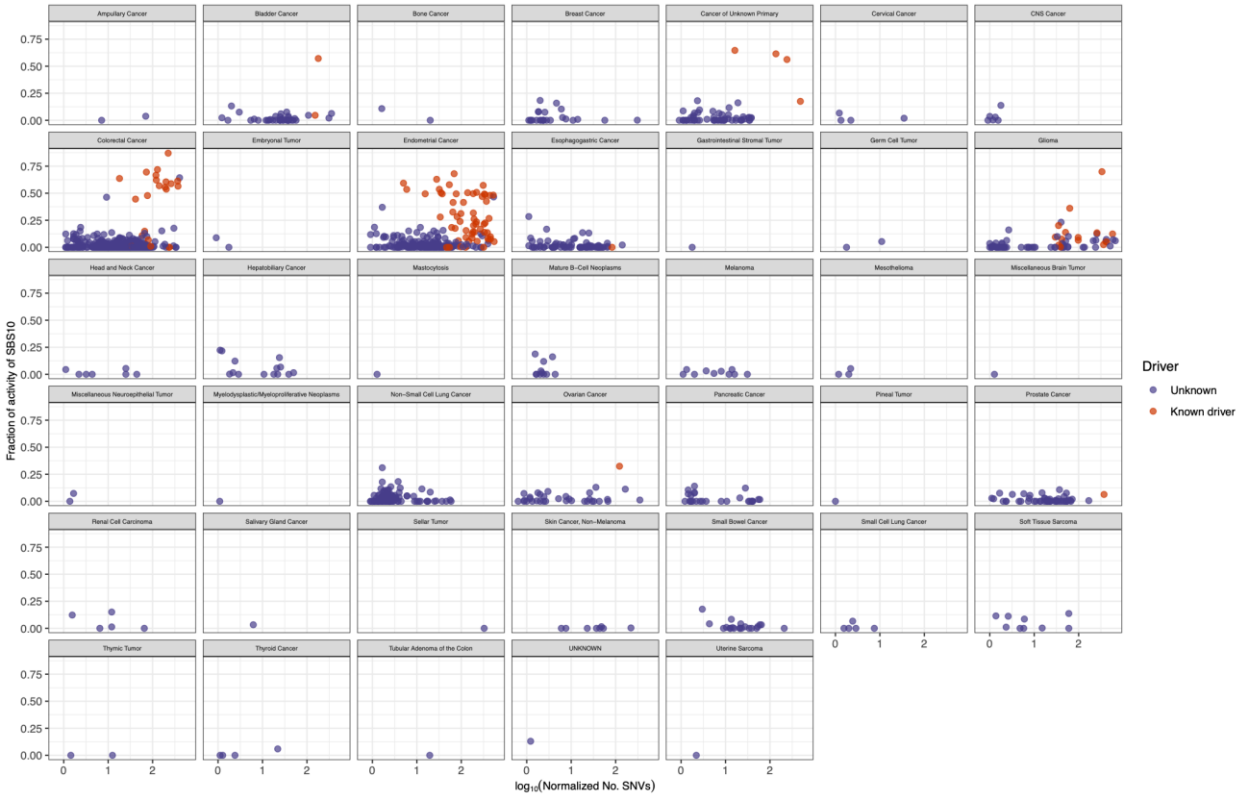

Figure S9a
The clustering of the PPD status of the MMRd tumors separated per cancer type (for all cancer types)

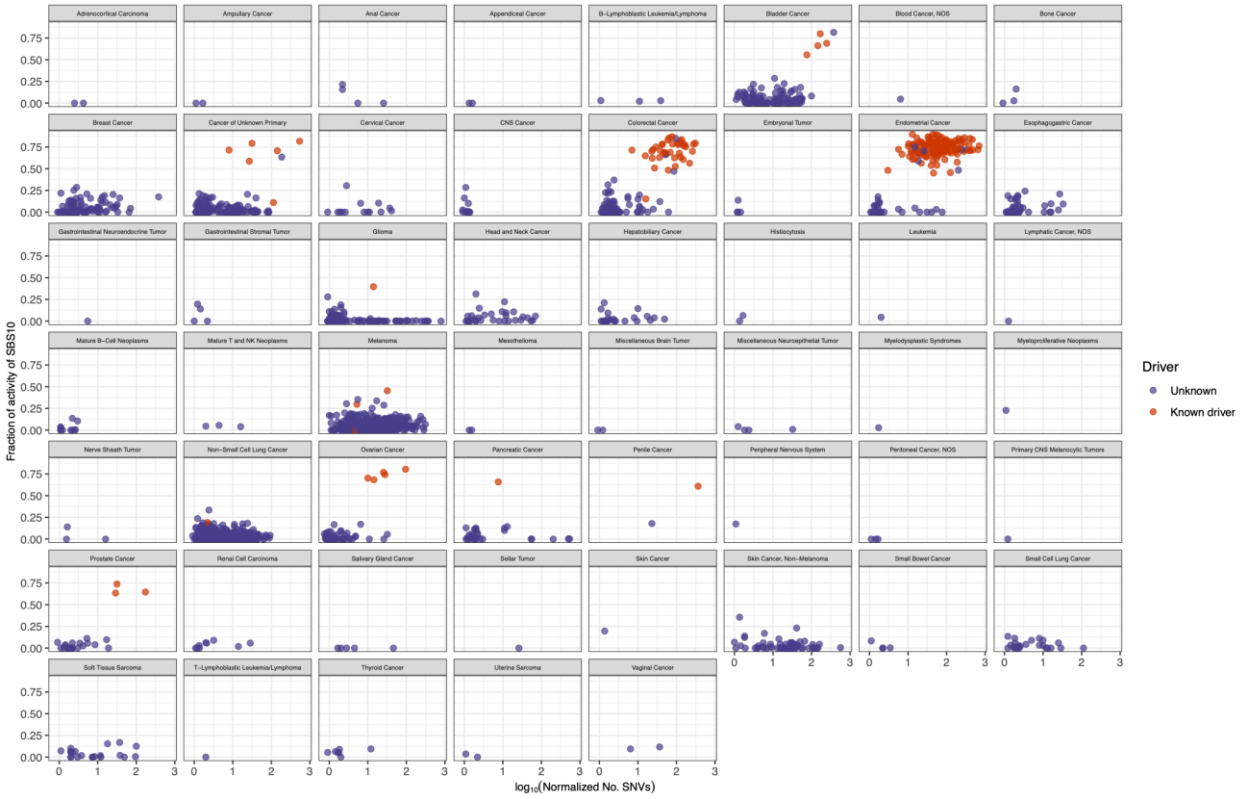

Figure S9b
The clustering of the PPD status of the MMRp tumors separated per cancer type (for all cancer types)

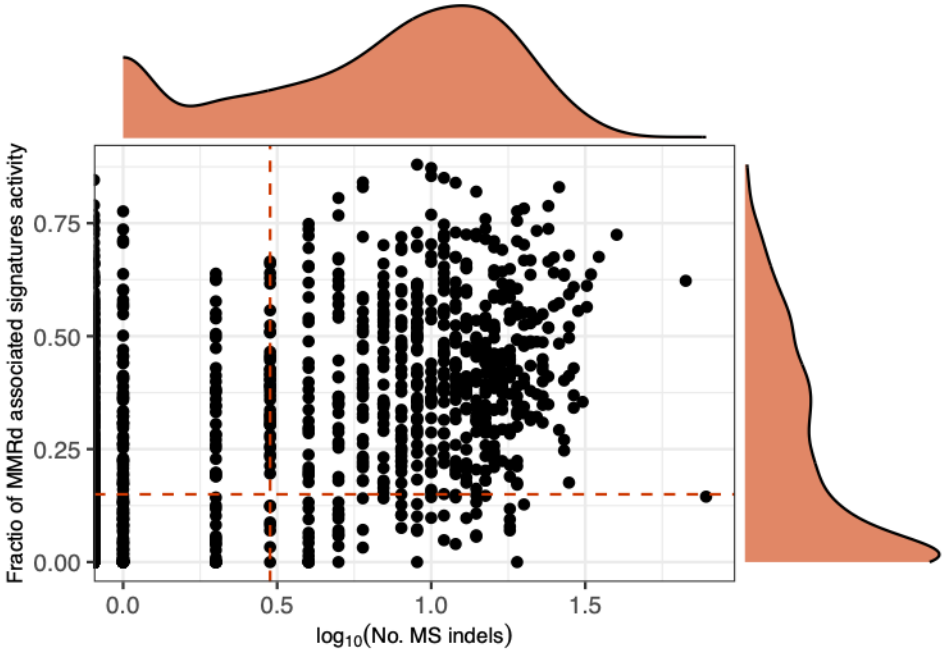

Figure S10. The log10 number MS-indels vs the fraction of MMRd signature activity each sample has. The distribution of the log10 number MS-indels and the fraction of MMRd signature activity is plotted on each side. The red dashed lines reflect the thresholds between MMRd and MMRp, and where chosen based on the shift in the distributions

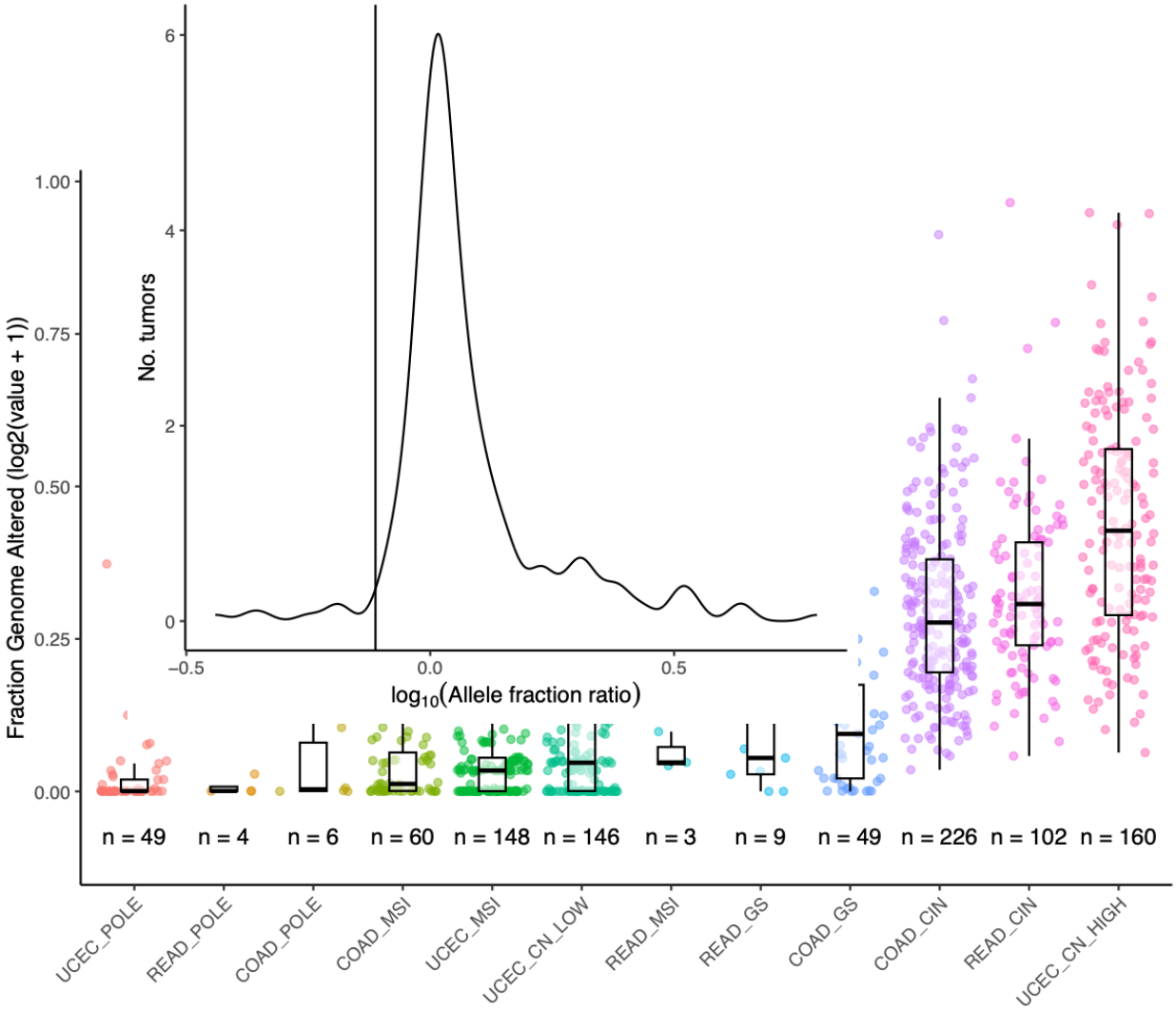

Figure S11a. The log10 of the fraction of the genome that is altered by copy number change, either deletion or amplification for each tumor, in the TCGA colon (COAD), rectal (READ) and endometrial (UCEC) cancers. Samples are grouped per molecular subtypes, and the x-axis is ordered by the median fraction of each group.

Figure S11b The distribution of the AF ratio between each common driver variants in a sample and its median AF in the GENIE dataset. The vertical line represents the threshold between clonal and sub-clonal variants.

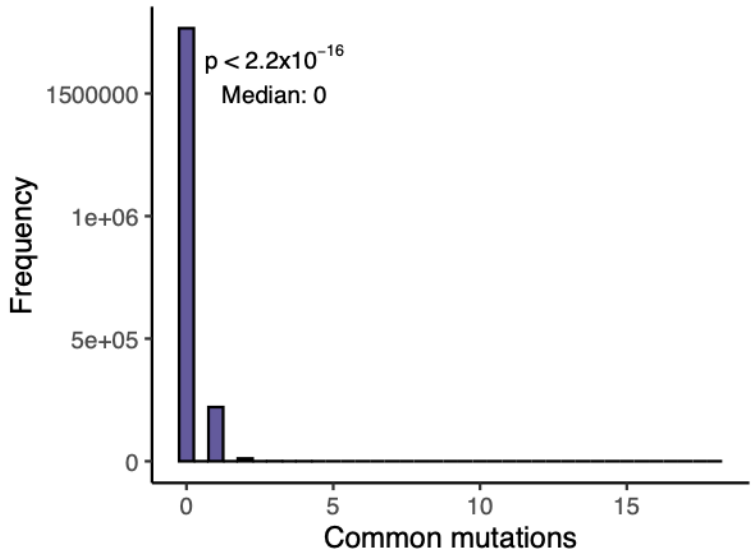

Figure S12. The distribution of the number of shared mutations between a set of 2000 unrelated CRC samples from the GENIE dataset.

Tables legends

Table S1. List of All POLE/D1 Variants Observed in the GENIE, TCGA, and CPC Datasets. Each variant is included with its previous classification as both somatic and germline variants, and the classification we obtained. Additionally, we provide MutPred2 and AlphaMissense pathogenicity predictions, along with FoldX ΔΔG predictions. We also include information on whether a variant is present in the GnoMad database as a germline variant and its allele fraction.

Table S2. Previous information about POLE/D1 driver variants based on ClinVar and literature review, both as somatic drivers and as germline drivers.

Table S3. P-values of pairwise comparison, related to Fig. 2b.

Table S4. List of genes that exhibit significantly different mutation rates between the P286R and V411L variants.

Table S5. The MMRd classification of the GENIE samples with a POLED1 variant and more than 10 SNVs.