

SUPPLEMENTARY INFORMATION

Whole genome sequencing of acral melanoma reveals genomic complexity and diversity, Newell *et al*

Supplementary Figure 1: Associations of clinicopathological features.

Supplementary Figure 2: Rearrangement and CNV signature probabilities and cosine similarity.

Supplementary Figure 3: Associations with aneuploidy

Supplementary Figure 4: Associations with localized complex genomic rearrangements

Supplementary Figure 5: Examples of localized complex events on recurrent chromosomes

Supplementary Figure 6: Recurrent linking translocations

Supplementary Figure 7: Significantly mutated genes

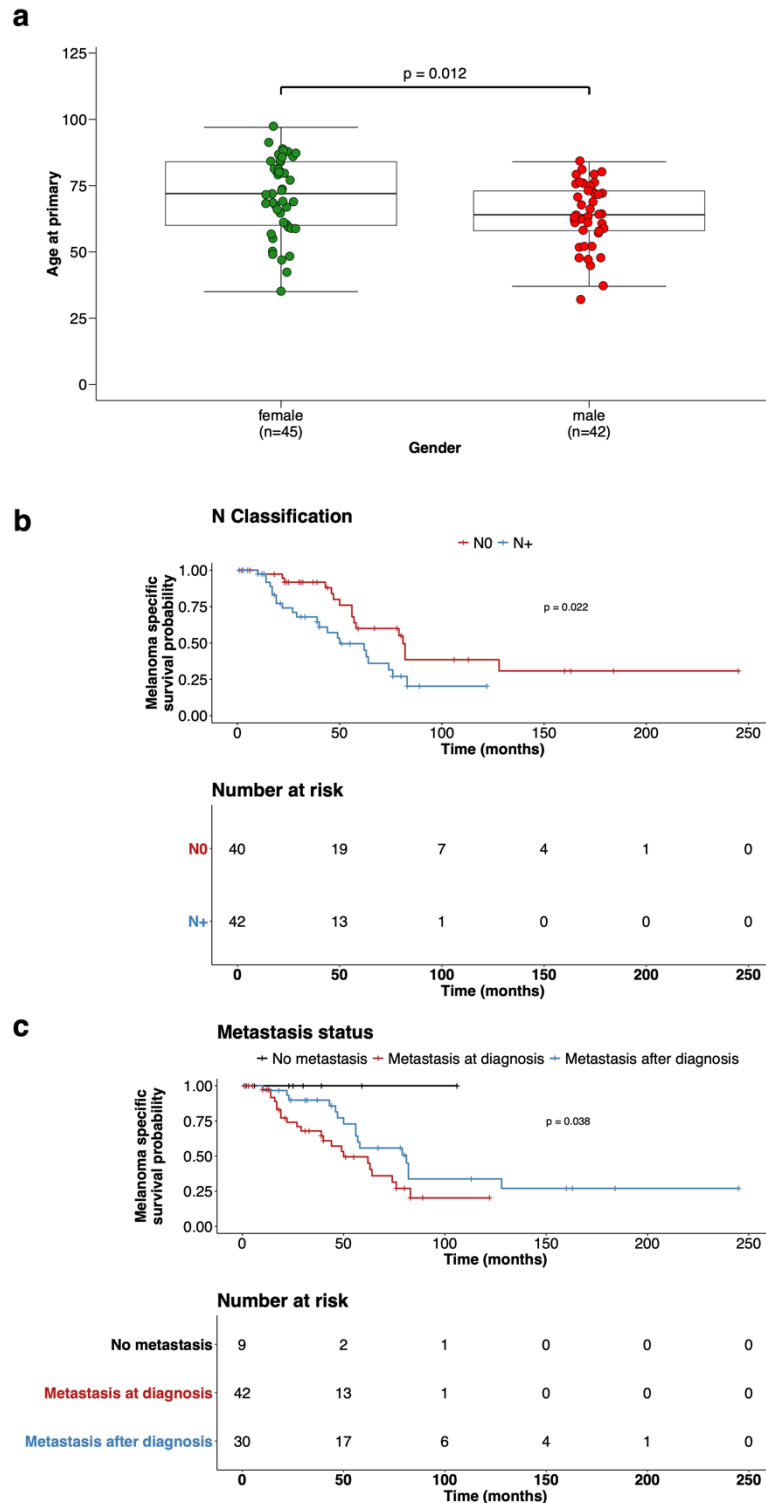
Supplementary Figure 8: GISTIC analysis in tumor subgroups

Supplementary Figure 9: Regions of recurrent rearrangement breakpoints in tumor subgroups

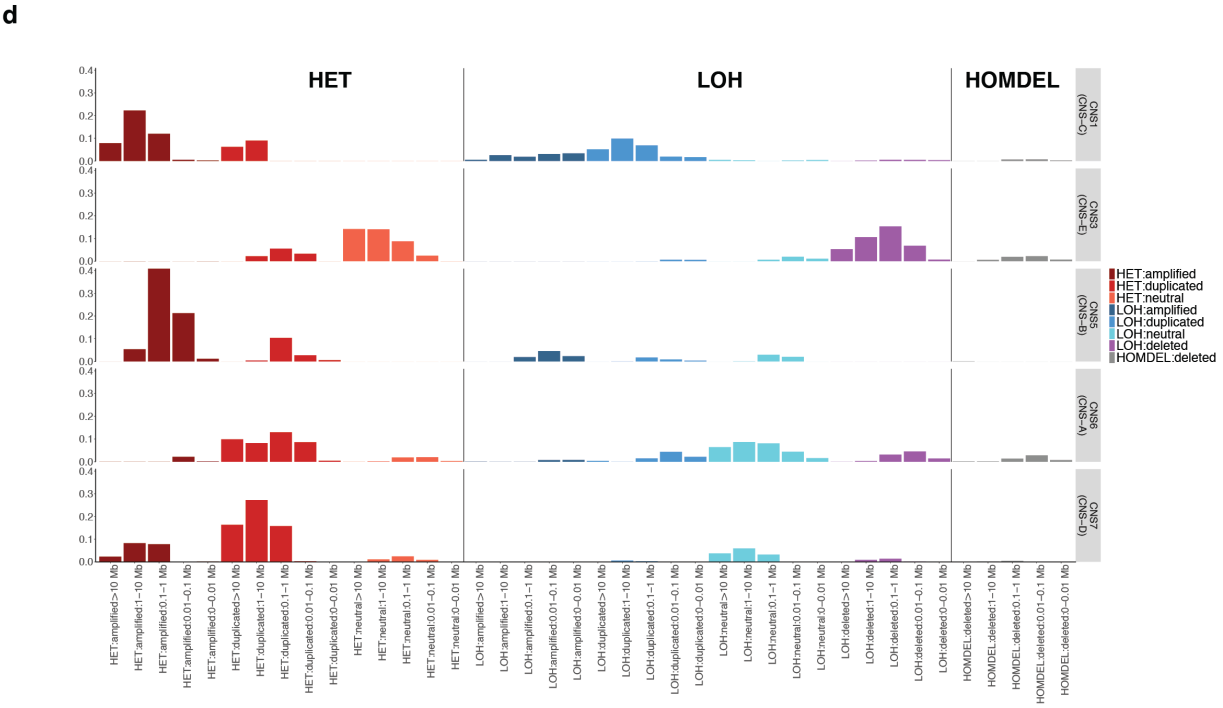
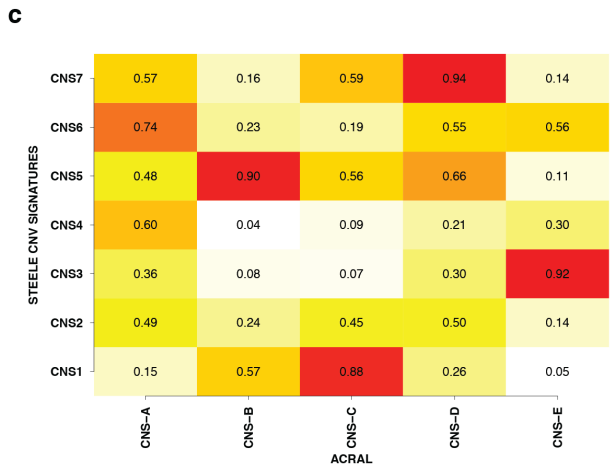
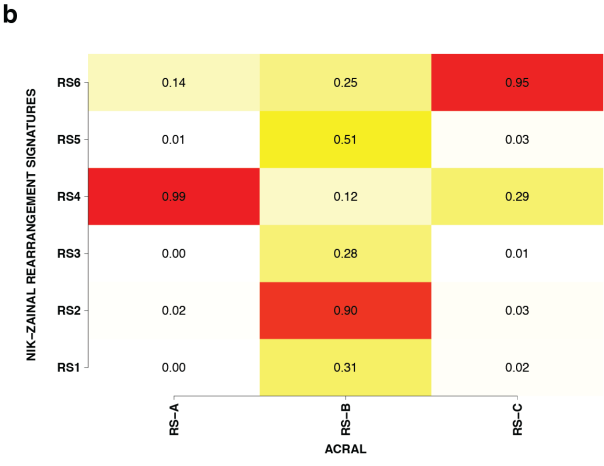
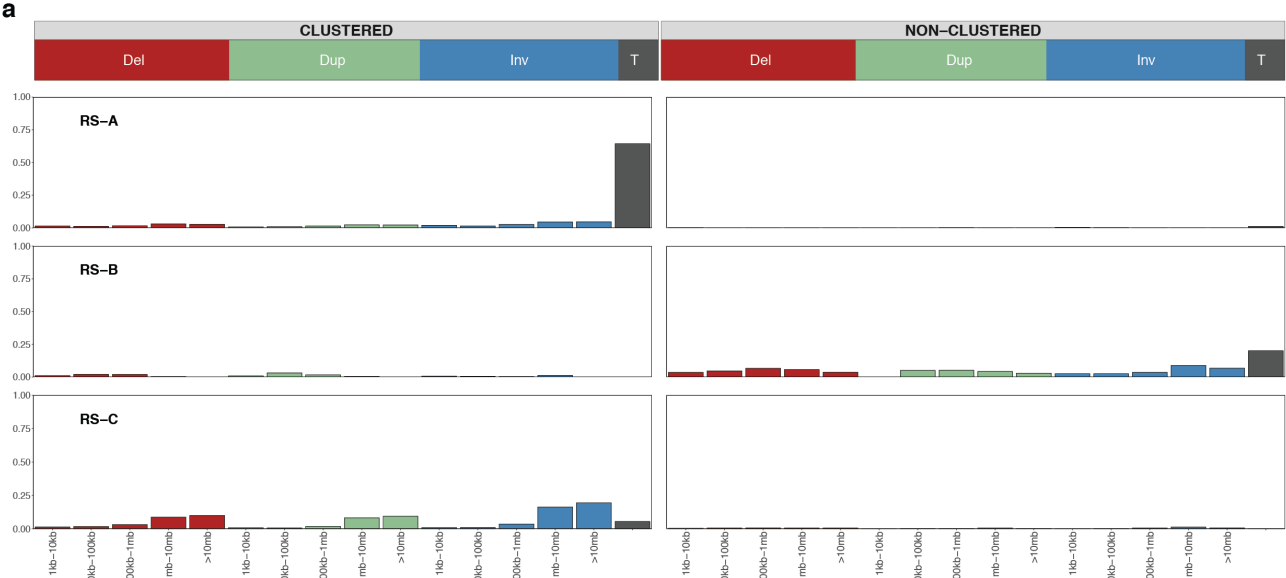
Supplementary Figure 10: Correlation of gene mutations with expression

Supplementary Figure 11: Survival in primary acral melanomas with and without *PTEN* aberrations

Supplementary Figure 12: Neoantigens and the tumor microenvironment

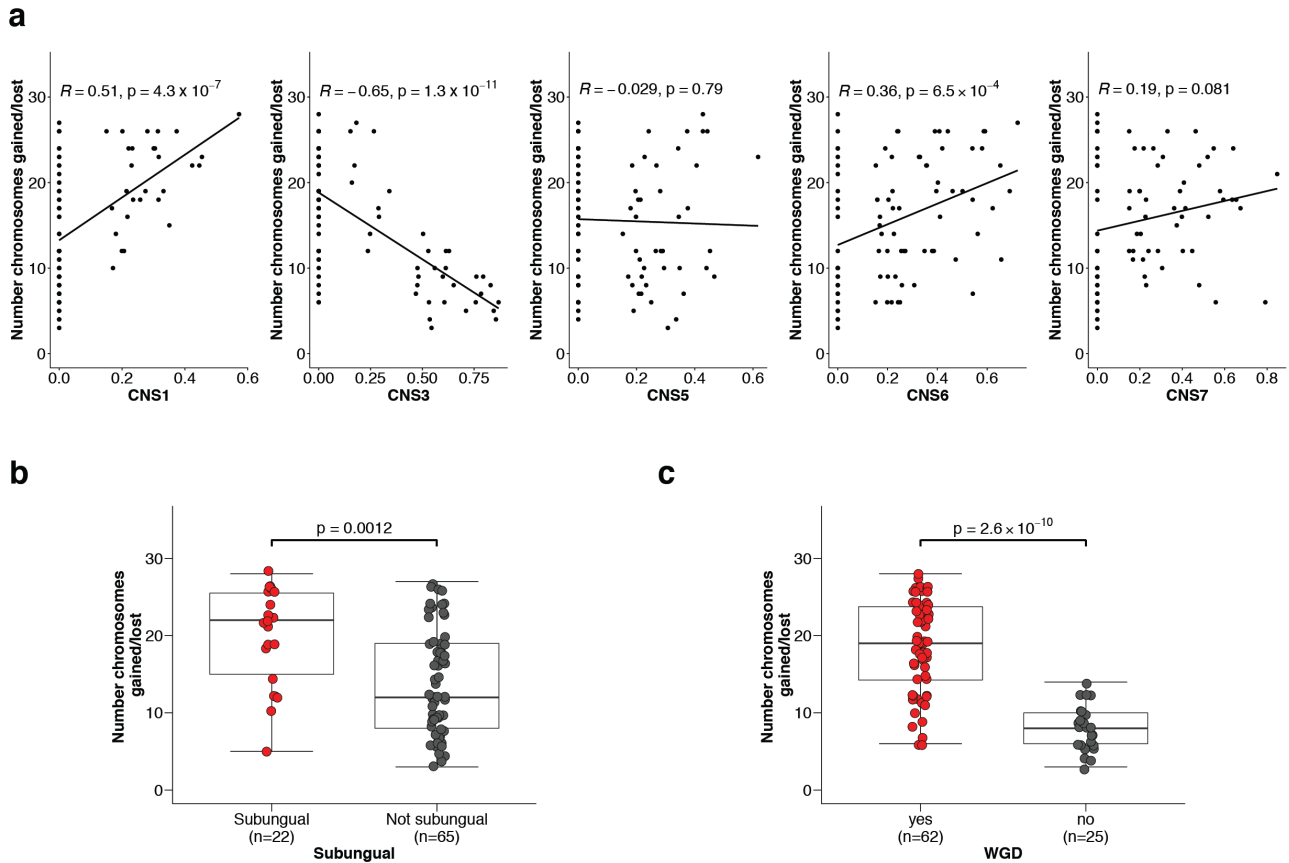


Supplementary Figure 1: Associations of clinicopathological features. **a**, Box plot of age at diagnosis of primary in males and females (Mann-Whitney U test). The box boundaries show the first to third quartiles, the median is the centre line and the whiskers represent 1.5 times the inter-quartile range. **b** and **c**, Kaplan-Meier curves showing melanoma-specific survival using a log-rank test. **b**, N classification. **c**, Metastasis status comparing patients with no metastasis, metastasis present when diagnosed and metastasis after diagnosis.

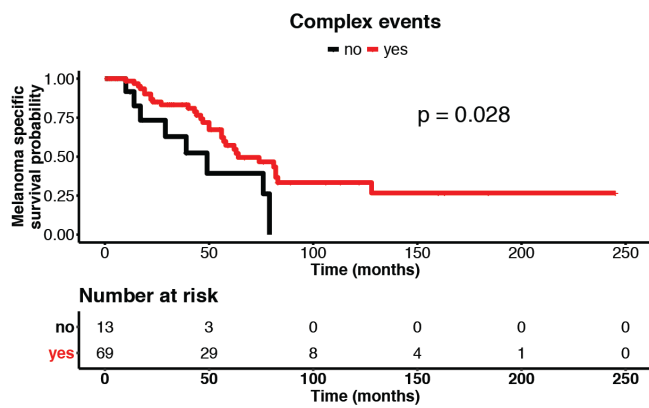
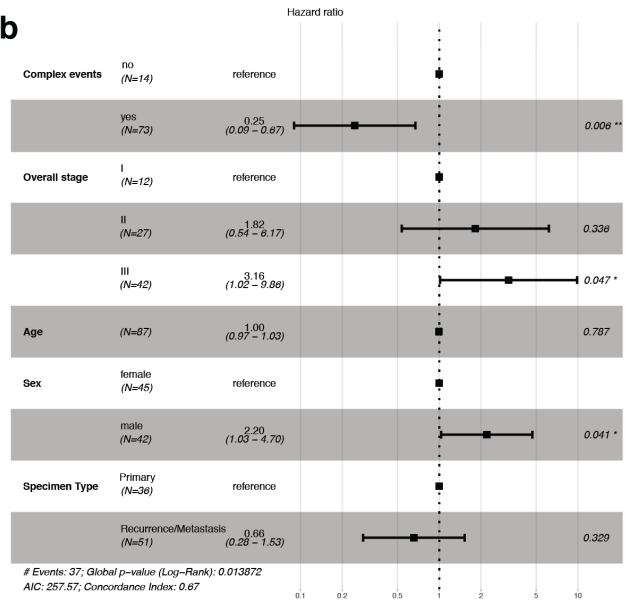
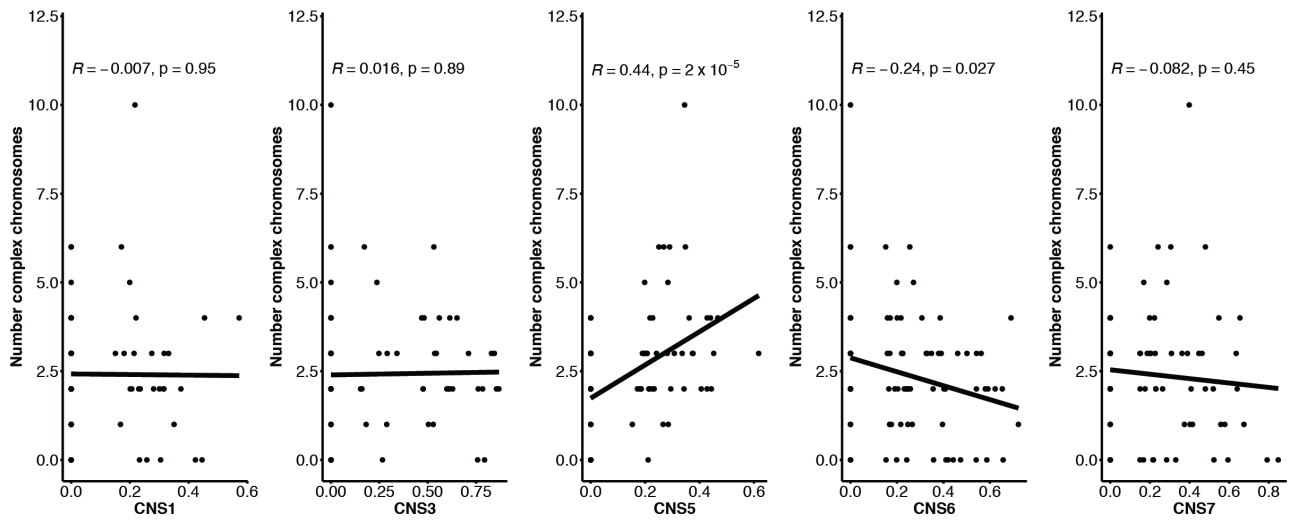
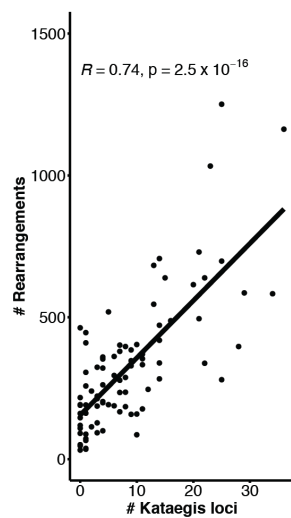
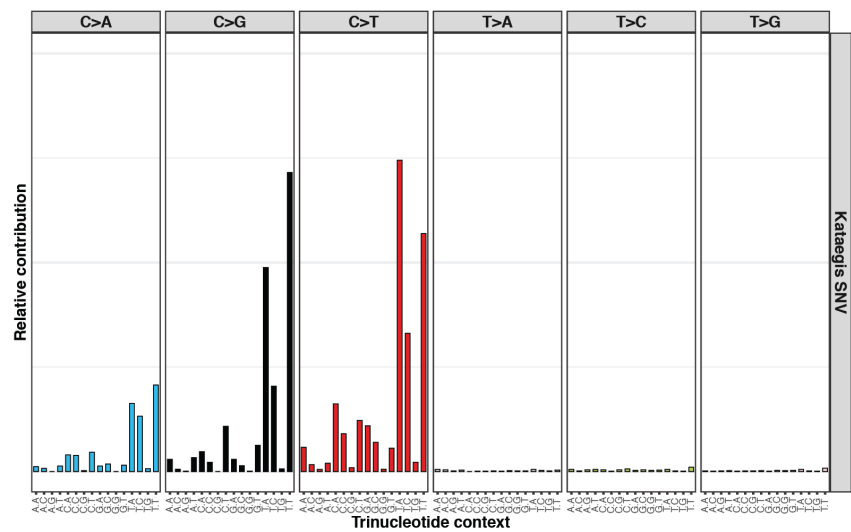


Supplementary Figure 2: Rearrangement and CNV signature probabilities and cosine

similarity. a, Three rearrangement signatures were extracted using non-negative matrix factorization (NMF) and are grouped by clustered and non-clustered deletions (Del), duplications (Dup), Inversions (Inv) of different sizes and translocations (T). The y axis shows the probability of each category type. **b,** Cosine similarity of rearrangement signatures with published breast cancer signatures. **c,** Cosine similarity of CNV signatures with published sarcoma CNV signatures. **d,** Five CNV signatures were extracted using NMF and are grouped into 40 categories. HET, heterozygous; LOH, loss of heterozygosity; HOMDEL, homozygous deletion. The y axis shows the probability of each category type. CNS1 is a signature of amplified LOH and CNS3 occurs in samples with a large proportion of unaltered segments with some small amplifications and large deletions. CNS5 is associated with amplified regions with retention of heterozygosity and CNS6 is a signature of large heterozygous duplicated neutral LOH segments. CNS7 is a signature with retention of heterozygosity with low level gains.

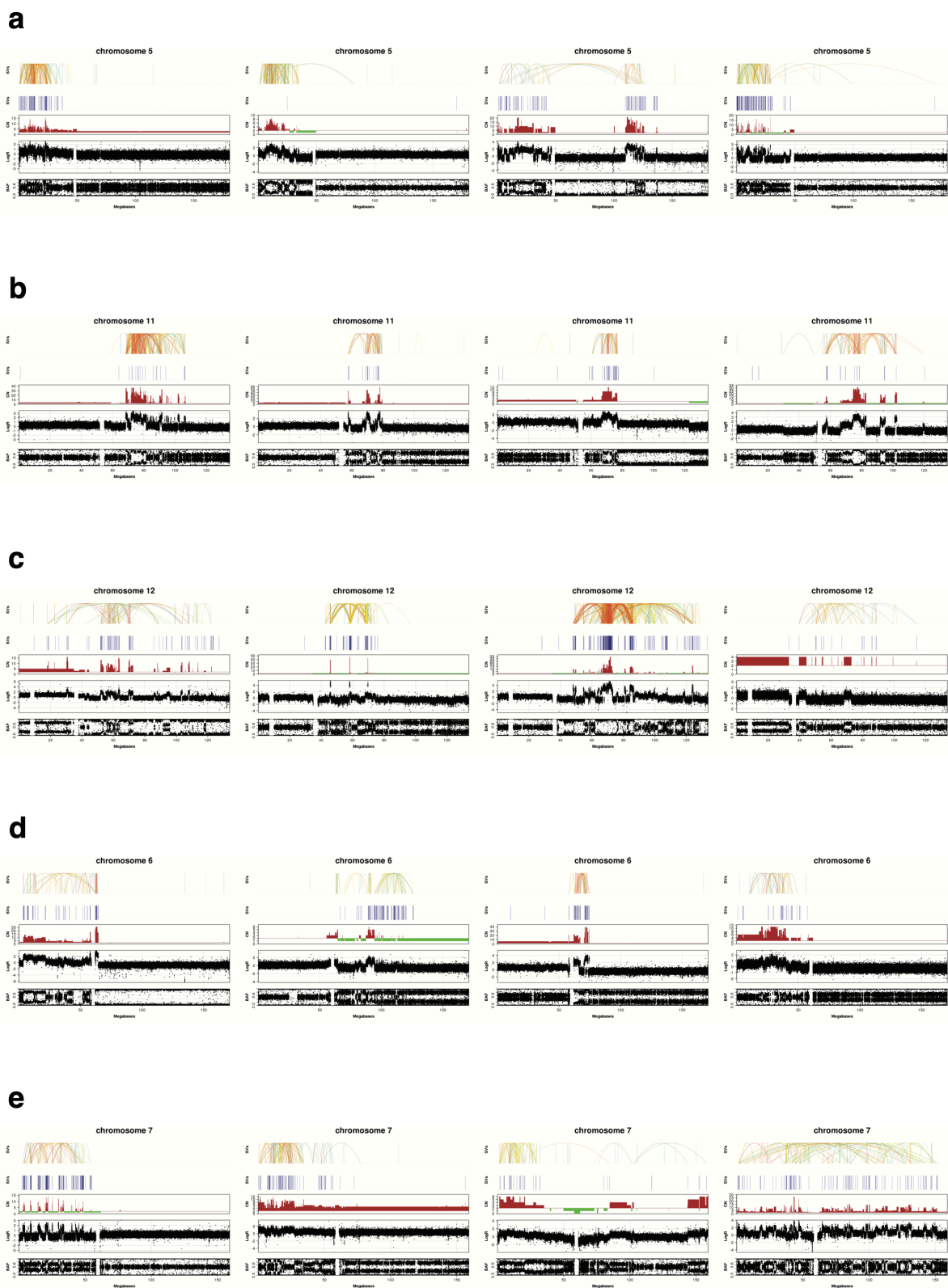


Supplementary Figure 3: Associations with aneuploidy. **a**, Pearson's correlation of CNV signatures CNS1, CNS3, CNS5, CNS6, CNS7 with the number of chromosome arms gained and lost. **b**, Box plot of aneuploidy in tumors that are subungual compared with tumors from other sites (Mann-Whitney U test). **c**, Box plot of aneuploidy in tumors that are diploid or have undergone whole genome duplication (Mann-Whitney U test). In each box plot, the box boundaries show the first to third quartiles, the median is the centre line and the whiskers represent 1.5 times the inter-quartile range.

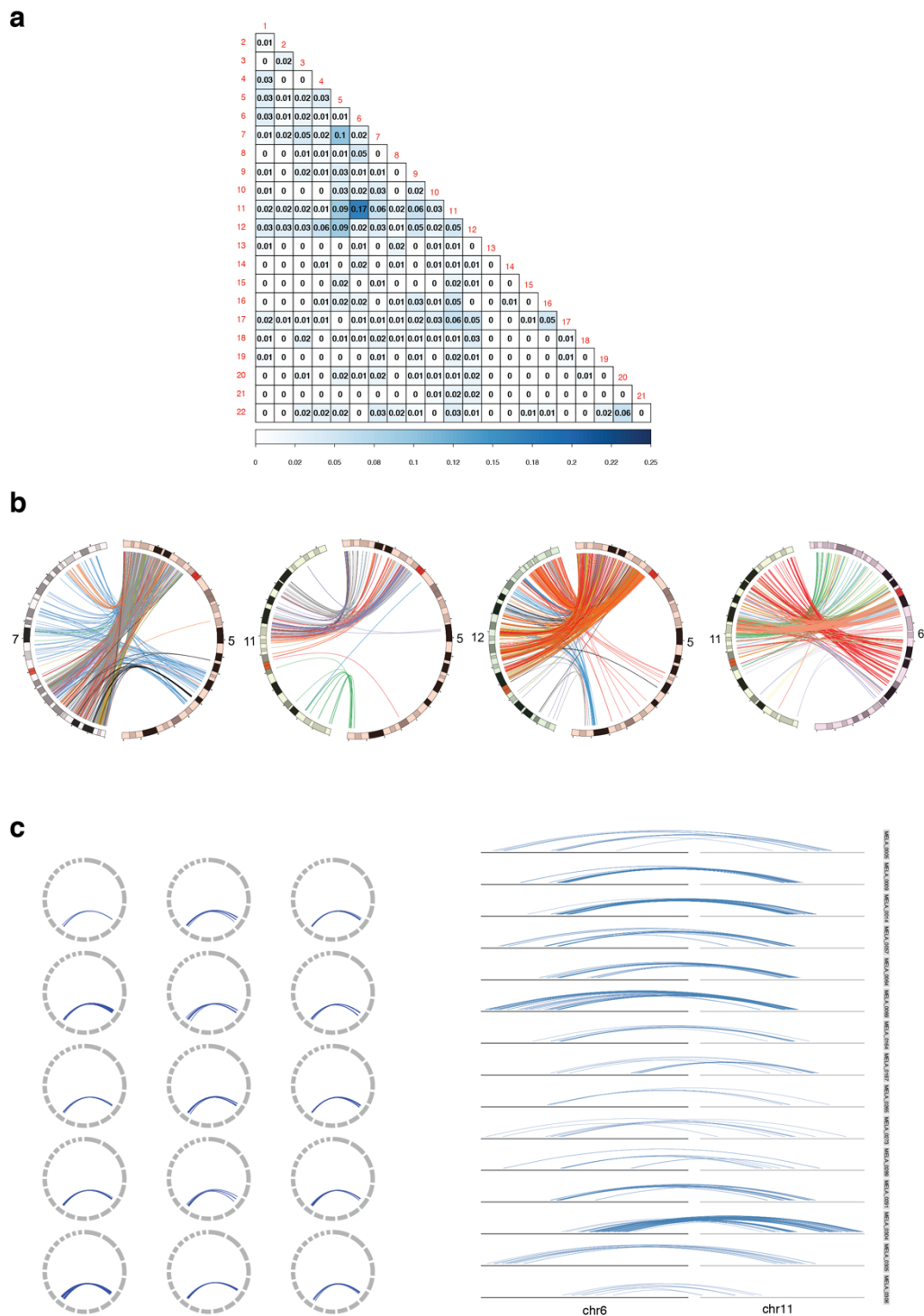
a**b****c****d****e**

Supplementary Figure 4: Associations with localized complex genomic rearrangements

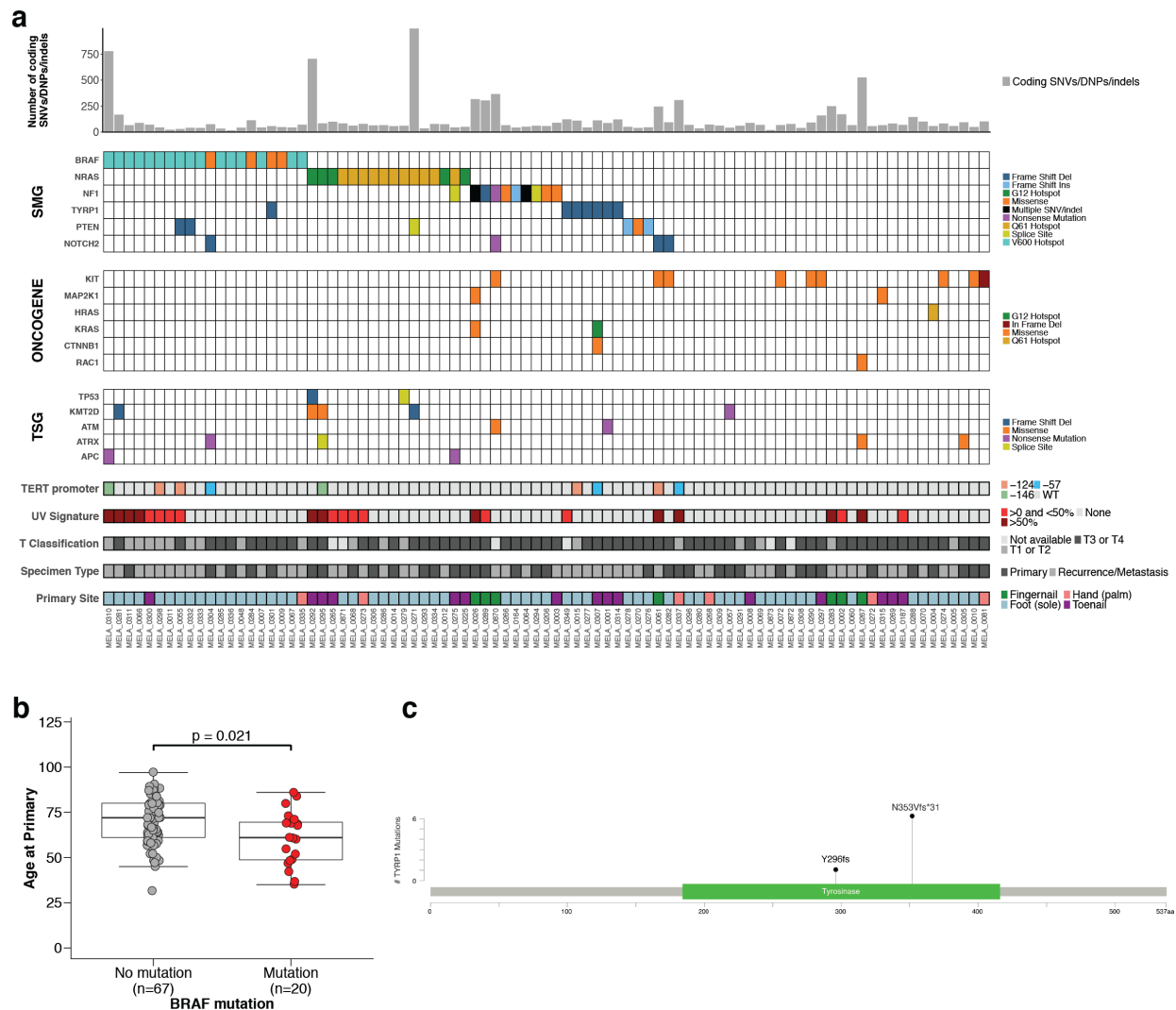
a, Kaplan-Meier plot of melanoma-specific survival with log-rank test in tumors with or without localized complex rearrangements. **b**, Forest plot for a multivariable Cox survival model based on presence of localized complex rearrangements, overall stage, patient age at diagnosis of primary, sex and specimen type (primary or recurrence/metastasis). **c**, Pearson's correlation of CNV signatures CNS1, CNS3, CNS5, CNS6, CNS7 with the number of chromosomes per tumor with localised complex rearrangements **d**, Pearson's correlation of the number of kataegis loci per tumor with structural rearrangement count. **e**, Trinucleotide context of SNVs that fall within identified kataegis loci tumors showing that there is a similarity to APOBEC signatures SBS2 and SBS13.



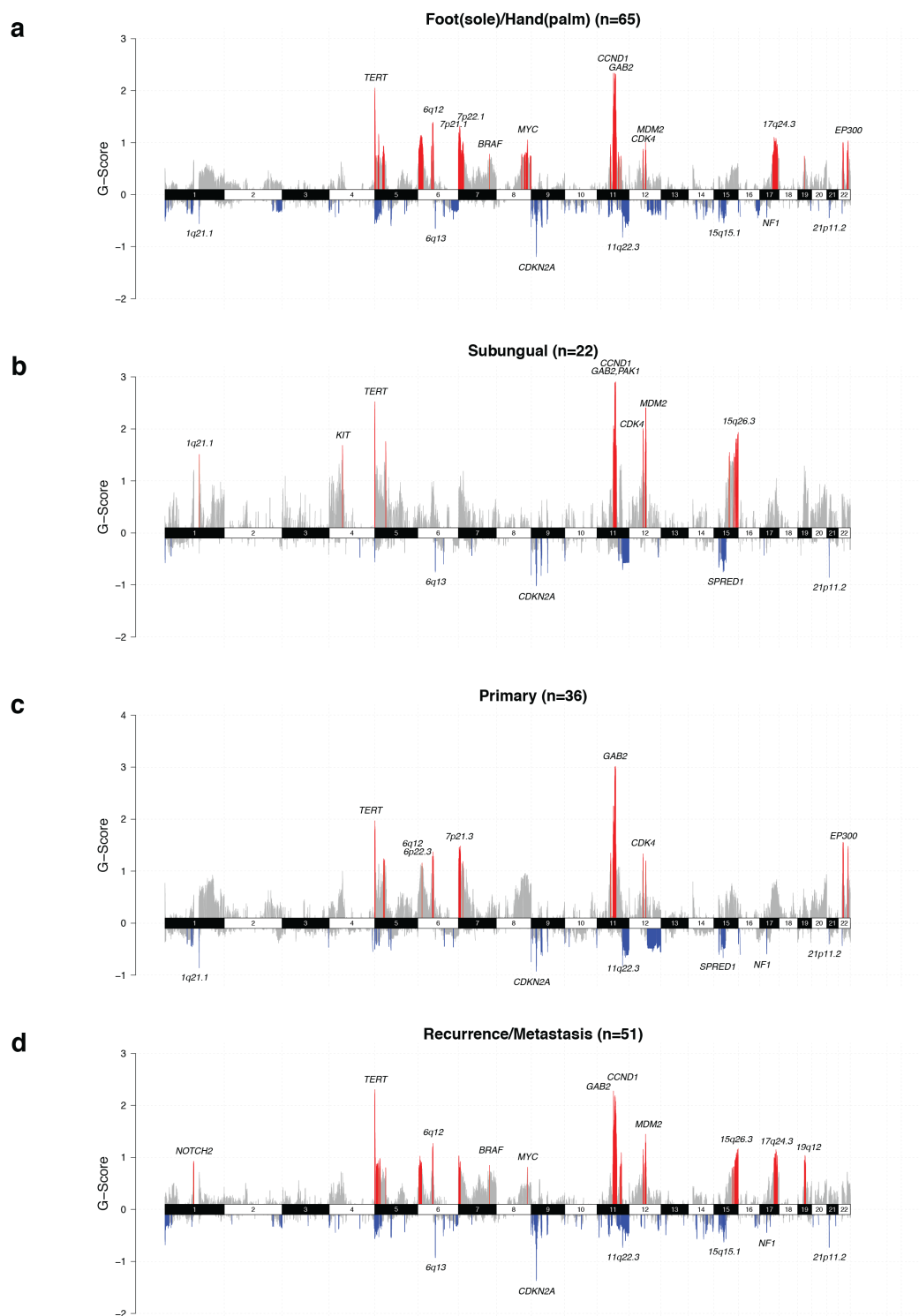
Supplementary Figure 5: Examples of localized complex events on recurrent chromosomes.
a, chromosome 5. **b**, chromosome 11. **c**, chromosome 12. **d**, chromosome 6. **e**, chromosome 7.



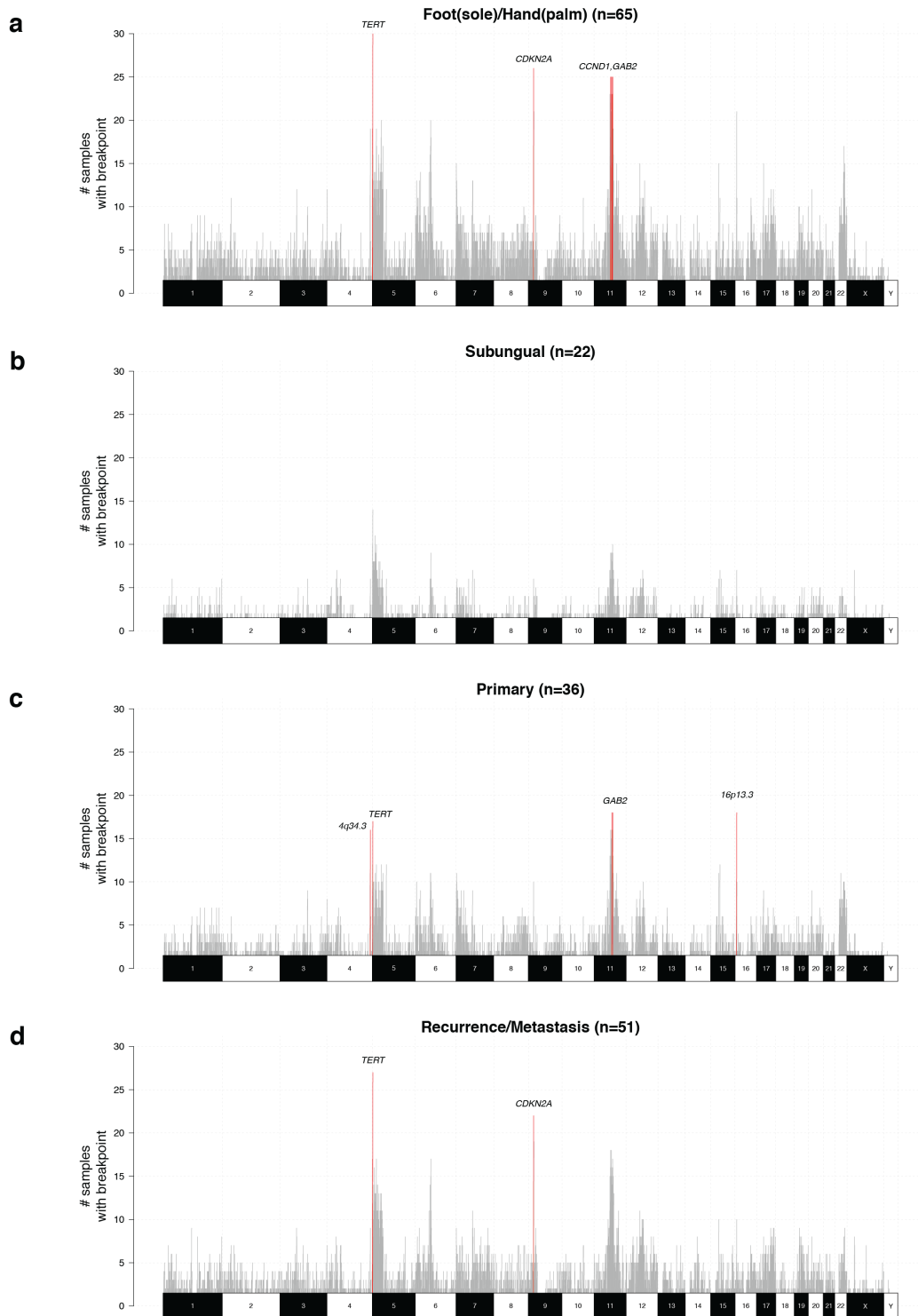
Supplementary Figure 6: Recurrent linking translocations. **a**, Proportion of tumors with 5 or more linking translocations between chromosomes. **b**, Circos plots showing translocations between (left to right) chromosomes 5 and 7; chromosomes 5 and 11; chromosomes 5 and 12 and chromosomes 6 and 11 for tumors with 5 or more linking translocations. Each colour represents the translocations for one tumor. **c**, Circos plots for chromosome 6 to chromosome 11 translocations for each tumor (left) and the location on the chromosome of each linking translocation.



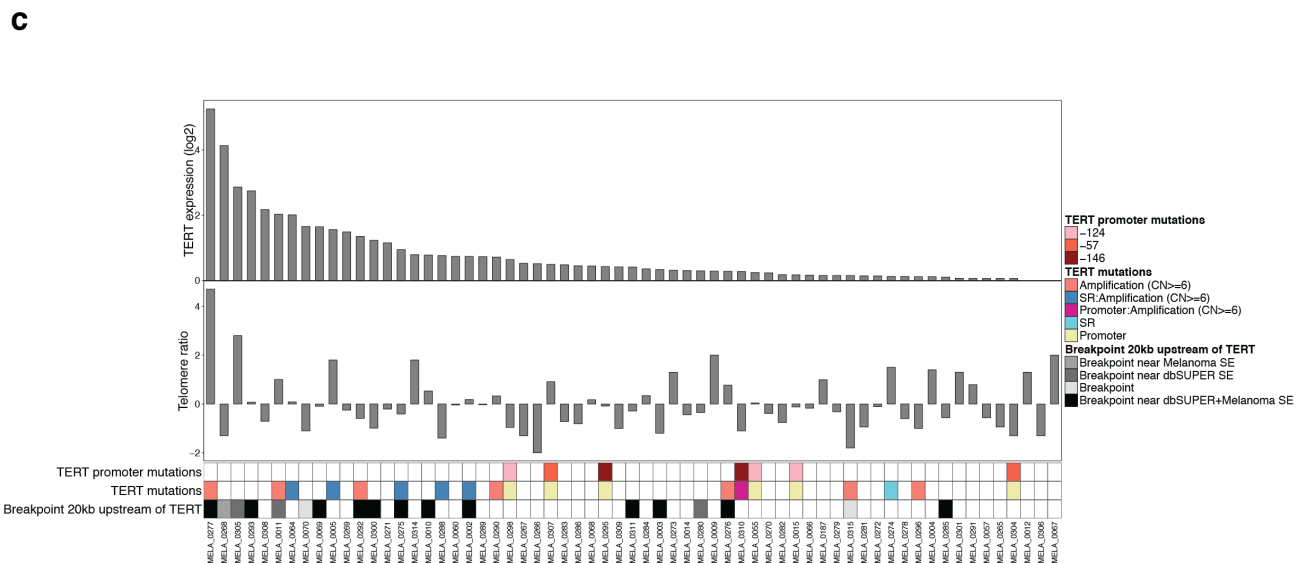
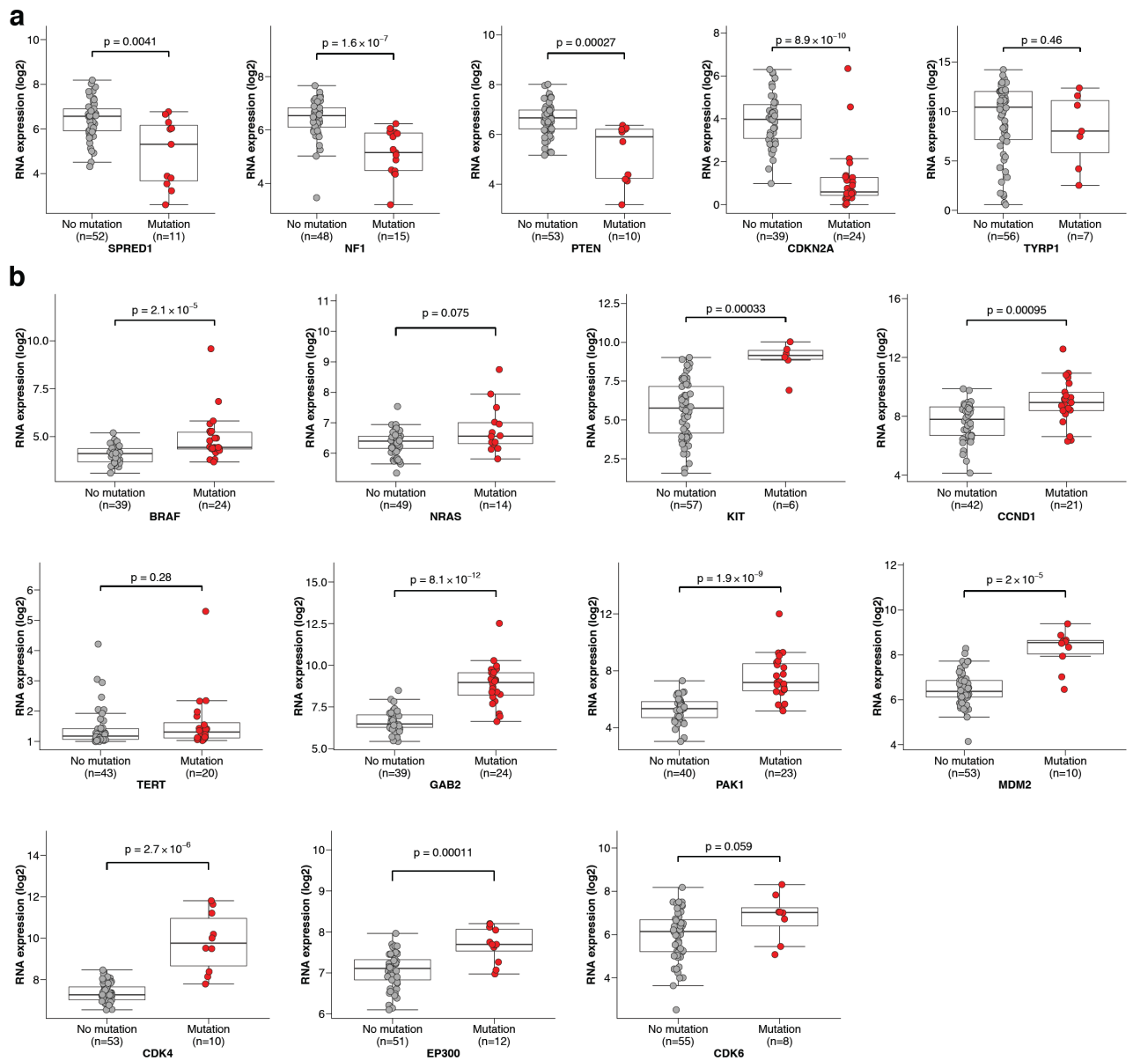
Supplementary Figure 7: Significantly mutated genes. **a**, Oncoplot showing from top to bottom: number of coding SNV/indels; mutations in significantly mutated genes; mutations in oncogenes; mutations in tumor suppressor genes (TSG); *TERT* promoter mutations; presence of UVR signature; T classification; specimen type and primary site per sample. The barchart for the number of coding SNV/indels in the upper panel represents the counts from $n=87$ tumors, where each bar represents the counts for a single tumor. For oncogenes, nine tumors had a *KIT* mutation, two tumors had a hotspot mutation in *MAP2K1* (2 at p.C21S) and two with *KRAS* mutations (1 at the Q61 hotspot and 1 at p.D33E) and single mutations occurred in *HRAS* (G12 hotspot), *RAC1* (P29S) and *CTNNB1* (P37C) were also identified. Mutations, including putatively protein truncating loss of function (LoF) mutations, also occurred in tumor suppressors such *TP53*, *KMT2D*, *ATRX*, *ATM* and *APC*. **b**, Box plot of age at diagnosis of the primary lesion in the presence or absence of *BRAF* mutations (Mann-Whitney U test). In each box plot, the box boundaries show the first to third quartiles, the median is the centre line and the whiskers represent 1.5 times the inter-quartile range. **c**, Position of mutations in TYRP1 protein. Mutations are coloured black (frame shift) and the tyrosine protein domain is coloured green.



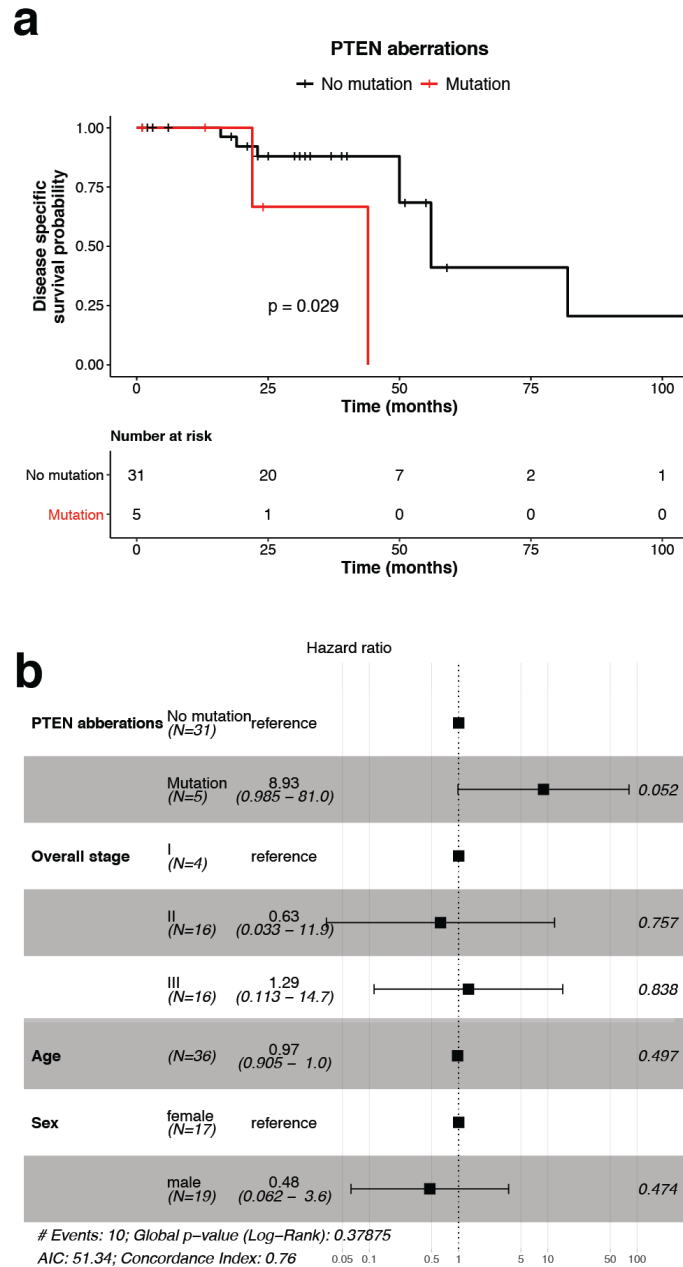
Supplementary Figure 8: GISTIC analysis in tumor subgroups. Focal regions of recurrent amplification (red) and deletion (blue) as identified by GISTIC2 in tumor subgroups. Genes and chromosomal cytobands of interest are annotated in the plot. **a**, Sole of the foot or palm of the hand primary tumor site, **b**, Subungual primary tumor site, **c**, Primary tumors, **d**, Recurrence and metastasis tumors.



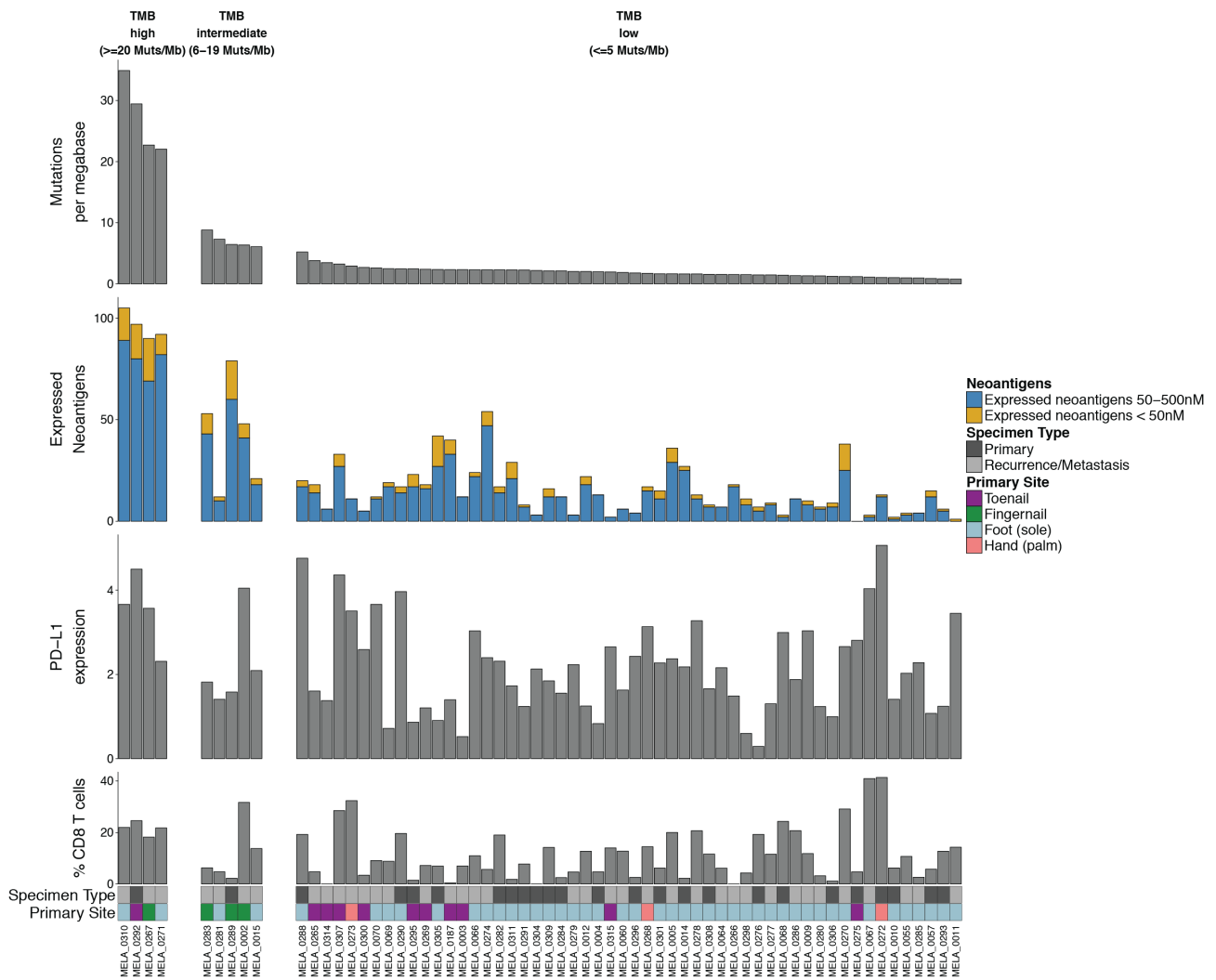
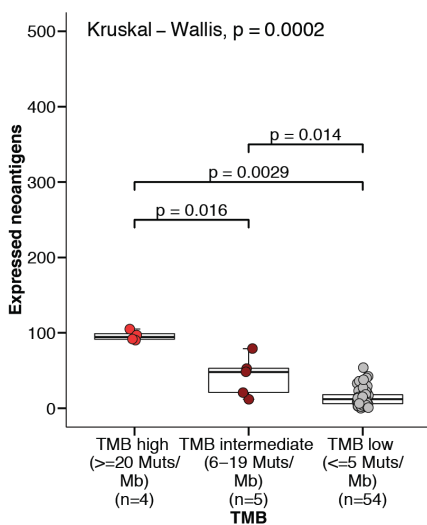
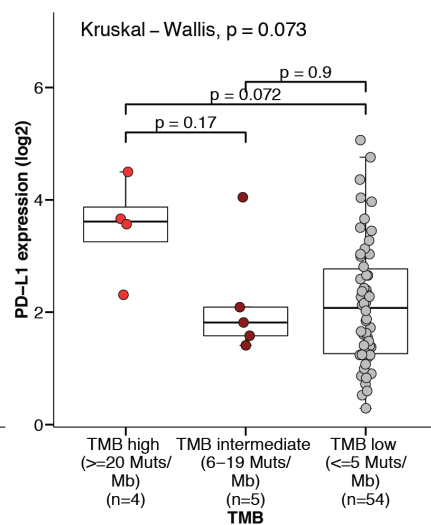
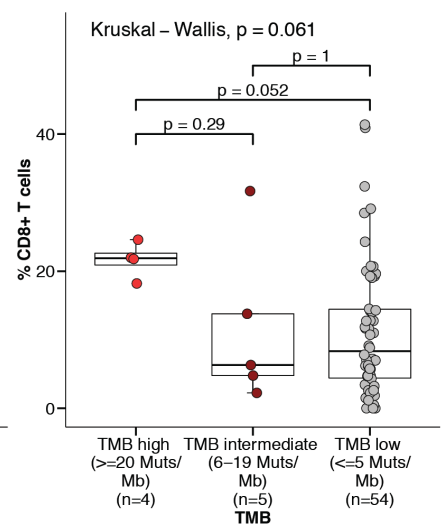
Supplementary Figure 9: Regions of recurrent rearrangement breakpoints in tumor subgroups. Regions of recurrent rearrangement breakpoints as identified by RETREAD in tumor subgroups. The plot shows 1Mb windows that have one or more rearrangements. Bars in gray represent windows that are not significant ($q > 0.2$), and red bars indicate regions that are significantly enriched ($q < 0.2$). **a**, Sole of the foot or palm of the hand primary tumor site. **b**, Subungual primary tumor site. **c**, Primary tumors. **d**, Recurrence and metastasis tumors.



Supplementary Figure 10: Correlation of gene mutations with expression. **a**, Box plots of gene expression in tumors with and without mutations that could potentially lead to decreased expression (SV, Homozygous deletion - copy number 0, SNV/indel) (Mann-Whitney U test). **b**, Expression of genes with and without mostly putatively activating mutations (amplification, SNV/indel mutations) (Mann-Whitney U test). In each box plot, the box boundaries show the first to third quartiles, the median is the centre line and the whiskers represent 1.5 times the inter-quartile range. **c**, *TERT* expression, telomere ratio ($\log_2$ tumor telomere length/matched normal telomere length), *TERT* mutations and rearrangement breakpoints upstream of *TERT*. Upstream breakpoints are coloured based on whether or not they are close (within 100kb) to a super enhancer. All gene expression values are shown in $\log_2(\text{TMM normalized RNA-seq values}+1)$.



Supplementary Figure 11: Survival in primary acral melanomas with and without *PTEN* aberrations. **a**, Kaplan-Meier plot of melanoma-specific survival with log-rank test in primary tumors with or without *PTEN* aberrations. **b**, Forest plot for a multivariable Cox survival model based on *PTEN* aberrations, overall stage, patient age at diagnosis of primary and sex).

a**b****c****d**

Supplementary Figure 12: Neoantigens and the tumor microenvironment. **a**, From top to bottom: tumor mutational burden (TMB) in mutations per megabase; number of expressed neoantigens with predicted HLA binding affinity of < 500 nM; PD-L1 (*CD274*) expression shown as $\log_2(\text{TMM normalized RNA-seq values}+1)$; percent of CD8 positive T cells estimated by immune cell deconvolution using CIBERSORT; specimen type group and primary site. Each bar chart represents data from n=63 tumors, where each bar represents the values for a single tumor. Tumors are divided into TMB high (≥ 20 mutations per megabase, n=4), TMB intermediate (6-19 mutations/Mb, n=5) and TMB low (1-5 mutations/Mb, n=54). **b-d**, Box plots comparing tumors with differing TMB. In each box plot, the box boundaries show the first to third quartiles, the median is the centre line and the whiskers represent 1.5 times the inter-quartile range. The overall test for significance is a Kruskal-Wallis test and p-values for pairwise comparisons represent Mann-Whitney U tests with adjustment for multiple test correction by FDR. **b**, Expressed neoantigens **c**, PD-L1 expression ($\log_2$ TMM normalized RNA-seq counts+1) **d**, Percent of CD8 positive T cells estimated by immune cell deconvolution using CIBERSORT.