## Supplementary material and methods

### Copy number assessment by shallow whole genome sequencing

First, all relative copy number plots were assessed for the presence of any degree of somatic copy number alterations (SCNAs). Cases without any SCNAs were assigned 'undetectable copy number alterations' (supplementary figure S1). For these cases, strict quality control criteria were applied to exclude that the latter was related to insufficient input (tumor cell percentage and DNA concentration) or quality of the case (number of reads). Regarding tumor cell percentage, the variant allele frequency of established driver mutations, i.e., *TP53* mutations, or *PTEN*, *PIK3CA, KRAS* in the absence of a *TP53* mutation in NSMP controls, was used as a surrogate of tumor cell percentage. Samples with a *TP53* VAF less than 30% were considered ineligible for reliable copy number assessment. For cases without a driver mutation, the tumor cell percentage was estimated by estimated by a pathologist (TB) on the representative H&E slide. In addition, cases with undetectable SCNAs with a measured DNA concentration ≤3 ng/μL and cases with less than one million reads were excluded from further copy number analysis. Next, we manually counted the number of SCNAs (gains and/or losses) for each case by visual inspection of the generated relative copy number plots. We subsequently divided all cases into three predefined categories based on the number of SCNAs (e.g. <5, 5-10, and >10). Representative examples of cases with undetectable SCNAs, cases with <5 SCNAs, and cases with >10 SCNAs are presented in Supplementary Figure S1.

Additionally, for cases with detectable copy number alterations, a purity/ploidy-solution was predicted, as described previously[1]. Briefly, the Rascal tool (relative to absolute copy number scaling) was used for absolute copy number fitting. The possible purity/ploidy-solutions were ranked based on the smallest distance (i.e., mean absolute deviation (MAD) applied to segments). A surrogate for tumor cell percentage was used to select the most suitable solution, e.g. the VAF of a driver gene or estimated purity by a pathologist. If both surrogates were absent, manual copy number fitting was performed. The predicted DNA ploidy was used for further analyses.

1 Sauer, C. M., Eldridge, M. D., Vias, M., Hall, J. A., Boyle, S., Macintyre, G. et al. Absolute copy number fitting from shallow whole genome sequencing data. bioRxiv 10.1101/2021.07.19.452658, 2021.2007.2019.452658 (2021).