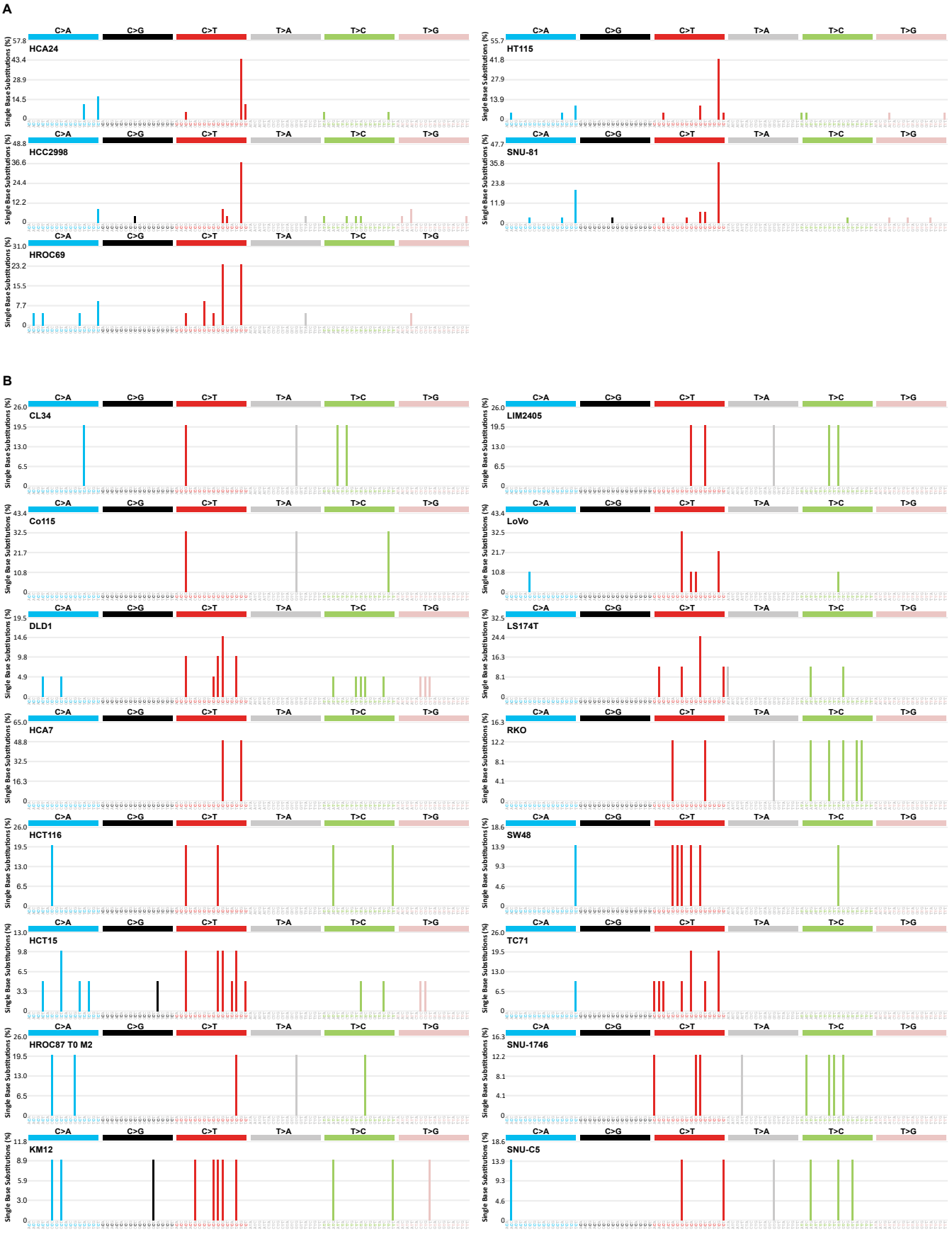
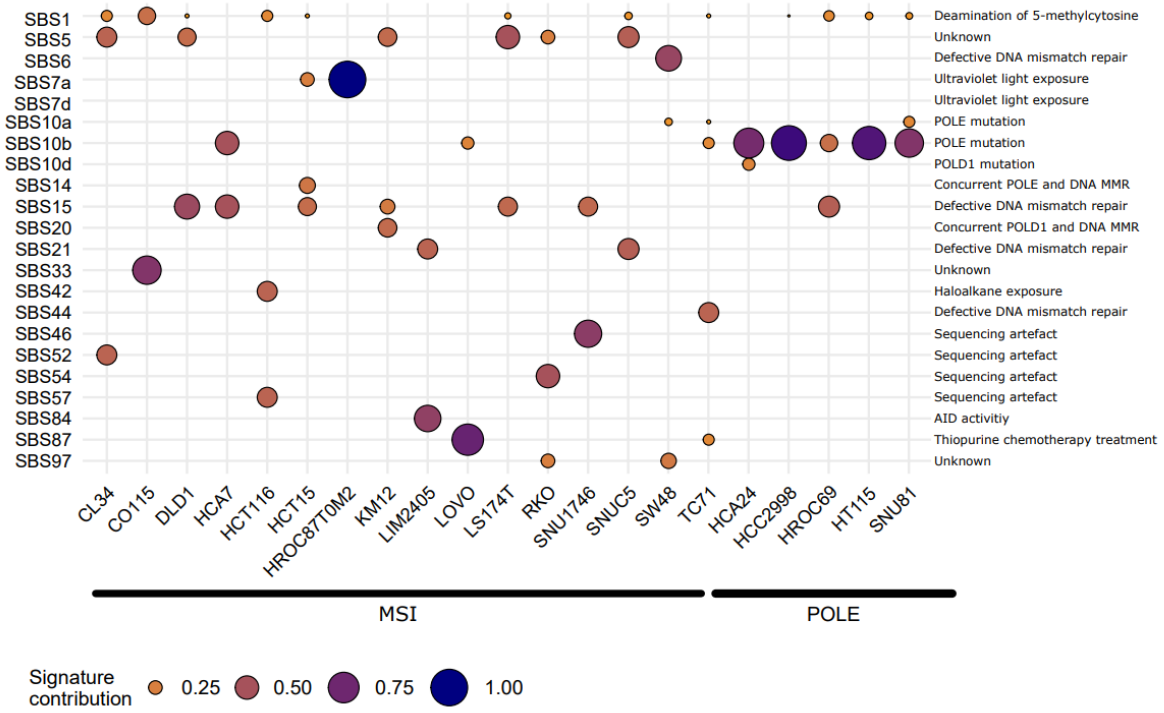
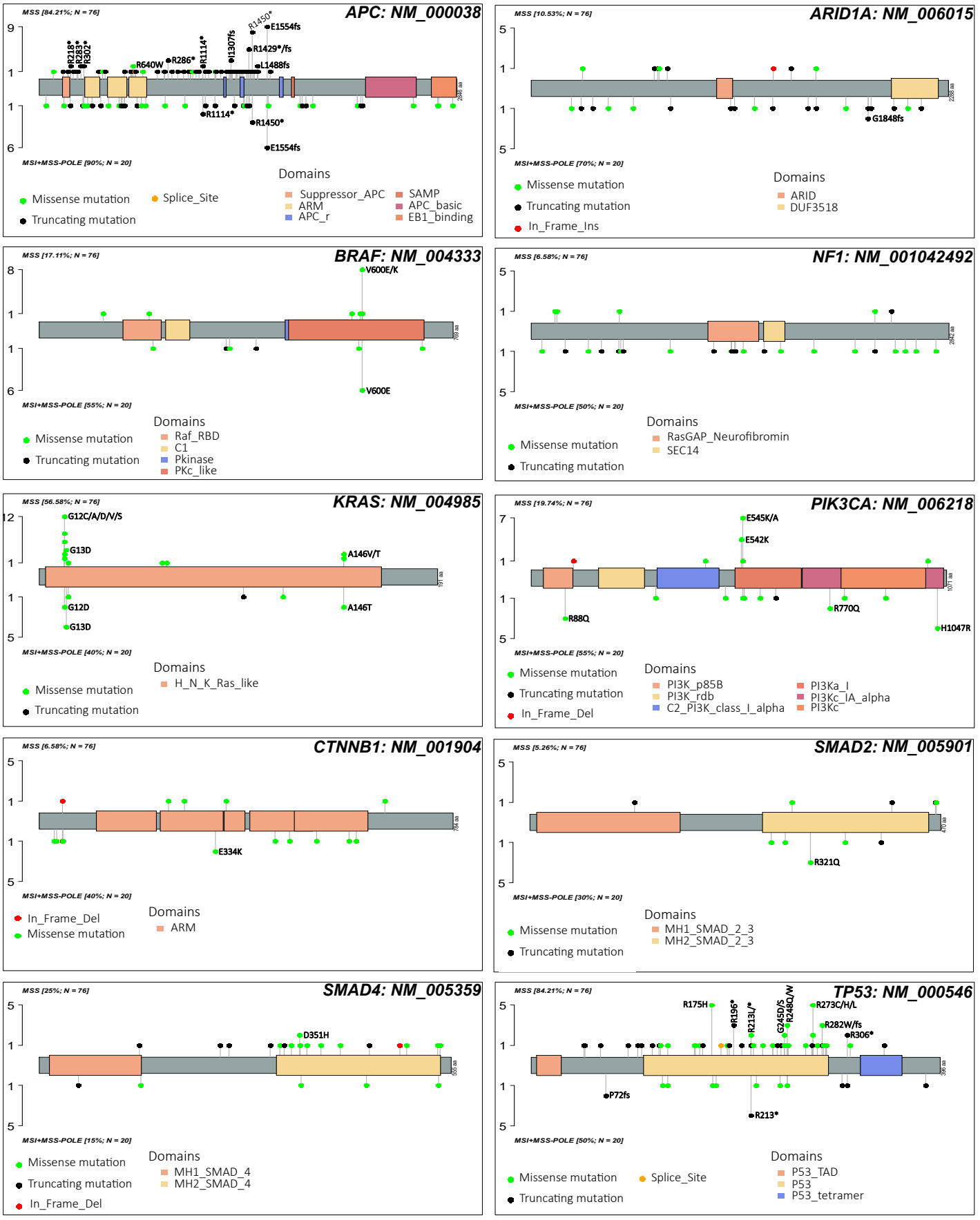


Supplementary Figure 1. Mutational signatures associated with hypermutation

phenotypes. Frequencies (vertical axes) of base substitutions plotted for each cell line with pathogenic *POLE* mutations (n = 5) **A** or MSI (n = 16) **B** according to substitution type (indicated in the top panels and with color codes) and sequence context (flanking nucleotides are indicated on the horizontal axes).

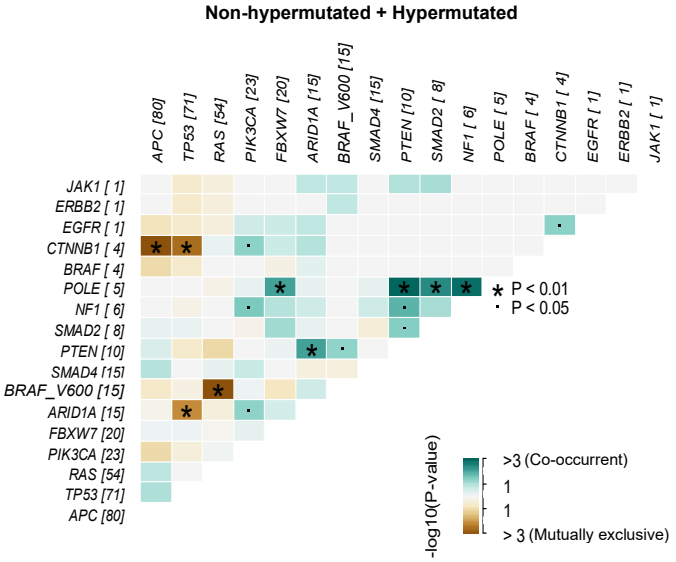


Supplementary Figure 2. Mutational signature contributions in hypermutated cell lines. Bubble plot showing the contribution of the listed single base substitution (SBS) signatures (left side) to the SNV load of each of the MSI and MSS *POLE* mutated cell lines. The proposed aetiology for each signature is shown on the right side.

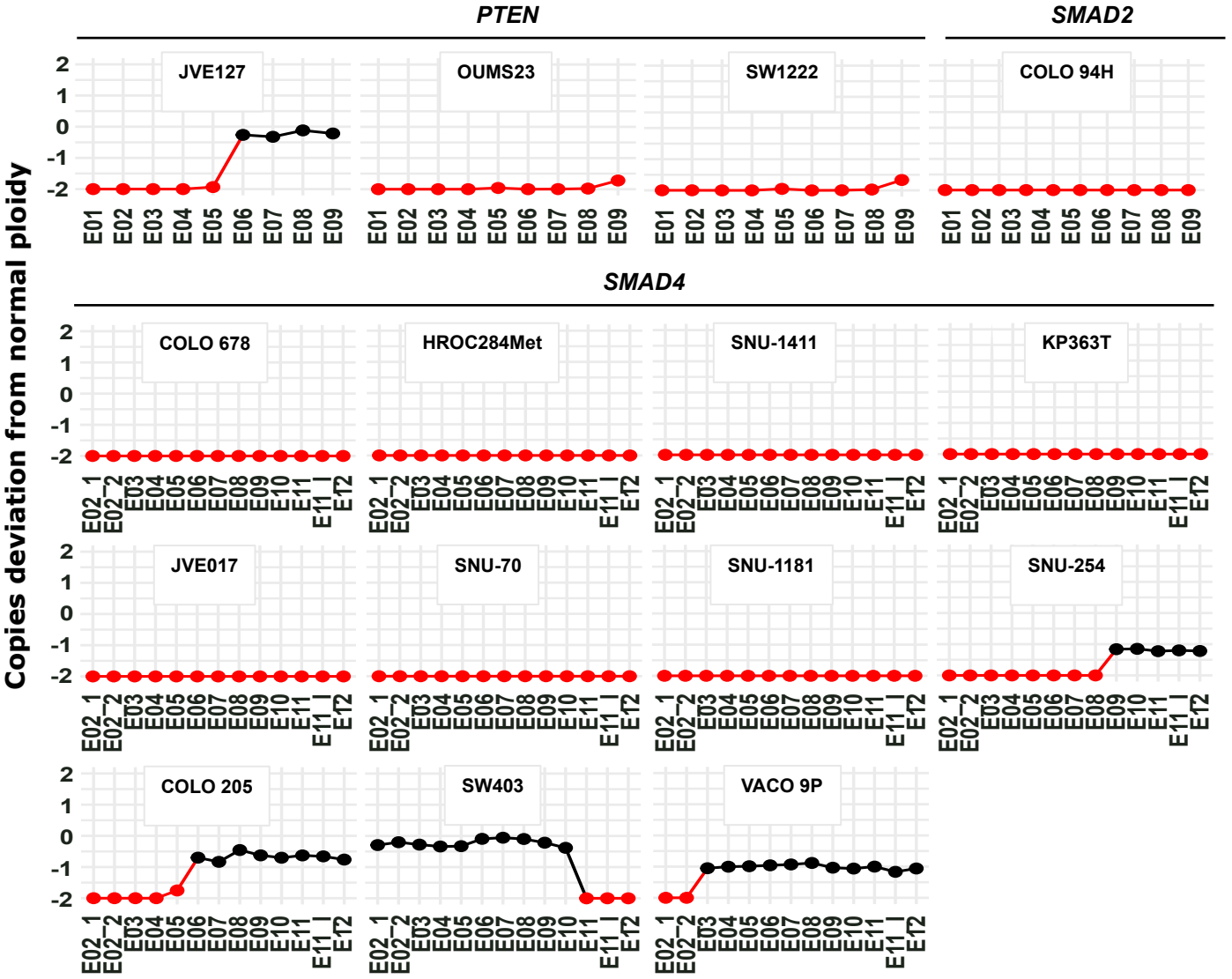


Supplementary Figure 3. Distribution of non-synonymous mutations along the encoded

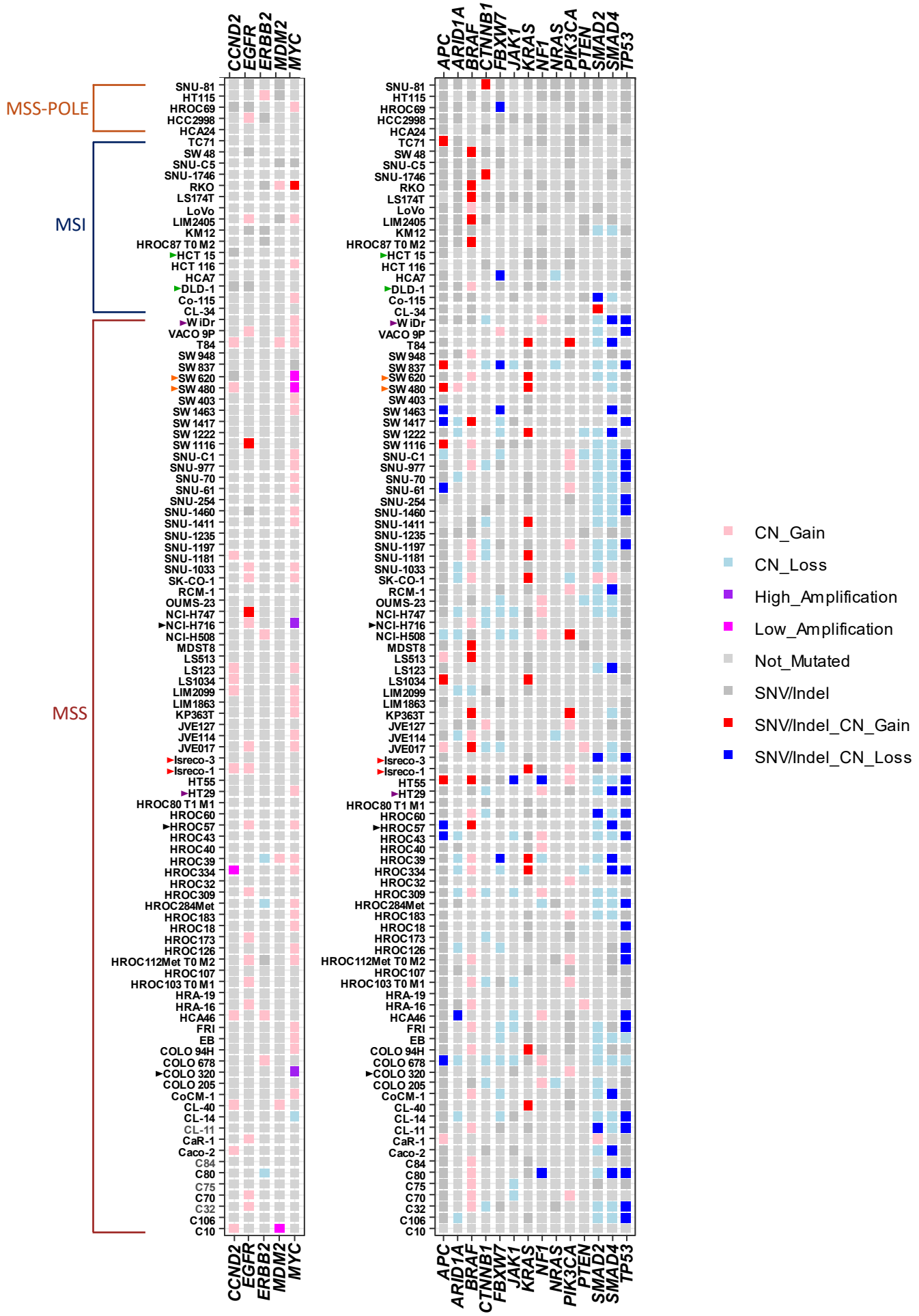
protein sequence of frequently mutated genes. Lollipop plots of mutations in non-hypermutated cell lines (MSS; top) and hypermutated cell lines (MSI+MSS-POLE; bottom) for each of ten frequently mutated genes. Mutations are colored according to the mutation type, as indicated in each plot. Recurrent mutations occurring in 2 or more samples are specified. Protein domains are color-coded and described for each gene (Supplementary Table 10). Mutations in cell lines derived from non-unique patients (DLD-1, Isreco-3, WiDr, SW620) and in the neuroendocrine cell lines (COLO 320, HROC57, NCI-H716) are not included. MSI/MSS: microsatellite instable/stable; MSS-POLE: *POLE* mutated samples.



Supplementary Figure 4. Somatic interactions of pathogenic mutations across phenotypes. Heatmap of significance levels for mutation interactions between gene pairs in hypermutated and non-hypermutated cell lines (*n* = 96) together. Co-occurrence versus mutually exclusivity is color-coded. Statistical significance of co-occurrence and mutually exclusivity was determined as described in Figure 2. Significant events are marked with asterisks and dots as indicated. Numbers next to the gene names correspond to the number of mutated samples. The *KRAS* and *NRAS* genes were grouped as *RAS* and mutations targeting the *BRAF* V600 hotspot (p.600V>E/K) were grouped separately from other *BRAF* mutations. Cell lines derived from non-unique patients (DLD-1, Isreco-3, SW620 and WiDr) and neuroendocrine cell lines (COLO 320, HROC57, NCI-H716) were excluded from the analysis. MSI/MSS: microsatellite instable/stable; MSS-POLE: *POLE* mutated samples.

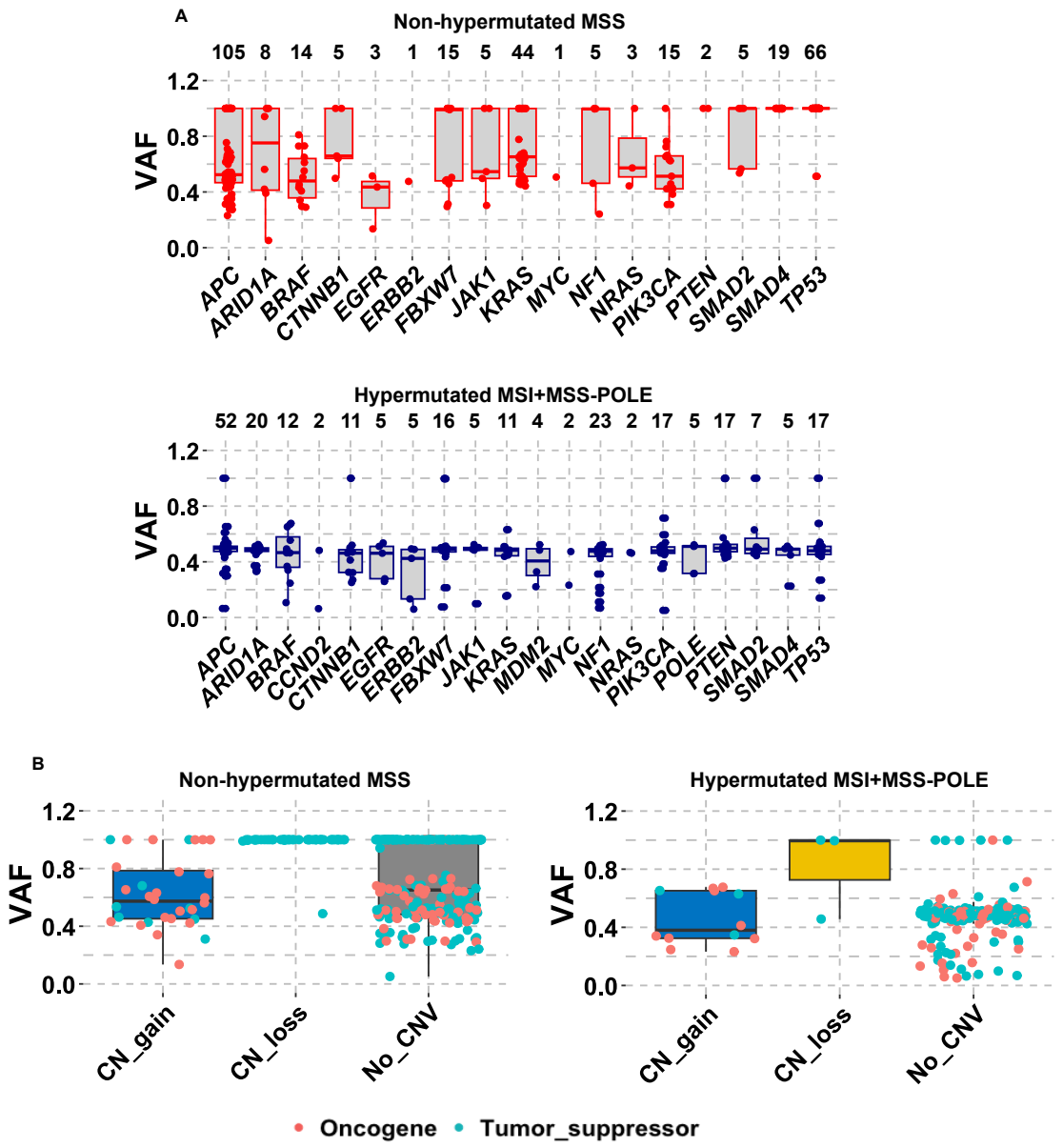


Supplementary Figure 5. Overview of homozygous gene deletions in the gene panel. The copy number deviation from the normal diploid genome (normal ploidy = 0 on y-axis) is displayed for *PTEN* and *SMAD2* (top four panels) and *SMAD4* (bottom set of panels) in cell lines for which partial or complete homozygous deletions have been detected (copy numbers close to -2, range -1.84 to -1.99, in Supplementary Table 12). The dotted line in each plot corresponds to the exons (E) annotated for each gene and shown on the x-axis. Red segments mark the gene areas affected by the homozygous deletions.



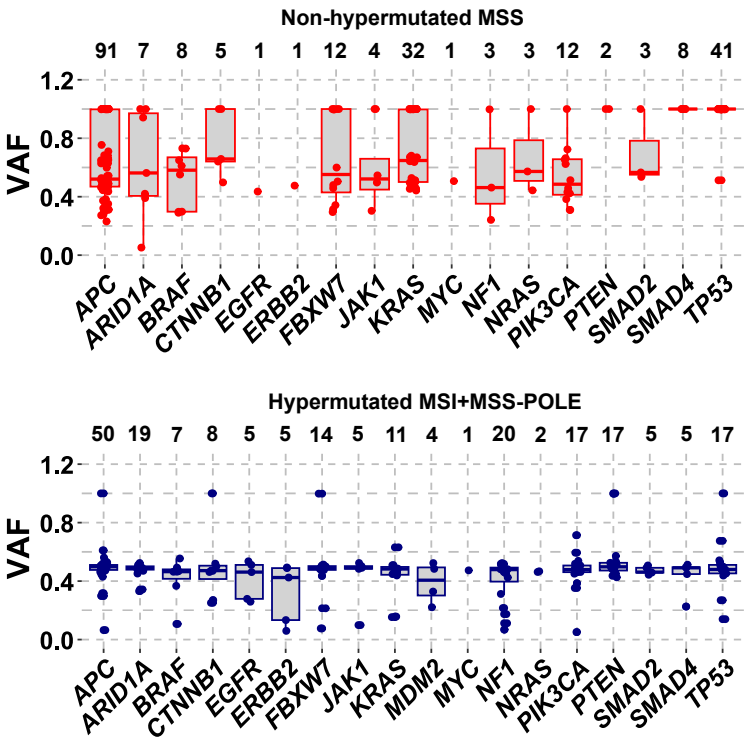
Supplementary Figure 6. Overview of concurrent DNA copy number alterations and

SNVs/indels in the gene panel. Copy number alterations detected in *MYC*, *CCND2*, *EGFR*, *MDM2* and *ERBB2* (left) and the rest of the gene panel (right) are color-coded according to the type of aberration and whether they co-occur with SNVs/indels in each sample. Cell lines are ordered alphabetically within each hypermutation phenotype. Cell lines derived from the same patient (DLD-1:HCT 15, Isreco-1:Isreco-3, HT-29:WiDr, SW480:SW620) are marked by arrowheads of the same color. Cell lines derived from neuroendocrine tumors (COLO 329, HROC57, NCI-H716) are marked with a black arrowhead. MSI/MSS: microsatellite instable/stable; MSS-POLE: *POLE* mutated samples; CN_Gain: copy number gain; CN_Loss: copy number loss; SNV: single nucleotide variant; Indel: insertion or deletion.



Supplementary Figure 7. Variant allelic frequencies across the gene panel. A Boxplots of the VAF of SNVs/indels in each gene plotted separately for non-hypermutated (MSS; top) and hypermutated (MSI+MSS-POLE; bottom) samples. Each dot in the plots represents a mutation and the number of mutations of each gene are indicated on the top. Mutations in cell lines derived from non-unique patients (DLD-1, Isreco-3, WiDr, SW620) and in the neuroendocrine cell lines (COLO 320, HROC57, NCI-H716) are not included. **B** Boxplot of the VAF of mutations according to the DNA copy number status of the mutated gene plotted separately for non-hypermutated (MSS; left) and hypermutated (MSI+MSS-POLE; right)

samples. Genes are classified as tumor-suppressors or oncogenes as described in Figure 1. Mutations in cell lines derived from the same patient (DLD-1, Isreco-3, WiDr, SW620) and in the neuroendocrine cell lines (COLO 320, HROC57, NCI-H716) are not included. MSI/MSS: microsatellite instable/stable; MSS-POLE: *POLE* mutated samples; CN_gain: copy number gain; CN_loss: copy number loss; No_CNV: no copy number variation.



Supplementary Figure 8. Variant allelic frequencies in genes not affected by CNVs. Boxplots of VAFs for mutations not associated with CNVs in each gene were plotted separately for non-hypermutated (MSS; top) and hypermutated (MSI+MSS-POLE; bottom) samples. Each dot in the plots represents a mutation and the number of mutations for each gene are indicated on the top. Mutations in cell lines derived from non-unique patients (DLD-1, Isreco-3, WiDr, SW620) and in the neuroendocrine cell lines (COLO 320, HROC57, NCI-H716) are not included. MSI/MSS: microsatellite instable/stable; MSS-POLE: *POLE* mutated samples.