# Supplementary figures legends

**Supplementary Figure X1. Mutations in 80 core DNA damage response genes across 108 tumours with microsatellite instability**

Pathogenic somatic and germline (grey dots) variants in MMR genes, *POLE,* and *POLD1* (rows; grouped by pathways, sorted by mutation rate) across 108 microsatellite instable tumours (columns). Variant classes predicted using snpEff [(Cingolani et al. 2012)](https://paperpile.com/c/RwXj5b/rIwJ). Top bars indicate the number of mono- and dinucleotide repeat indels. Figure is created using maftools [(Mayakonda et al. 2018)](https://paperpile.com/c/RwXj5b/Tx4x).

**Supplementary Figure X2. Mutations in MMR genes, *POLE*, and *POLD1* across 108 tumours with microsatellite instability**

Pathogenic somatic and germline (grey dots) variants in MMR genes, *POLE,* and *POLD1* (rows; grouped by pathways, sorted by mutation rate) across 108 microsatellite instable tumours (columns). Variant classes predicted using snpEff [(Cingolani et al. 2012)](https://paperpile.com/c/RwXj5b/rIwJ). Top bars indicate the number of mono- and dinucleotide repeat indels. Figure is created using maftools [(Mayakonda et al. 2018)](https://paperpile.com/c/RwXj5b/Tx4x).

**Supplementary Figure X3. Pathogenic mutations in multiple MMR genes associates with elevated numbers of mono- and dinucleotide repeat indels**

Number of indels in mono- and dinucleotide indels among tumours with one or more pathogenic mutations in MMR genes as well as *POLE* and *POLD1*, statistically explored by wilcoxon rank-sum test (un-paired, two-directional). We observed only a single tumor with mutations in five or more MMR genes, and did not statistically test these.

**Supplementary Figure X4. Correlation between per-tumour number mono- and dinucleotide repeat indels**

**a** Per-tumour (6,057 tumours) number of mononucleotide indels flanked by ≥3 bases similar to the deleted/inserted base (y-axis) correlate with no. of dinucleotide indels flanked by ≥2 di-mers similar to the deleted/inserted dimer (x-axis), and **b** the accumulated number of indels in mono- and dinucleotide repeat context (x-axis) correlates with the score obtained from msiseq analysis (y-axis). **c** Ratio of dinucleotide repeat indels (x-axis across >6,000 tumours (y-axis).

**Supplementary Figure X5. Rate of mono- and dinucleotide repeat sequence is preserved across genes**

**a** Per-gene (735 genes**)** number of bases in mono- and dinucleotide repeat sequences stretching three or more bases (x-axis; log10-transformed) shows **b** linear correlation (blue line) with the size of the gene (y-axis; log10-transformed). **c** The ratio of mono- and dinucleotide repeat sequences (x-axis) varies between genes.

**Supplementary Figure X6. Genome-wide ploidy changes do not explain the number of mono- and dinucleotide repeat indels**

**a** Genome-wide ploidy summarized by the median ploidy across all bases of the genome (x-axis)across ~6,000 tumors (y-axis), and **b** mapped against the per-tumour number of indels in mono- and dinucleotide repeat indels (y-axis), for tumours with microsatellite stability (MSS; blue) and microsatellite instability (MSi; red).

**Supplementary Figure X7. Indels are enriched in repeat sequences shorter than 15 bases**

**a** Number of mono- and dinucleotide repeat indels across 735 DDR genes in 6,057 tumours (y-axis) divided by the length of the repeat sequence they occur in (x-axis; including repeat sequences of ≥3 bases). **b** Number of repeat sequences (≥3 bases) across 735 DDR genes in the hg19/GRCh37 genome assembly (x-axis) against the length of each repeat sequence (x-axis, same as **a**). Bethesda panel marker regions annotated in red.

**Supplementary Figure X8. Tumours with microsatellite instability show increased rates of pathogenic mono- and dinucleotide repeat indels**

Number of genes (out of 735 DDR genes; y-axis) divided by the rate of pathogenic events that are mono- and dinucleotide repeat indels (x-axis). Bottom lines indicate the rate across MMR related genes (*PMS2, MSH3, MSH6, MLH1, MSH2, PMS1, MLH3, POLE, EXO1, POLD1*; red)