

Fig. 1

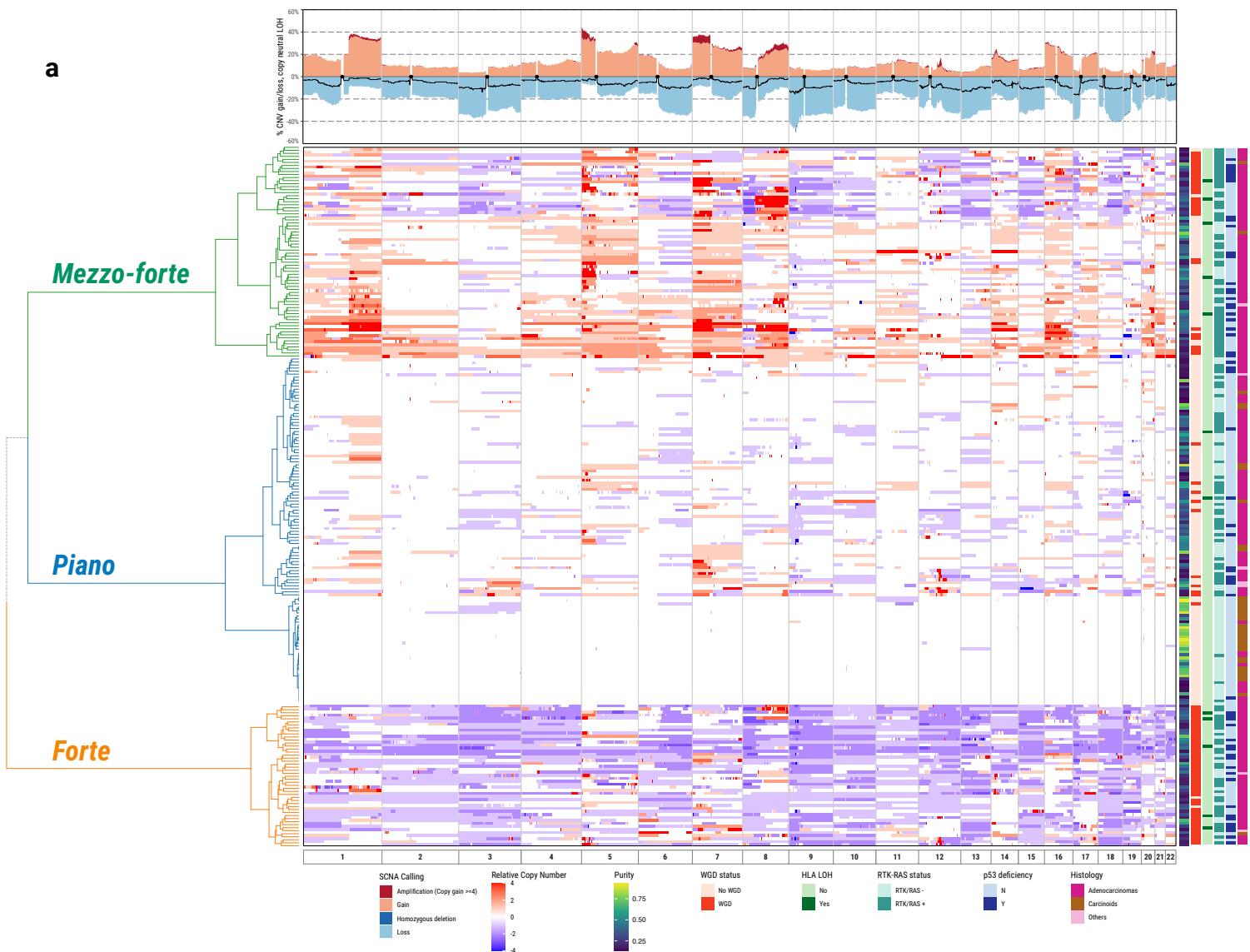


Fig. 1 Landscape of somatic copy number alterations in Sherlock-Lung.

Left panel shows unsupervised clustering of arm-level SCNA events: *piano*, *mezzo-forte* and *forte*. The relative copy number is calculated as: total copy number - ploidy (non-WGD=2 and WGD=4). Samples in rows are annotated by tumor purity, WGD status, HLA LOH, RTK-RAS status, TP53 deficiency, and tumor histological type. Top panel shows SCNA frequency including amplification, deletion and copy neutral LOH (black line).

Fig. 2

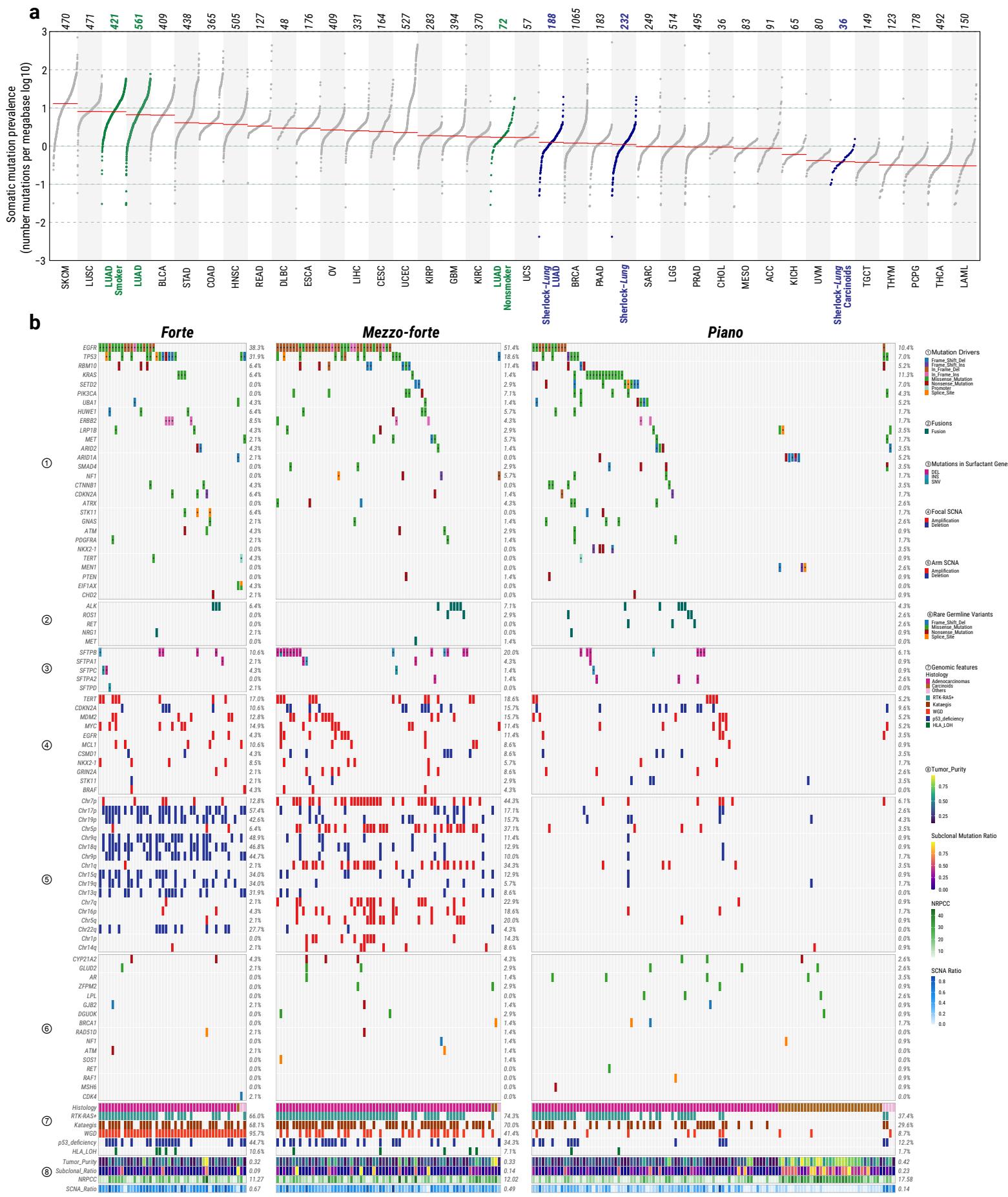


Fig. 2

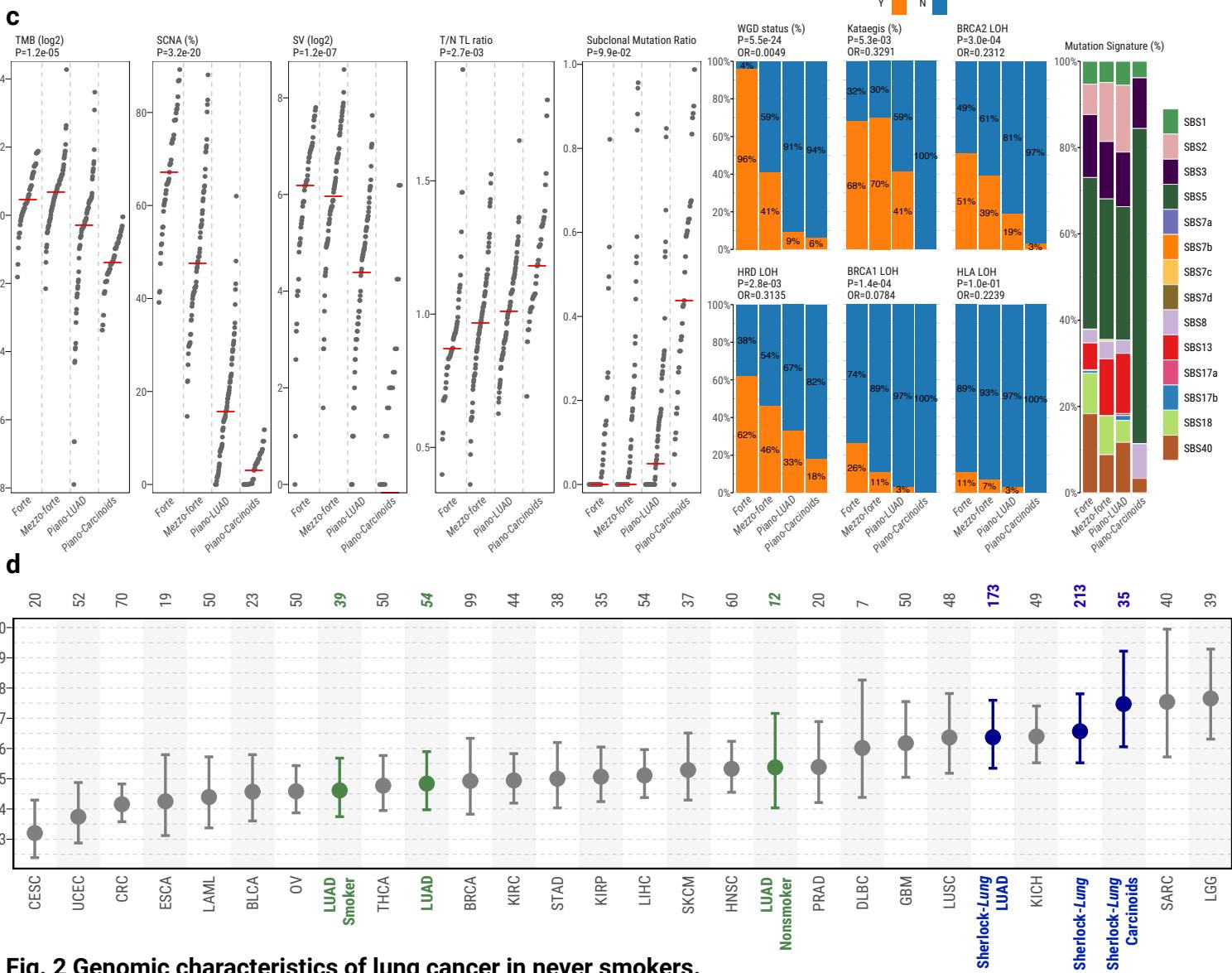


Fig. 2 Genomic characteristics of lung cancer in never smokers.

a, Tumor mutational burden (TMB) across lung cancer in never smokers from the Sherlock-Lung study (blue) and 33 cancer types from the TCGA study. The Sherlock-Lung samples are shown overall and by histological type. TCGA LUAD samples (green) are shown overall and by smoking status. Each dot represents a sample; total sample numbers for each type are shown at the top. The red horizontal lines are the median numbers of mutations per megabase (log10). On the bottom, acronyms of cancer types as in TCGA. **b**, Summary of genomic features in LCINS based on different SCNA clusters. Panels from top to bottom describe: 1) most frequently mutated or potential driver genes; 2) oncogenic fusions; 3) somatic mutations in surfactant associated genes; 4) significant focal SCNAs; 5) significant arm-level SCNAs; 6) genes with rare germline mutations; 7) and 8) different genomic features. The numbers on the right panel show the overall frequency (1-7) or median values (8). NRPCC: the number of reads per clonal copy. **c**, Comparison of genomic aberrations or features among *forte*, *mezzo-forte*, *piano-LUAD*, and *piano-Carcinoids* tumors. Left five panels: tumor mutation burden, percentage of genome with SCNAs, SV burden, T/N TL ratio and subclonal mutation ratio. *P*-values are calculated using two-sided Mann-Whitney U test. Middle six panels: enrichments for WGD, Kataegis, *BRCA2* LOH, *BRCA1* LOH, HRD LOH and *HLA* LOH. *P*-values and *OR* are calculated using two sided Fisher's exact test. Right one panels: signature decomposition based on Cosmic v3 signatures. All statistical analyses were performed between *forte* and *piano-LUAD*. **d**, Distribution of mean telomere lengths (TL) in Sherlock-Lung (dark blue, overall and by histological type), TCGA LUAD (green, overall and by smoking status) and TCGA other cancer types (Grey).

Fig. x

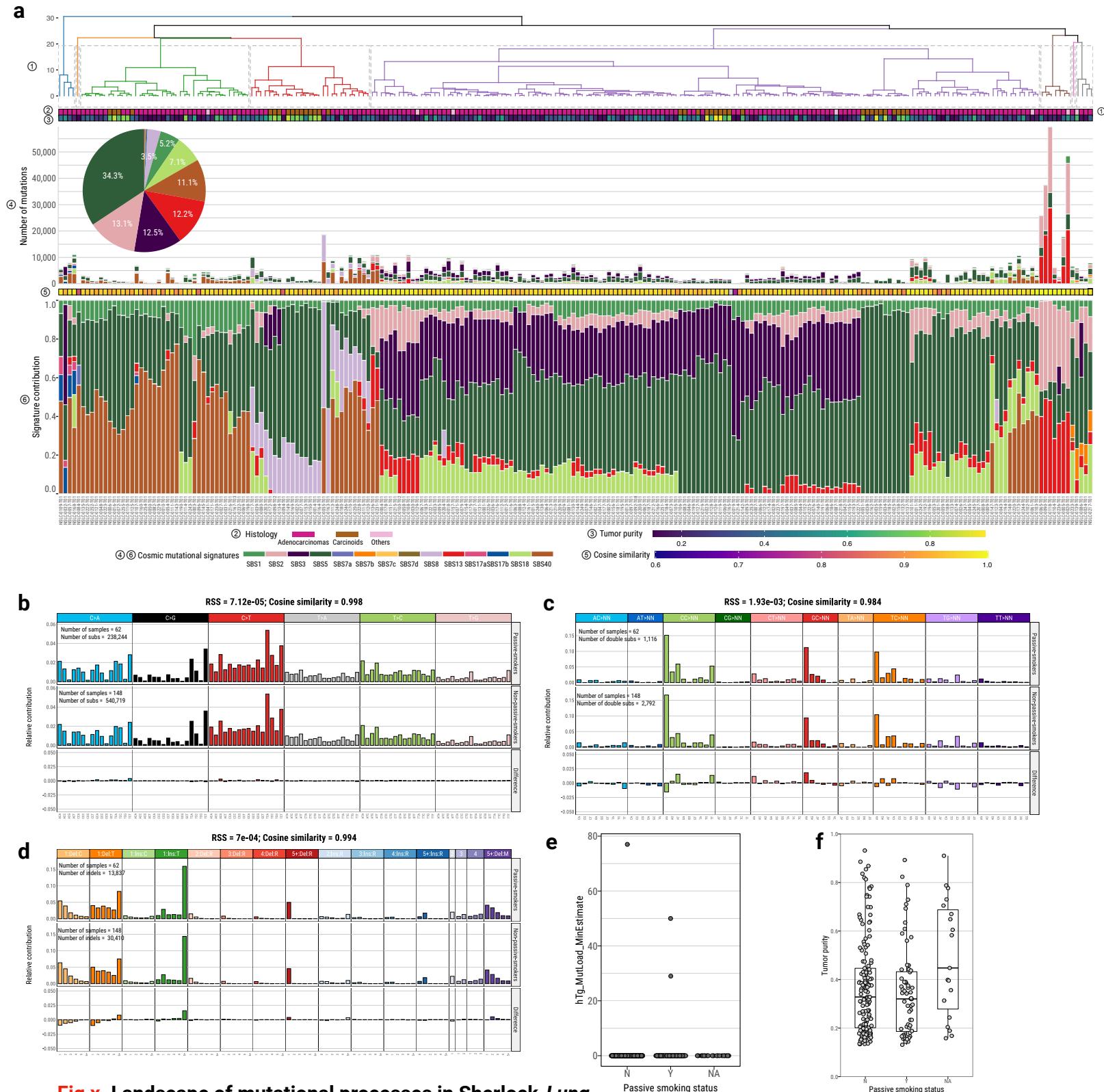
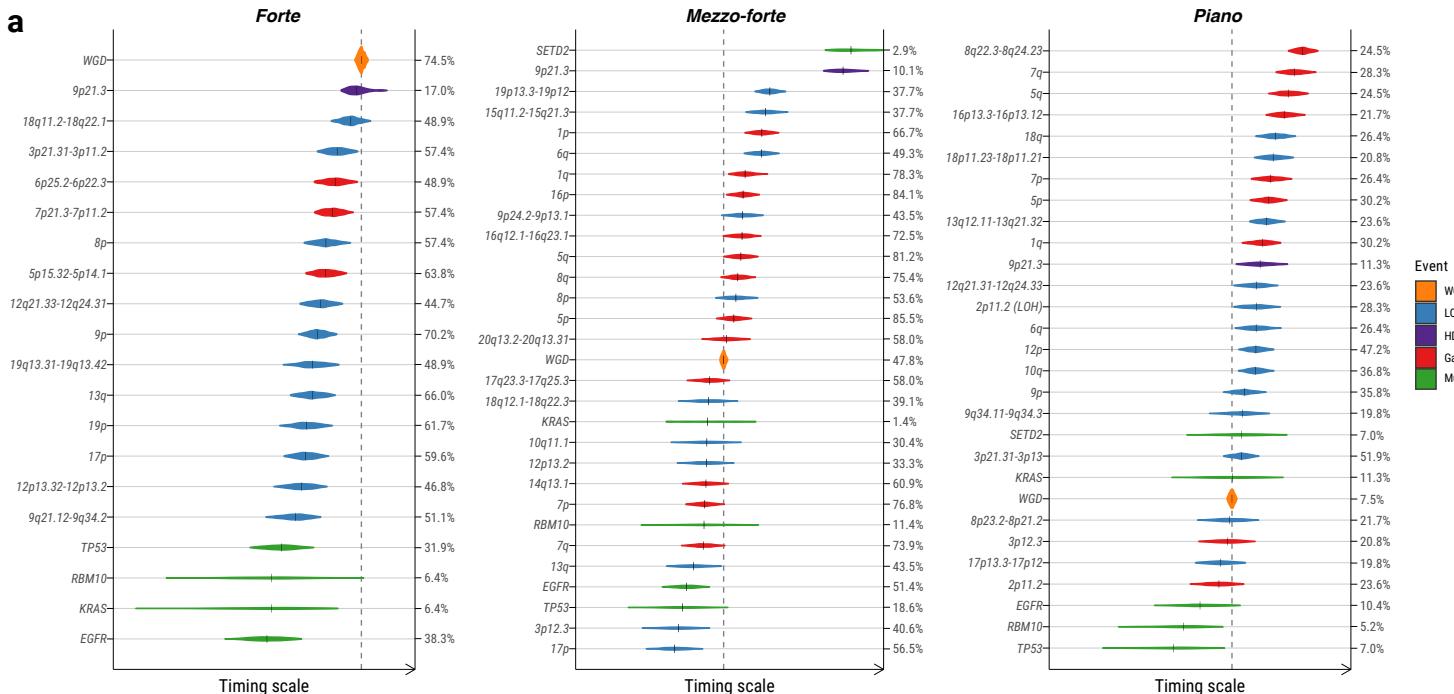


Fig x. Landscape of mutational processes in Sherlock-Lung.

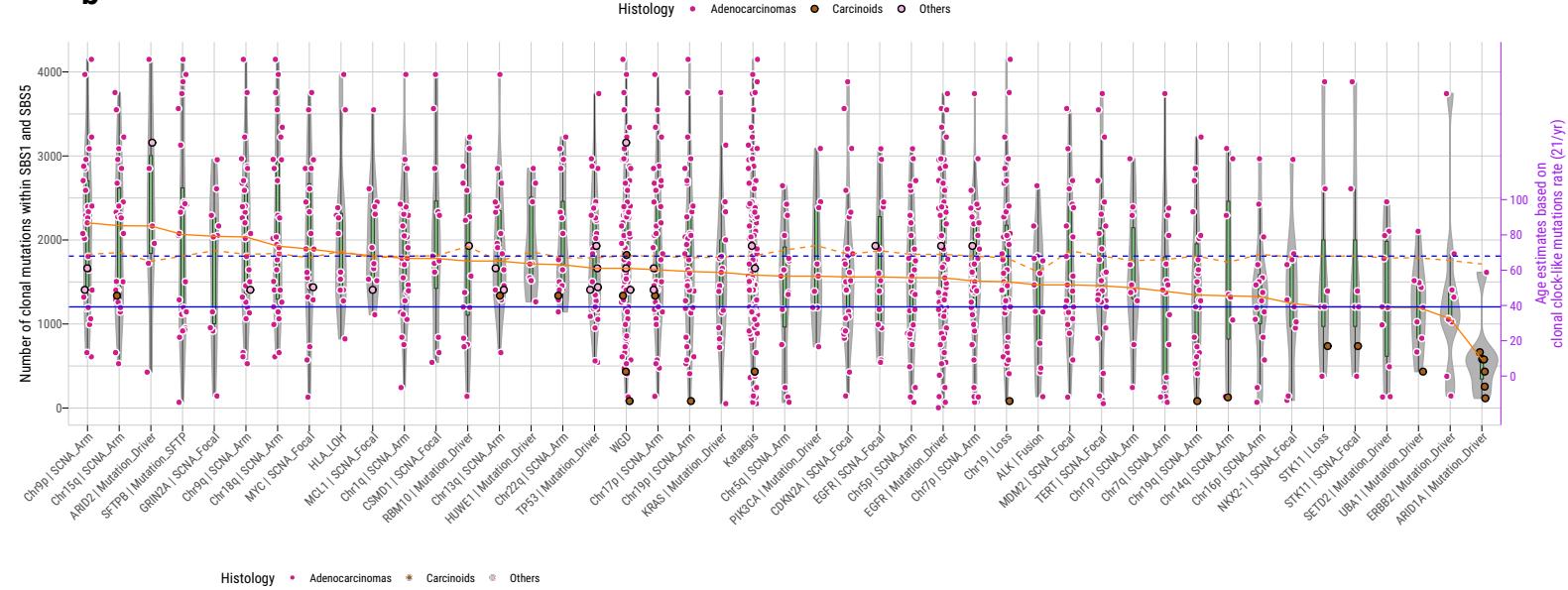
a) Mutational signature profile of single base substitutions (SBS) across 232 Sherlock-Lung samples. Panels from top to bottom: 1) Unsupervised clustering based on the proportion of SBS signatures; 2) Tumor histological type; 3) Tumor purity; 4) Pie chart showing the percentage of mutations contributed to each SBS signature and the barplot presenting the total number of SNVs assigned to each SBS signatures; 5) Cosine similarity between original mutational profile and signature decomposition result; 6) Proportions of SBS mutational signatures in each sample. b-d) Mutational spectra comparison of single base substitutions (b), double base substitutions (c) and indels (d) between passive-smokers and non-passive smokers. e-f) Identification of alkylation-induced mutagenesis ($hTg \rightarrow hGg$ signature) (e) and tumor purity (f) among passive smokers (Y), non-passive smokers (N) and rest of patients (NA).

Fig. 3

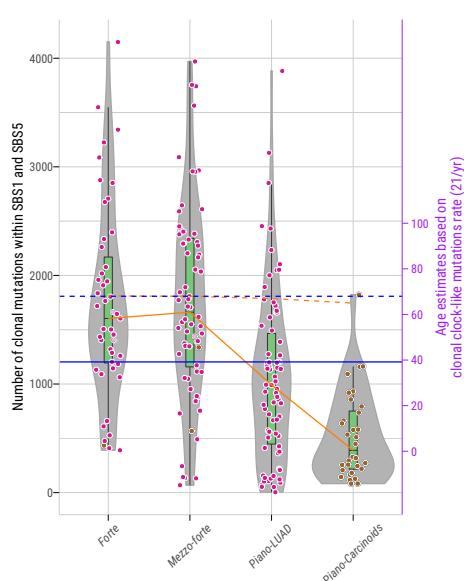
a



b



c



e

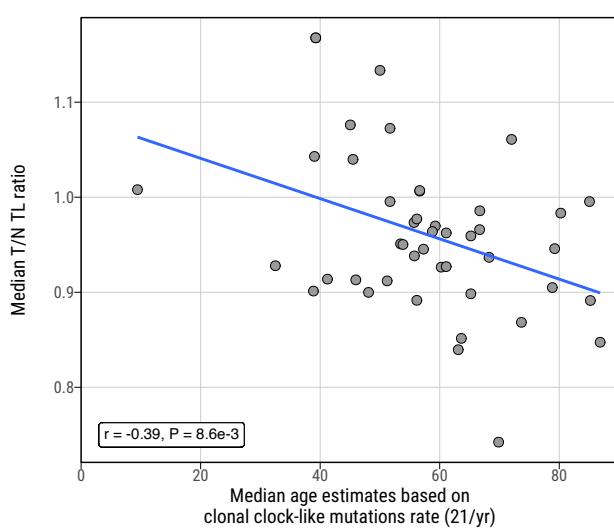
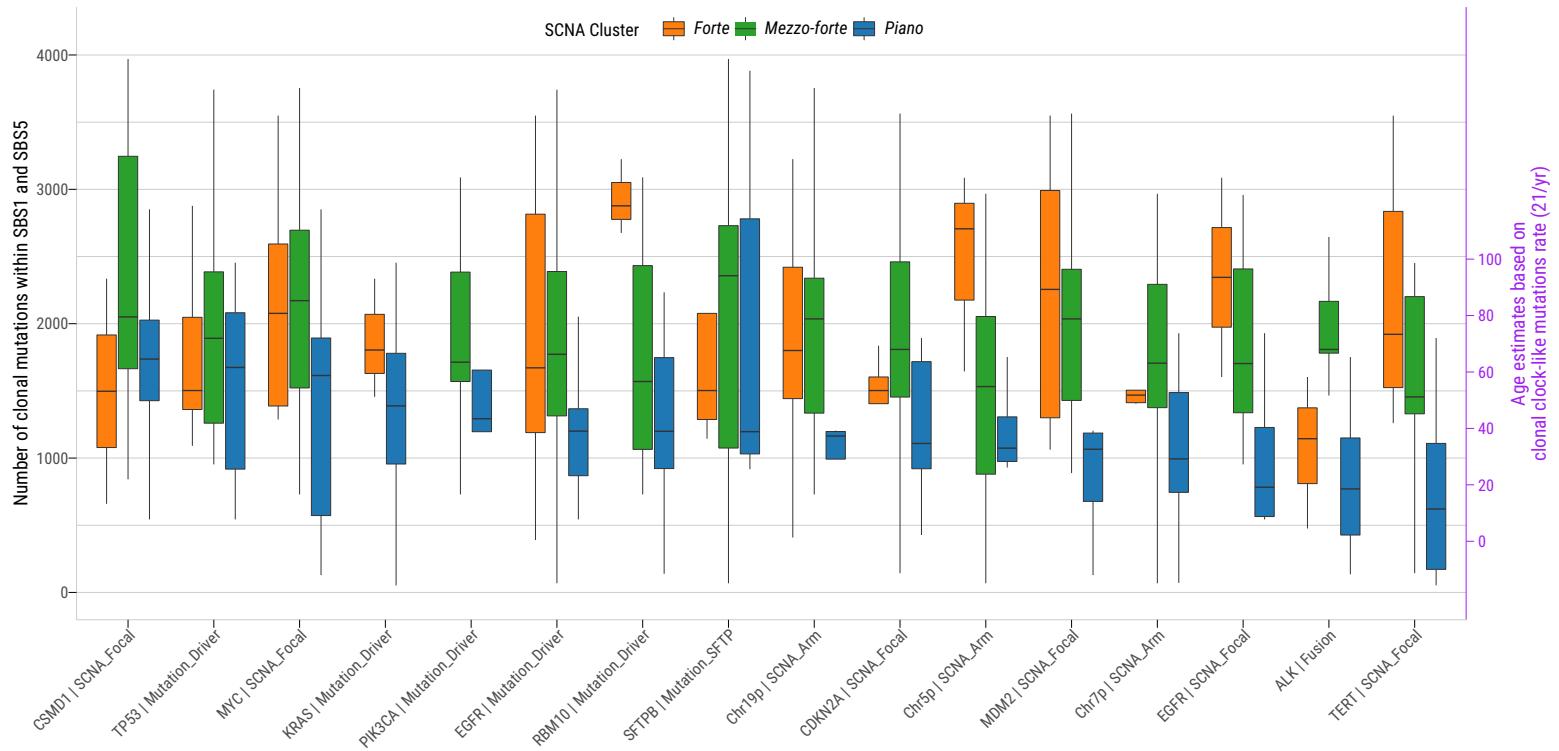


Fig. 3**d****Fig. 3 Reconstruction of the evolutionary history of lung cancer in never smokers**

a, Diagram of estimated ordering of significant SCNAs (including chromosome gains/losses and mutations) relative to WGD in three lung cancer subtypes based on SCNA clusters *forte*, *mezzo-forte* and *piano*. The size of violin plots denotes the uncertainty of timing for specific events across all samples and the short black solid lines show the median time. The vertical dashed line indicates the median time for WGD events. Ordering of genomic events was based on the PlacketLuce package model with 95% CI. The frequency of each event is labeled on the right y-axis. **b**, Estimation of tumor chronological time by genomic alterations or features as shown in **Figure 2b**. Each group (violin plot, boxplot, and scatter plot) represents a genomic feature with frequency > 3%. Each dot denotes the number of clonal clock-like mutations (assigned to SBS1 and SBS5) (left y-axis) in a sample that harbors the given genomic aberration (x-axis). The color of each dot represents the tumor histological subtype. The molecular age estimated based on clonal clock-like mutations (21/yr) is presented on the right y-axis. The orange solid and dashed line indicates the median estimated molecular age and the median age at diagnosis in the same group, respectively. The blue solid and dashed line indicates the median estimated molecular age and the median age at diagnosis in all samples, respectively. **c**, Similar to **b**, estimation of tumor chronological time among SCNA subtypes: *forte*, *mezzo-forte*, *piano-LUAD* and *piano-Carcinoids*. **d**, Boxplots showing the tumor chronological time among *forte*, *mezzo-forte*, and *piano* harboring the same genomic alterations in LUAD from *Sherlock-Lung*. **e**, Scatter plot shows the significant negative correlation between median age estimated by clonal clock-like mutations and the median T/N TL ratio within the same genomic alteration groups as in **b** with frequency >3%.

Fig. 4

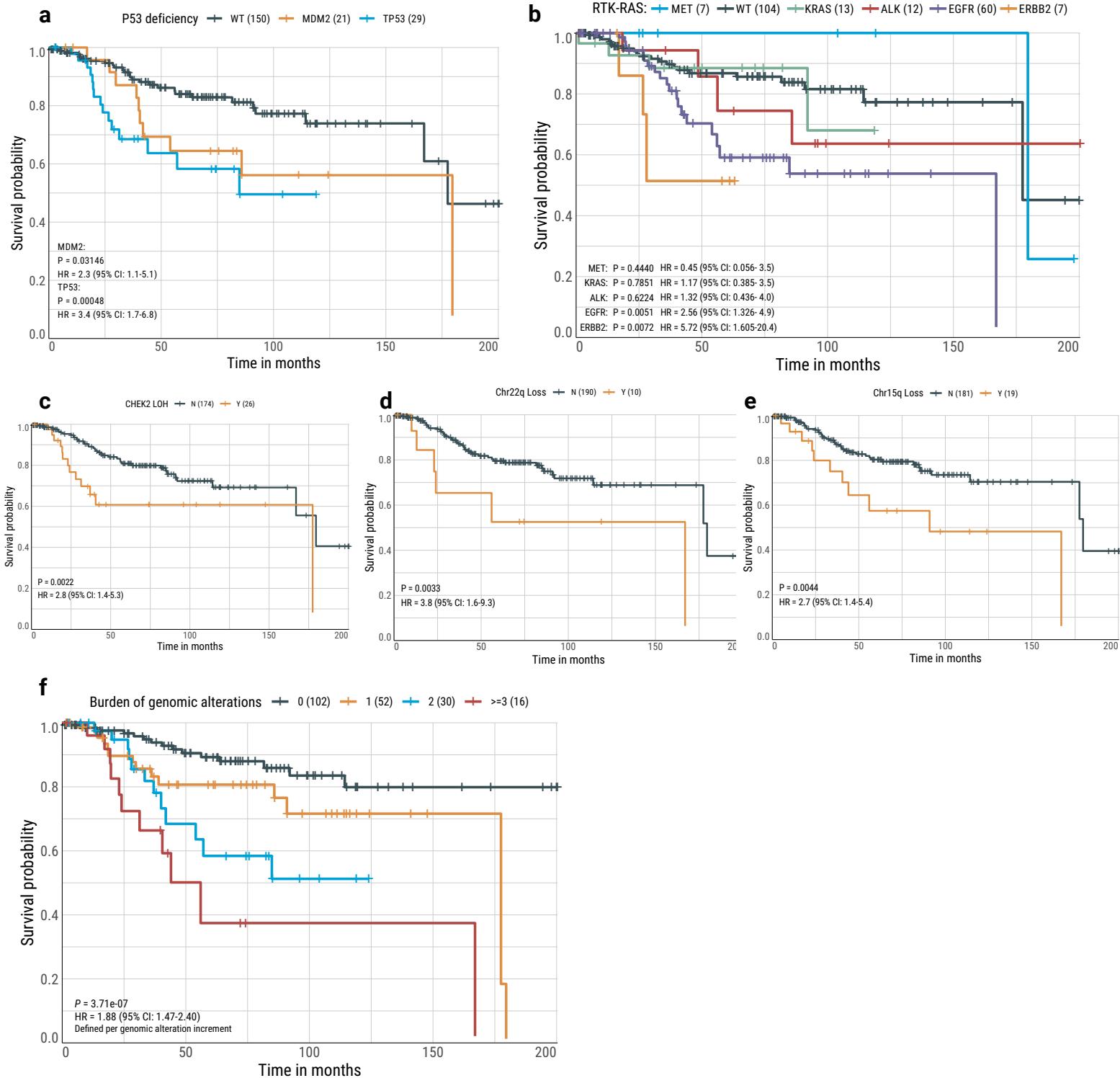
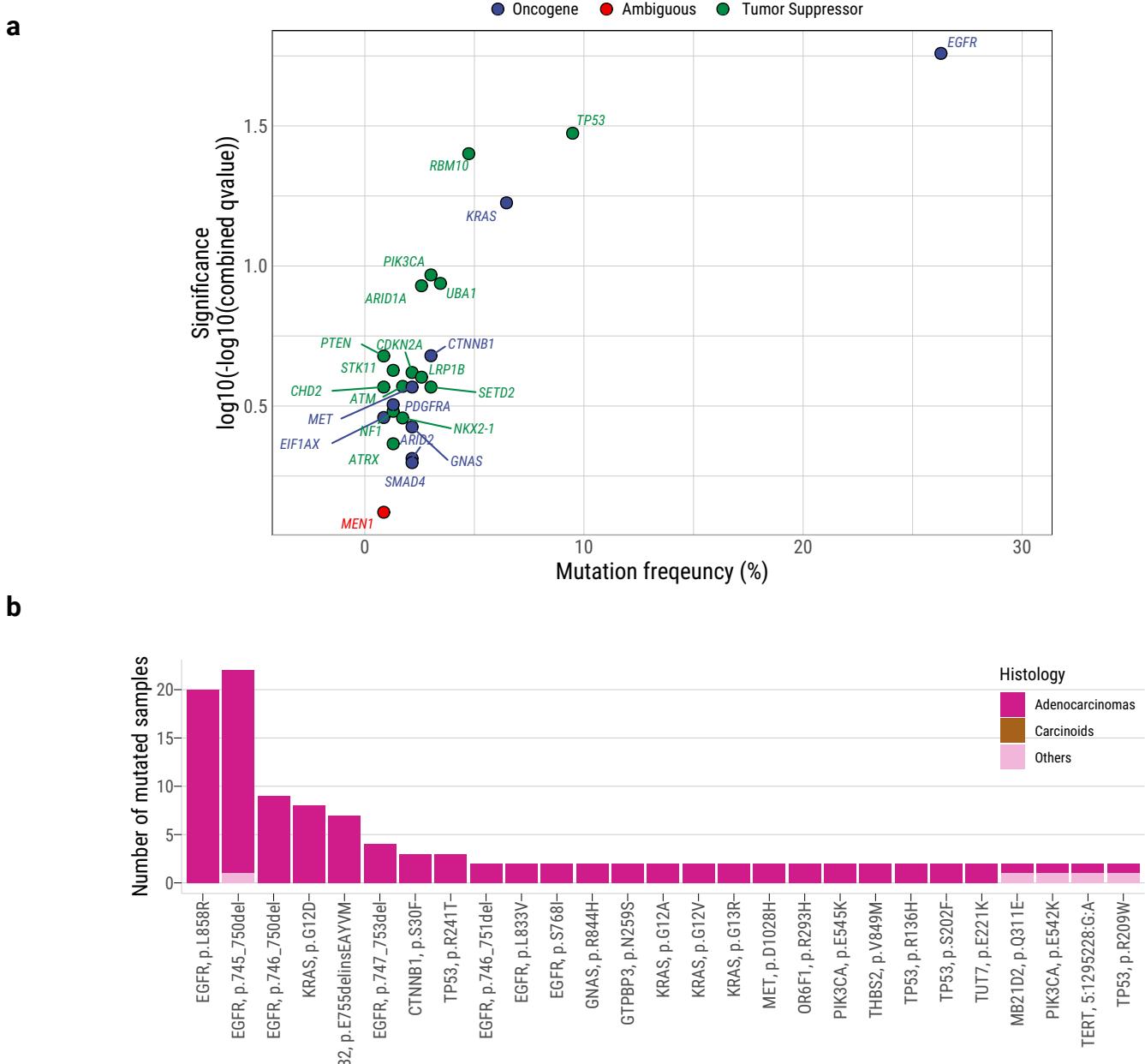


Fig. 4 Association between genomic aberrations and clinical outcomes in never smoker lung cancer patients.

Kaplan-Meier survival curves for overall survival stratified by (a) TP53 mutations and MDM2 amplification, (b) activation of individual driver genes in the RTK-RAS pathway, (c) CHEK2 LOH, (d) Chr22q loss, (e) Chr15q loss, and (f) Risk score based on the burden of five genomic alterations. P-values for significance and hazard ratios (HR) of difference are calculated using the log-rank test with adjustment for age, gender and tumor stage.

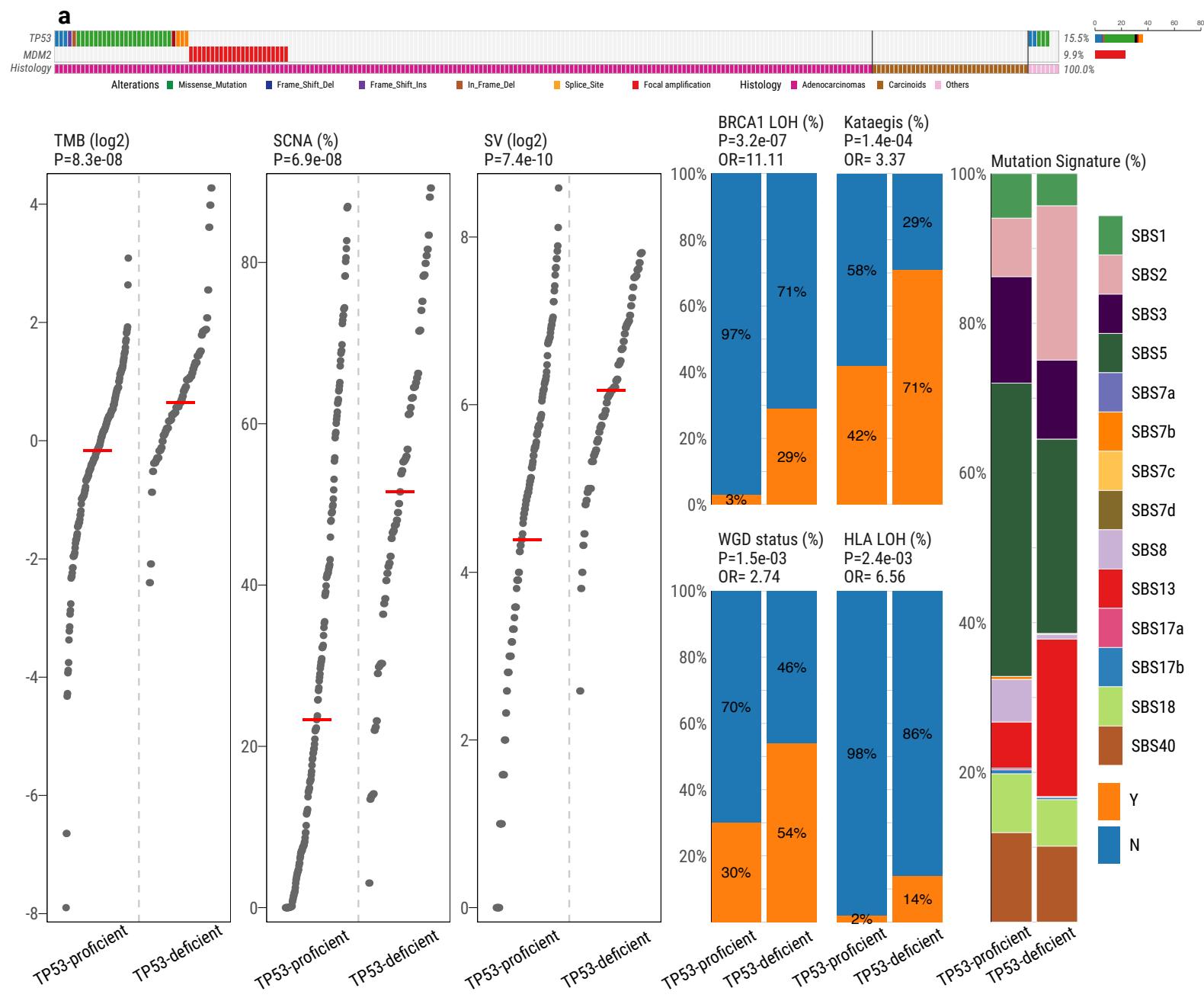
Extended Data Fig 1



Extended Data Fig. 1 Genes with signals of positive selection in Sherlock-Lung.

a, The scatter plot represents significance according to IntOGen q -value <0.05 (y-axis) and mutational frequency in the cohort (x-axis). Genes are colored according to their inferred mode of action in tumorigenesis. **b**, Recurrent non-synonymous driver mutations (in ≥ 2 patients).

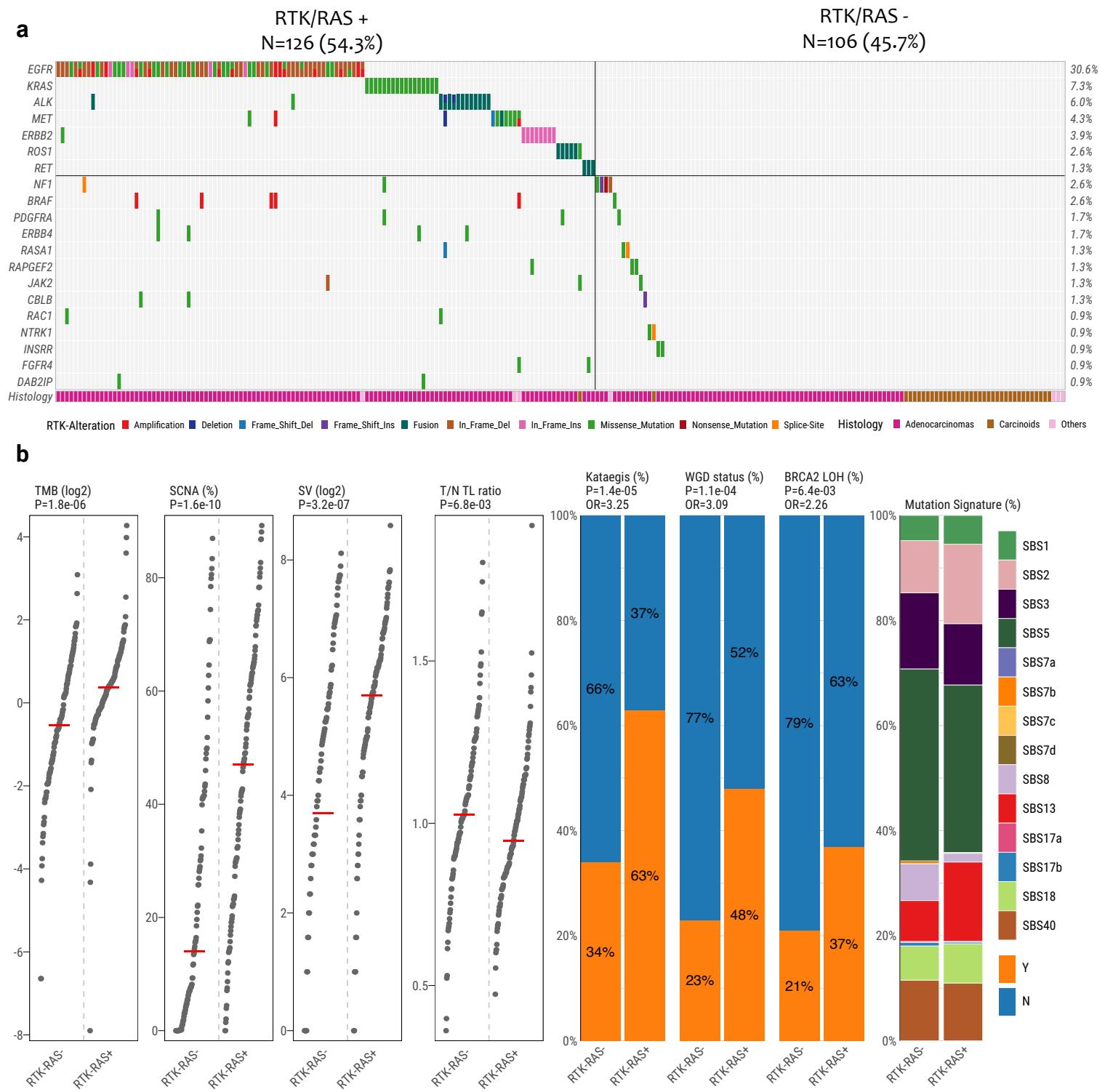
Extended Data Fig 2



Extended Data Fig. 2 Genomic alterations of TP53 pathway in Sherlock-Lung.

a, Oncoplot showing the mutual exclusivity between *TP53* mutations and *MDM2* amplification, which was used to define the *TP53* proficient and deficient groups. The bottom bar shows tumor histological types. **b**, Comparison of genomic features between *TP53*-proficient and *TP53*-deficient tumors. Left three panels: tumor mutation burden, percentage of genome with SCNA and SV burden. *P*-values are calculated using the two-sided Mann-Whitney U test. Middle four panels: enrichments for *BRCA1* LOH, Kataegis events, WGD events, and HLA LOH. *P*-values and *OR* are calculated using Fisher's exact test. Right panel: Contributions of each SBS signature.

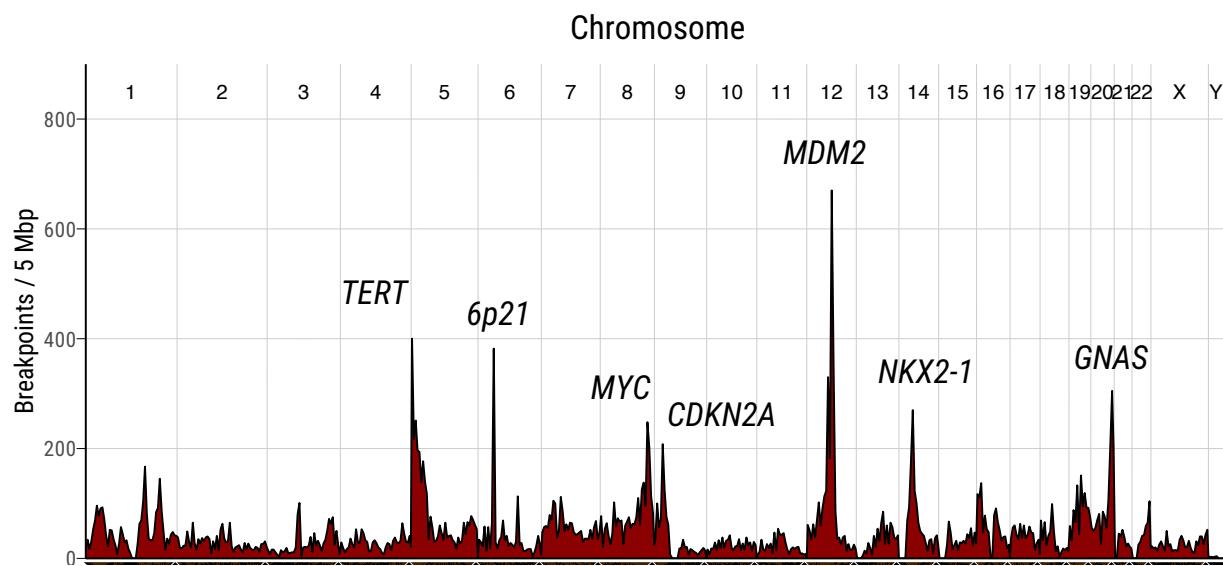
Extended Data Fig 3



Extended Data Fig. 3 Genomic alterations of RTK-RAS pathway in Sherlock-Lung.

a, Oncoplot showing mutual exclusivity of a few genes within the RTK-RAS pathway, which were used to define the RTK-RAS status. The bottom bar shows tumor histological types. **b**, Comparison of genomic features between RTK-RAS negative and positive tumors. Left four panels: tumor mutational burden, percentage of genome with SCNA, SV burden and T/N TL ratio. P -values are calculated using the two-sided Mann-Whitney U test; Middle three panels: enrichments for Kataegis events, WGD events, and *BRCA2* LOH. P -values and OR are calculated using Fisher's exact test; Right panel: Contributions of each SBS signature.

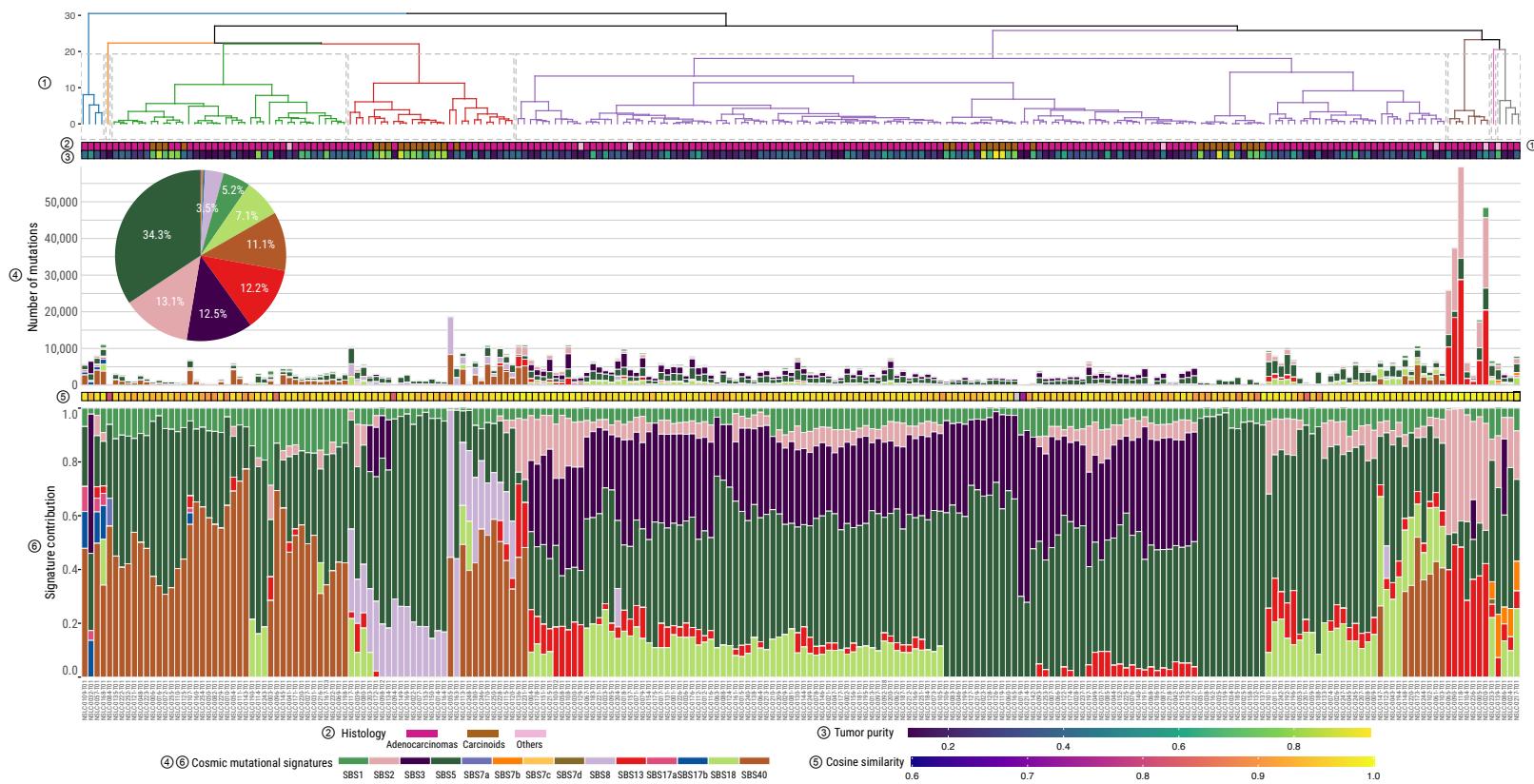
Extended Data Fig 4



Extended Data Fig. 4 Recurrence of SV breakpoints in Sherlock-Lung.

The frequencies of chromosomal breakpoints are calculated using 5 Mb as a window across the whole genome.

Extended Data Fig 5

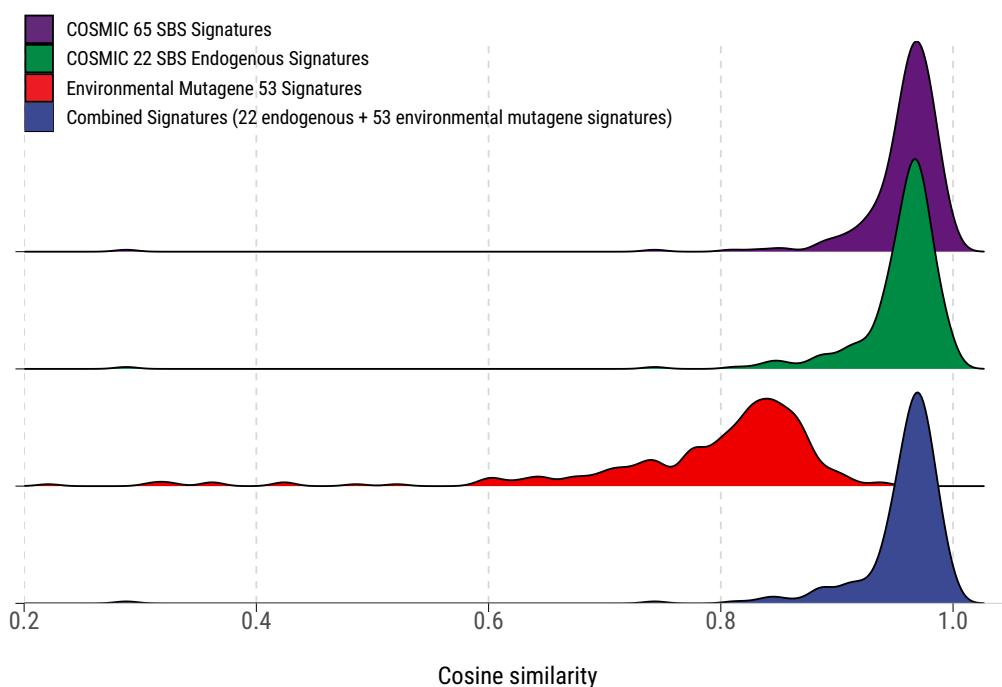


Extended Data Fig. 5 Landscape of mutational processes in Sherlock-Lung.

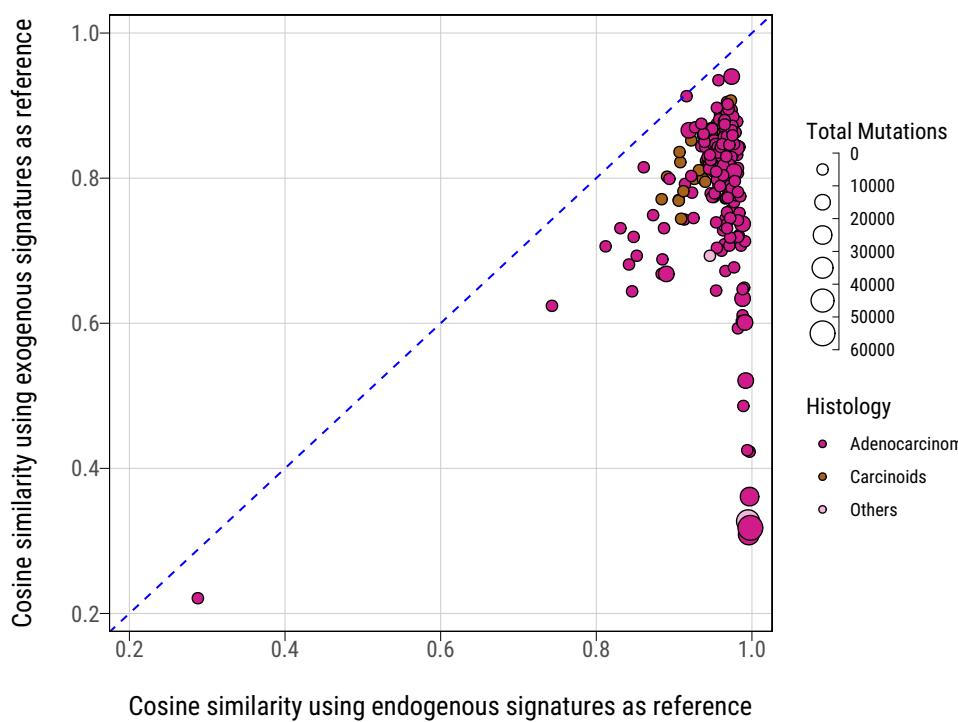
Mutational signature profile of single base substitutions (SBS) across 232 Sherlock-Lung samples. Panels from top to bottom: 1) Unsupervised clustering based on the proportion of SBS signatures; 2) Tumor histological type; 3) Tumor purity; 4) Pie chart showing the percentage of mutations contributed to each SBS signature and the barplot presenting the total number of SNVs assigned to each SBS signatures; 5) Cosine similarity between original mutational profile and signature decomposition result; 6) Proportions of SBS mutational signatures in each sample.

Extended Data Fig 6

a



b



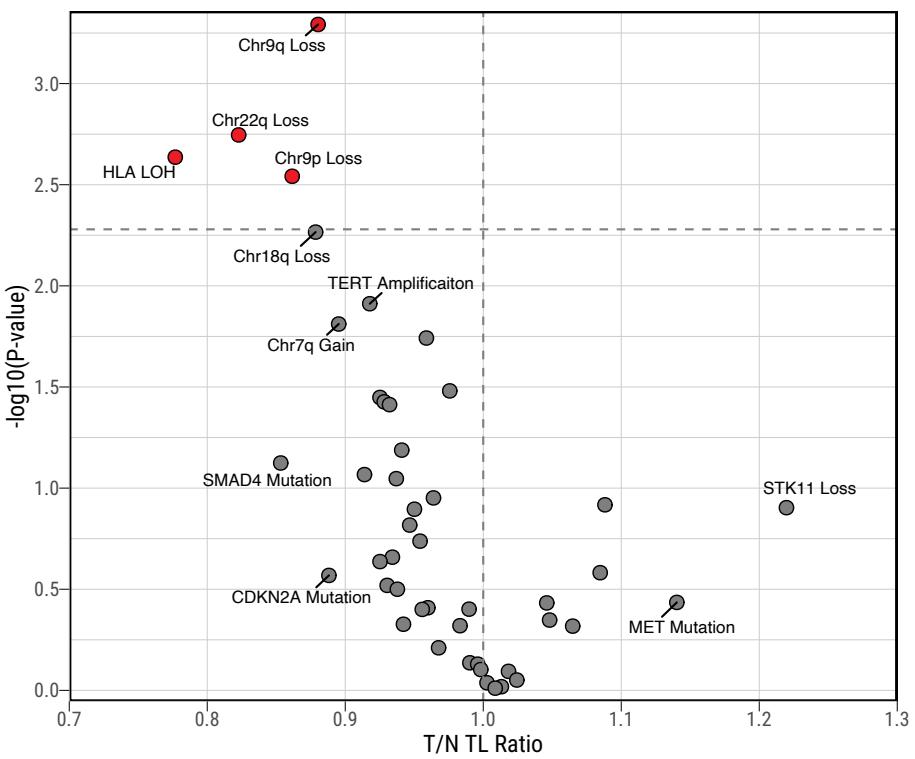
Extended Data Fig. 6 Dominant endogenous processes in Sherlock-Lung.

a, Density plot of cosine similarity between original mutational profile and reconstructed mutational profile using reference signatures from (top to bottom): 65 COSMIC SBS signatures, 22 COSMIC SBS signatures for endogenous processes, 53 Mutagene SBS signatures of environmental exposures, and a combined set of signatures including the 22 endogenous and 53 environmental exposure signatures.

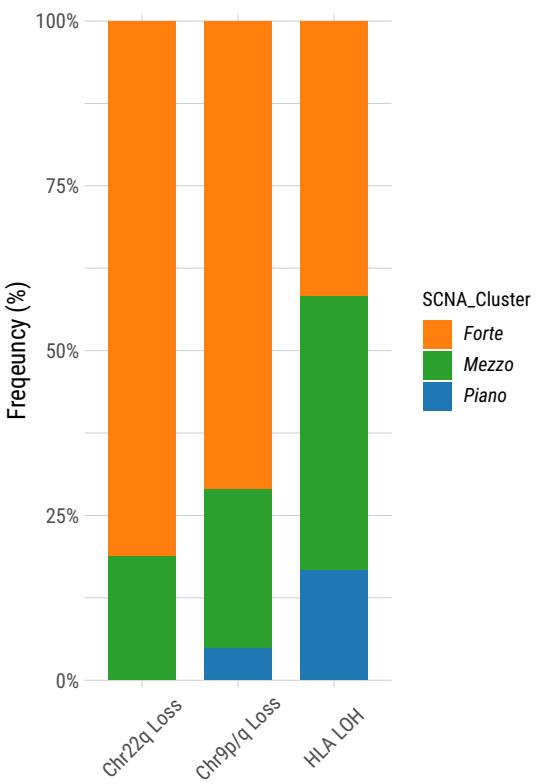
b, Comparison of the cosine similarity between the original mutational profiles and reconstructed mutational profiles using endogenous and exogenous signatures (similar to **a**). Each dot represents one sample. The size and color represent the total number of mutations and tumor histological type, respectively.

Extended Data Fig 7

a



b

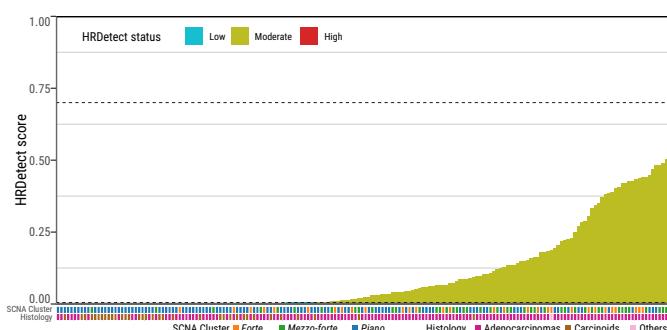


Extended Data Fig. 7 Association between T/N TL ratio and somatic alterations in Sherlock-Lung.

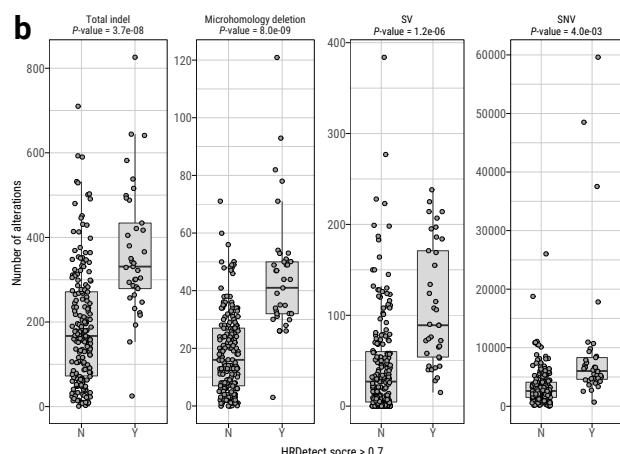
a, Scatterplot showing association between T/N TL ratio and somatic alterations. Association P -values (two-sided T-test; FDR adjusted using Benjamini-Hochberg method) are shown on the y-axis. Genomic alterations with $FDR \leq 0.1$ or $T/N TL$ ratio > 1.1 or < 0.9 are labeled and further highlighted in red when significant ($FDR=0.05$; horizontal dashed line). **b**, The proportion of each SCNA cluster among the group of tumors with somatic alterations significantly associated with shorten T/N TL including Chr22q Loss, Chr9p/q Loss or HLA LOH.

Extended Data Fig 8

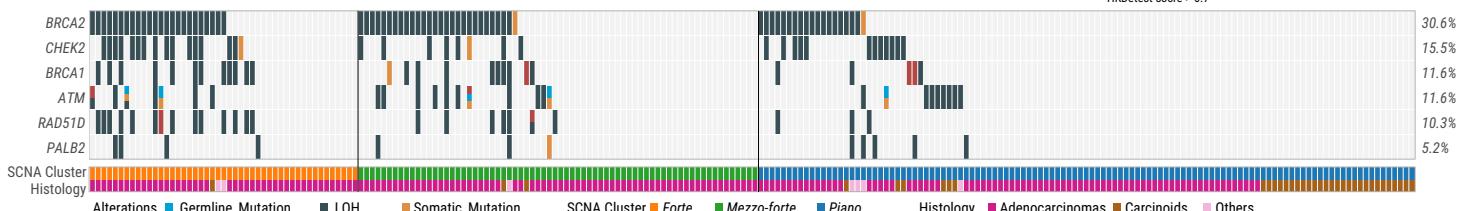
a



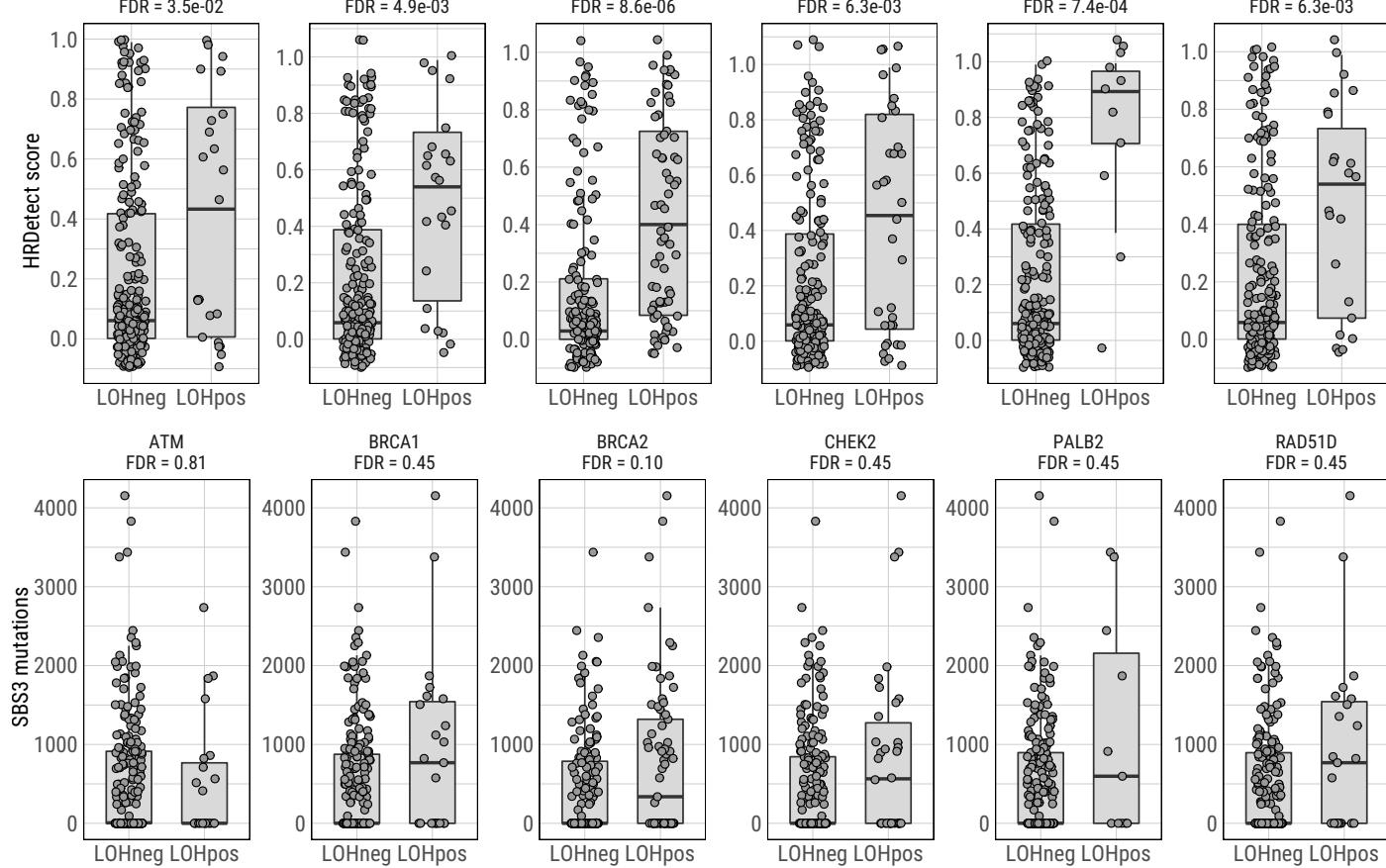
b



c



d

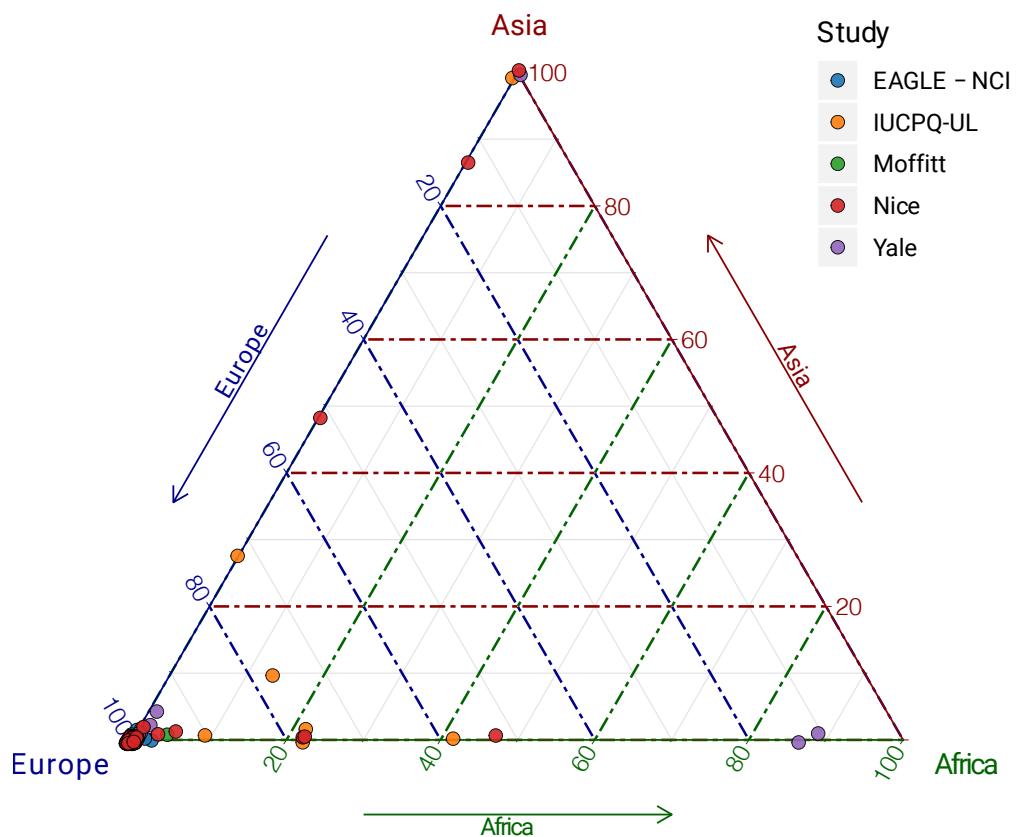


Extended Data Fig. 8 Homologous recombination deficiency (HRD) in Sherlock-Lung.

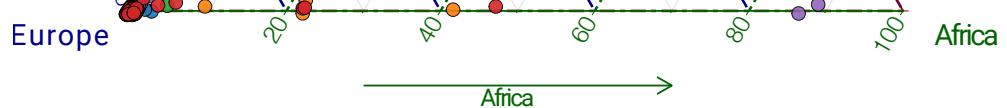
a, HRDetect scores of Sherlock-Lung samples. HRD-high: >0.7 , HRD-low: < 0.005 , based on Davies et al.⁴¹. **b**, Comparison of the number of total indels, microhomology deletions, SVs, and SNVs between samples with HRDetect score below 0.7 (group N) and above 0.7 (group Y). **c**, Oncoplot of genomic alterations in HRD associated genes, including germline mutations, somatic mutations and LOH. Samples with biallelic alterations are represented by bars with two different colors. The bottom bar shows tumor histological types. **d**, Boxplots of HRDetect scores (top) and SBS mutation loads (bottom) in tumors with and without LOH of six HR associated genes. *P*-values are calculated using the two-sided Mann-Whitney U test.

Supplementary Fig. 1

a



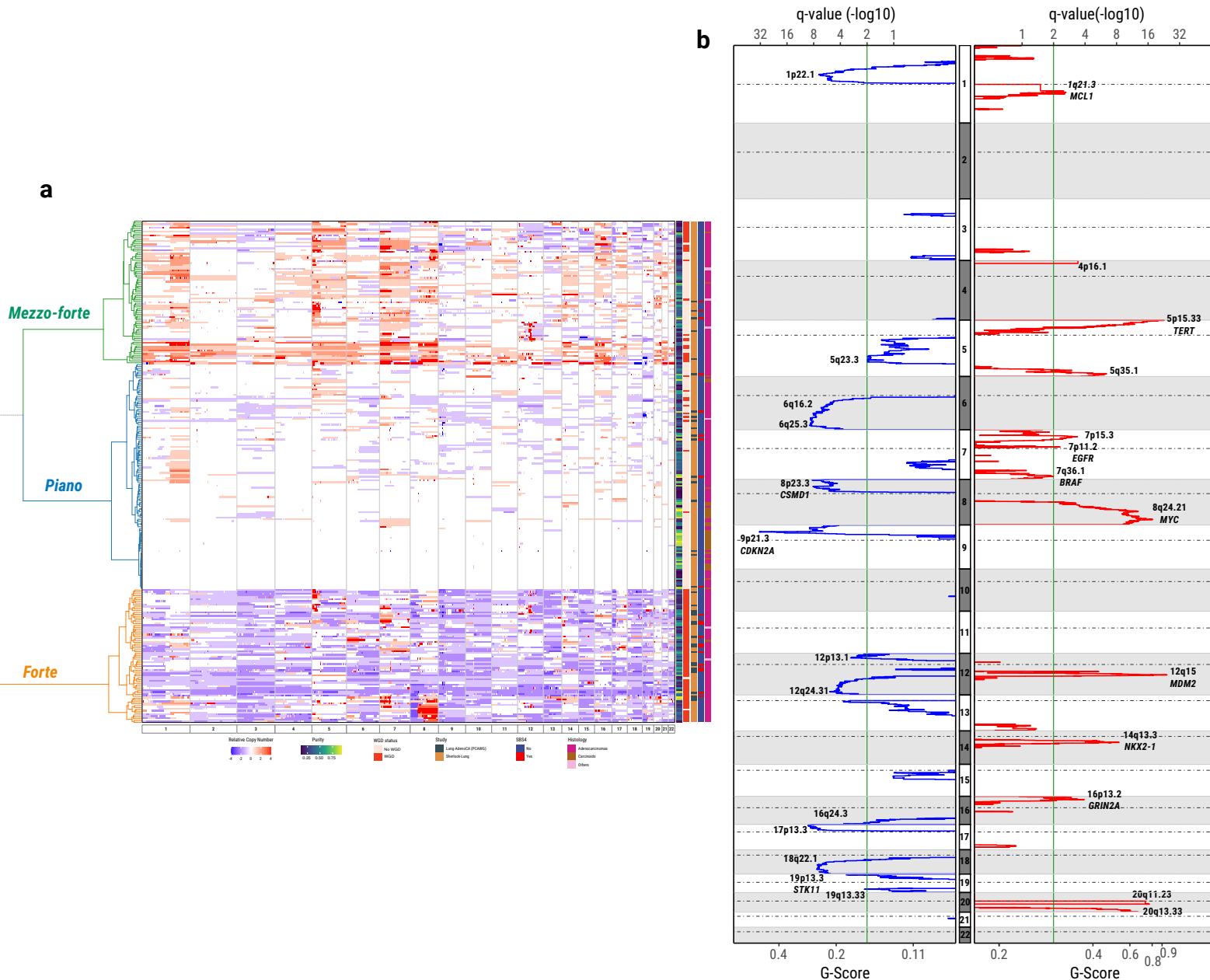
b



Supplementary Fig. 1 Patients' ancestry proportions based on ADMIXTURE analysis.

Each dot represents an individual color-coded by study centers.

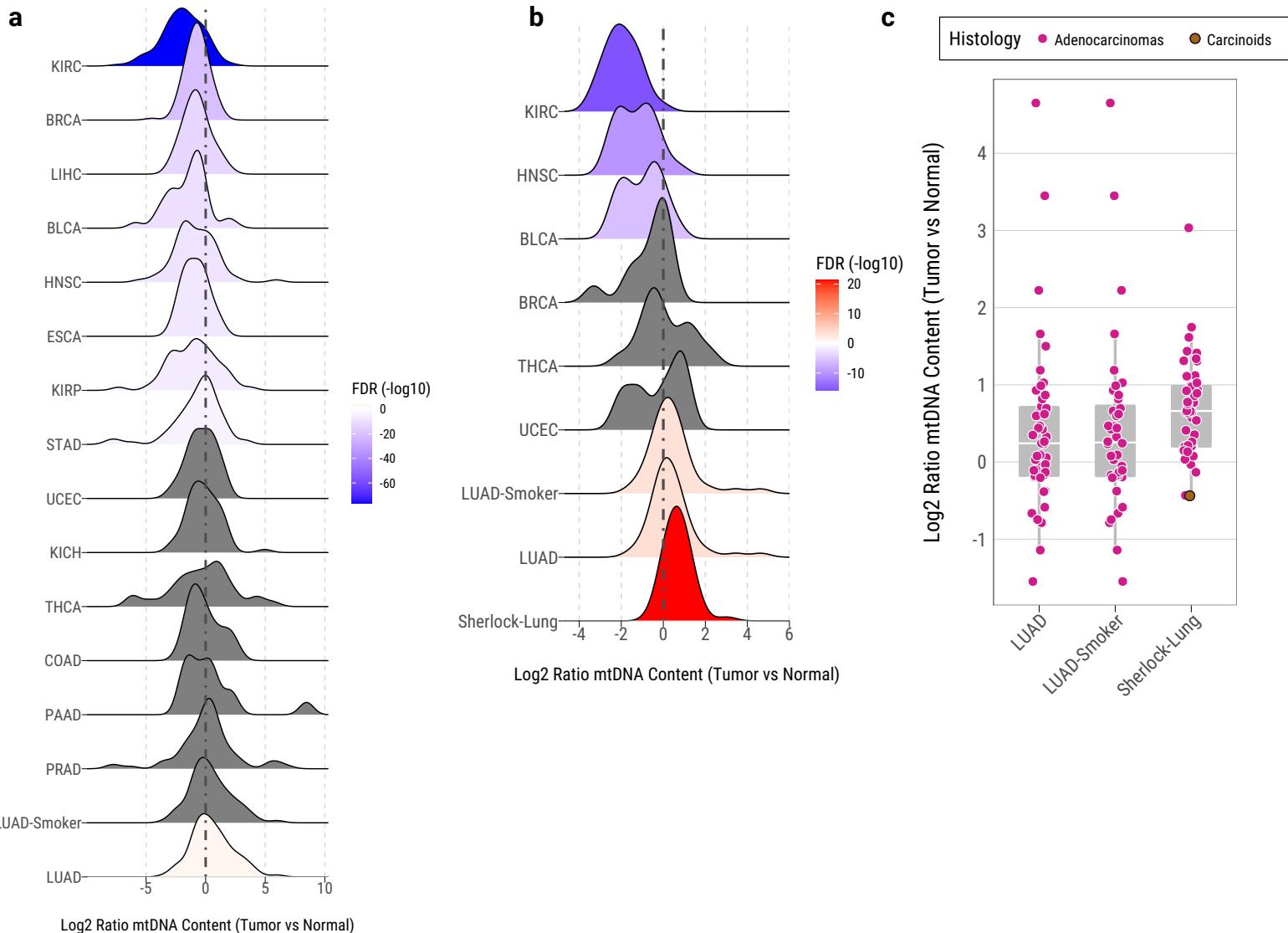
Supplementary Fig. 2



Supplementary Fig. 2 Somatic copy number alterations analysis in both arm- and focal-level.

a, The landscape of somatic copy number alterations combining 38 TCGA LUAD samples from PCAWG with the Sherlock-Lung samples. Left panel presents unsupervised clustering of arm-level SCNA events: forte, mezzo-forte and piano. Relative copy number is calculated as: total copy number - ploidy (non-WGD=2 and WGD=4). Samples in rows are annotated by tumor purity, WGD status, study, and presence of SBS4 signatures and tumor histological type. **b**, GISTIC analysis of focal amplifications (red) and deletions (blue) in Sherlock-Lung. Significantly amplified and deleted ($q\text{-value} < 0.01$, green line) regions are annotated with cytoband or gene names.

Supplementary Fig. 3

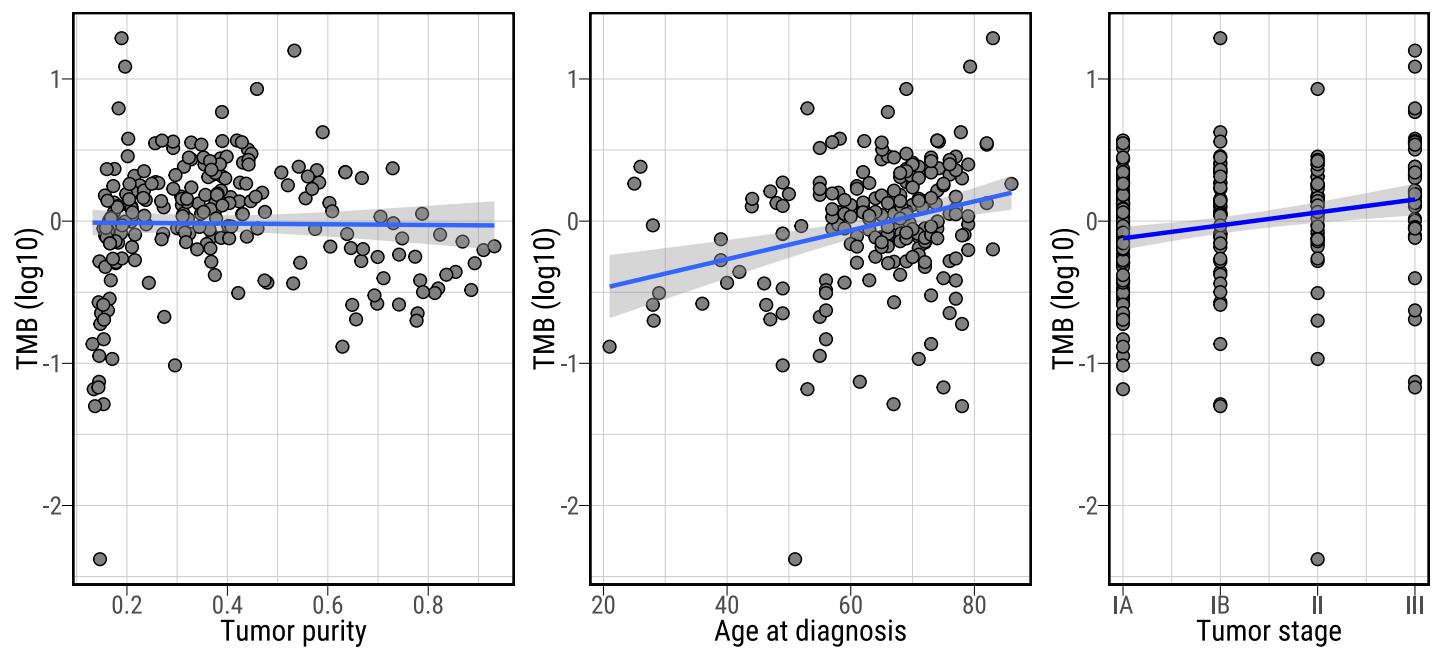


Supplementary Fig. 3 Mitochondrial DNA copy number accumulation in tumor compared to normal tissue.

a, Accumulation (light red) or depletion (blue) of mtDNA in tumor samples relative to the adjacent normal tissues based on WES data from the TCGA study. Normalized density plots illustrate log₂ tumor/normal ratio of mtDNA content. Each row represents a tumor type. Statistical significance of trends is assessed using Wilcoxon sign rank test, and P-values are corrected using the Benjamini-Hochberg procedure.

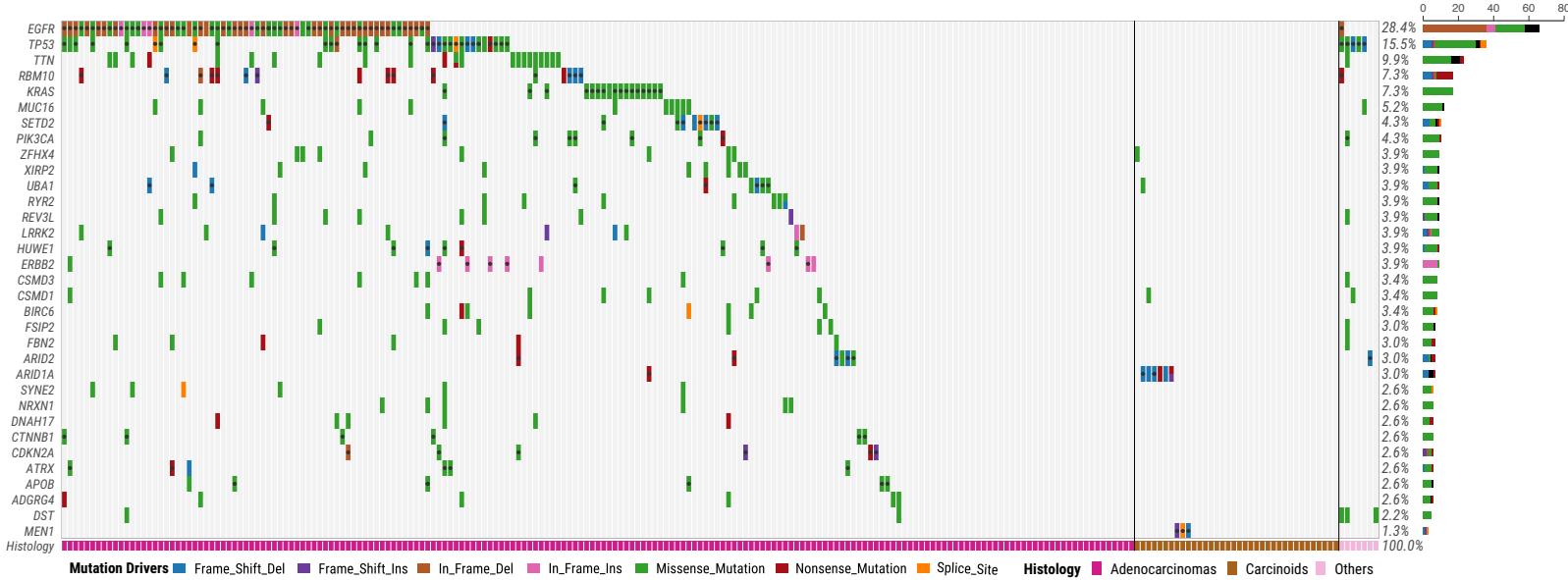
b, The accumulation (red) or depletion (blue) of mtDNA in tumor samples relative to adjacent normal tissue based on the WGS data from the TCGA and Sherlock-Lung studies. **c**, Comparison of T/N mtDNA copy number ratio (log₂) among TCGA LUAD, TCGA LUAD smokers only and Sherlock-Lung.

Supplementary Fig. 4

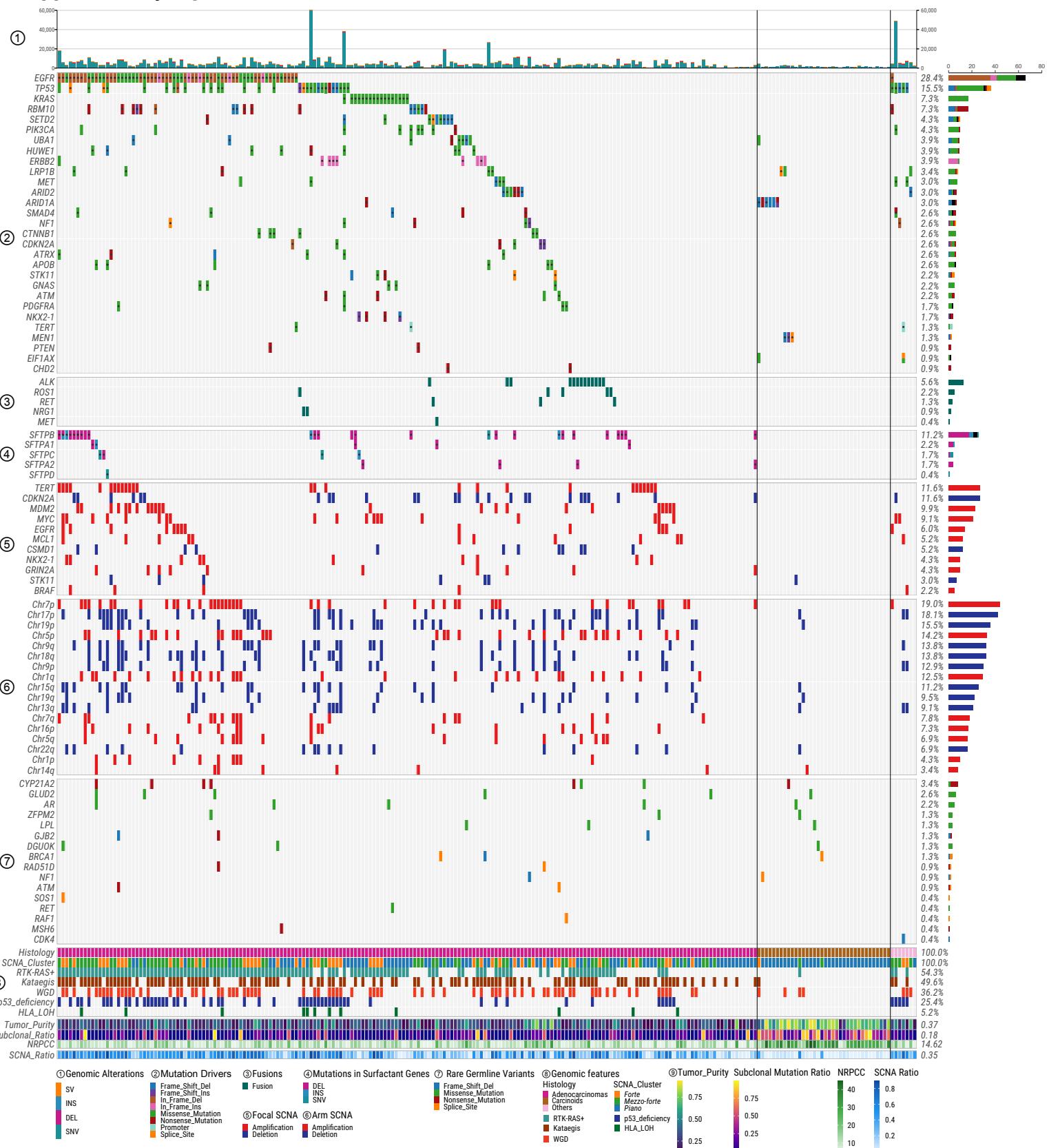


Supplementary Fig. 4 Scatter plot of the correlations between TMB (\log_{10}) and tumor purity, age at diagnosis, and tumor stage.

Supplementary Fig. 5



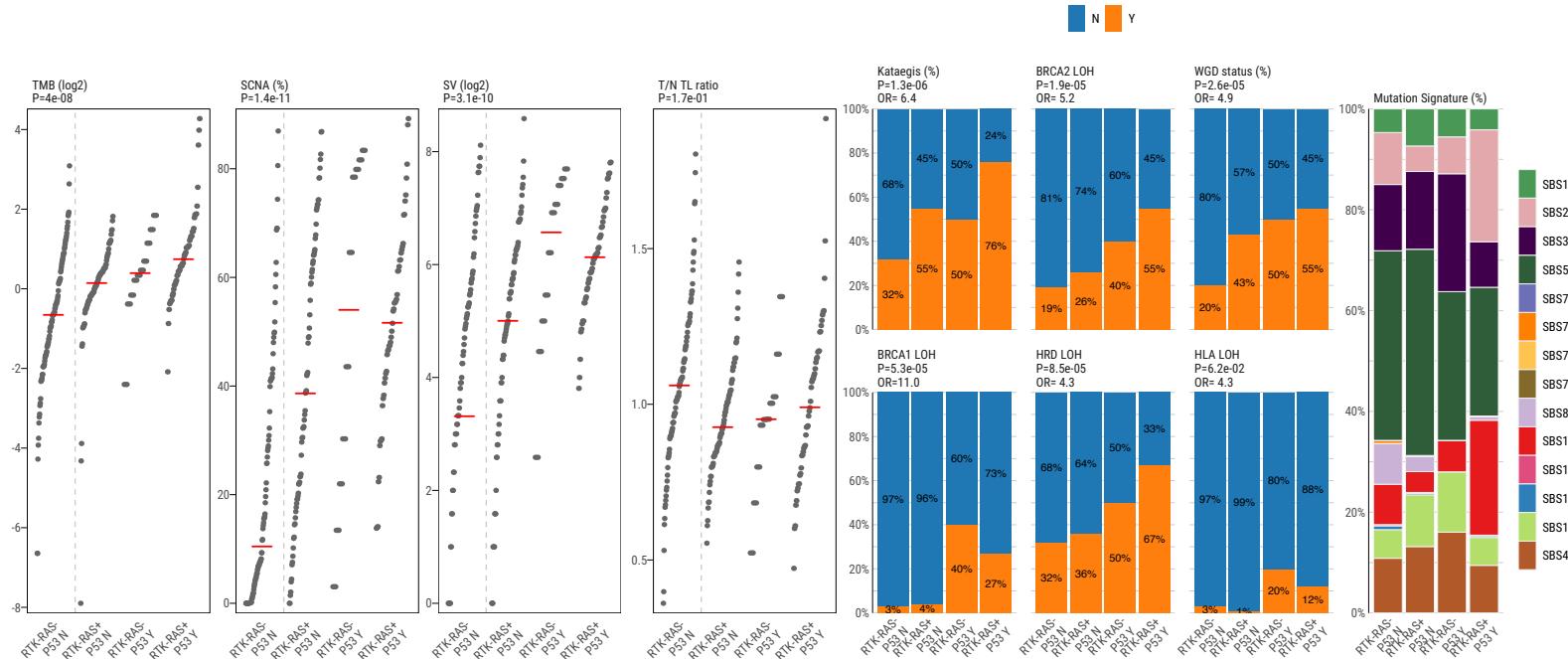
Supplementary Fig. 6



Supplementary Fig. 6 Summary of genomic alterations and features in Sherlock-Lung.

Panels from top to bottom describe: 1) distribution of genomic alteration numbers; 2) most frequently mutated or potential driver genes; 3) oncogenic fusions; 4) somatic mutations in surfactant associated genes; 5) significant focal SCNAs; 6) significant arm-level SCNAs; 7) genes with rare germline mutations; 8) and 9) different genomic features. The numbers on the right panel show the overall frequency (1-8) or median values (9). NRPCC: the number of reads per clonal copy.

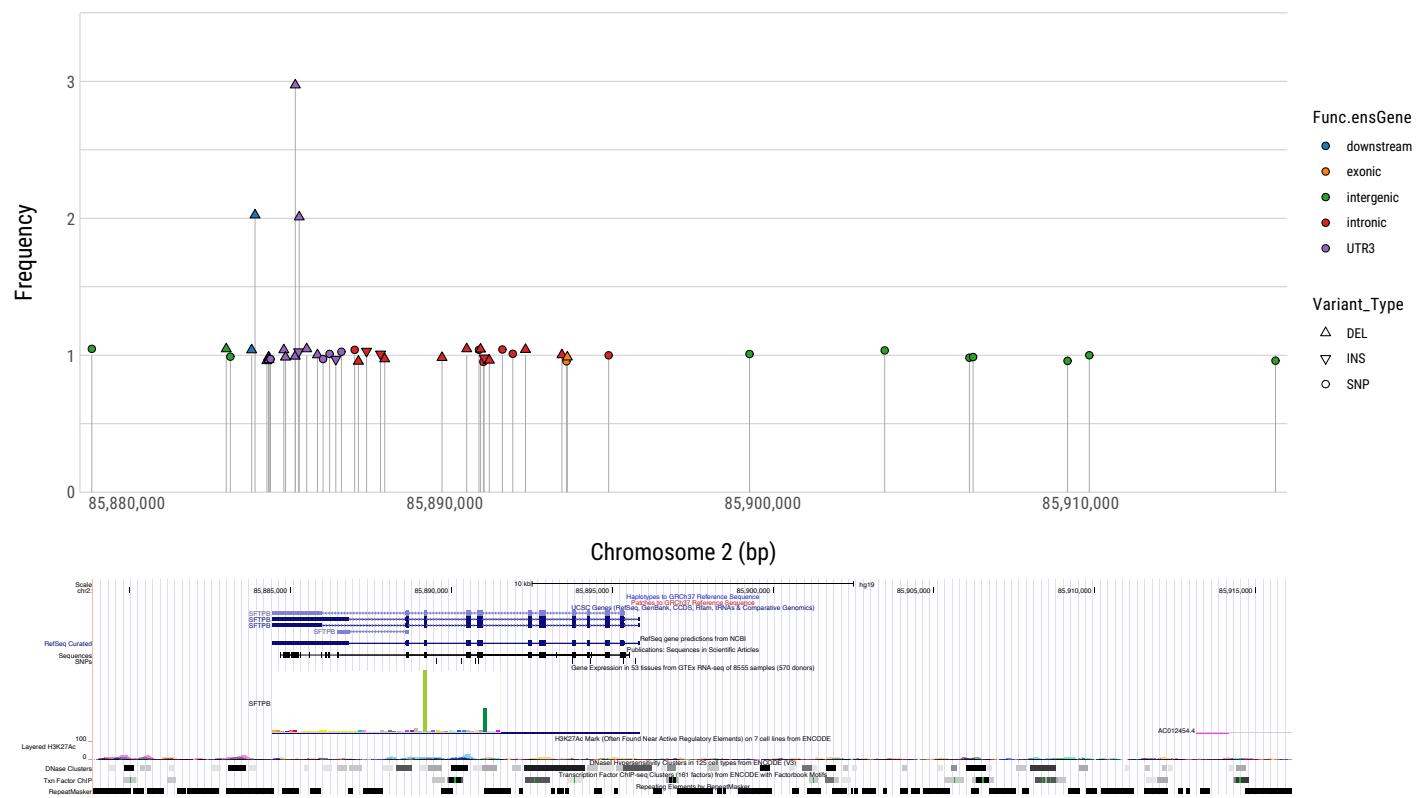
Supplementary Fig. 7



Supplementary Fig. 7 Comparison of genomic features combining TP53 and RTK-RAS status.

Left four panels: tumor mutation burden, percentage of genome with SCNA, SV burden and T/N TL ratio. P -values are calculated using the two-sided Mann-Whitney U test. Middle six panels: enrichments for Kataegis events, *BRCA2* LOH, WGD events, *BRCA1* LOH, HRD LOH, and HLA LOH. P -values and *OR* are calculated using Fisher's exact test Right panel: Contributions of each SBS signature. All statistical analyses were performed between "RTK-RAS- & TP53 N" and "RTK-RAS+ & TP53 Y".

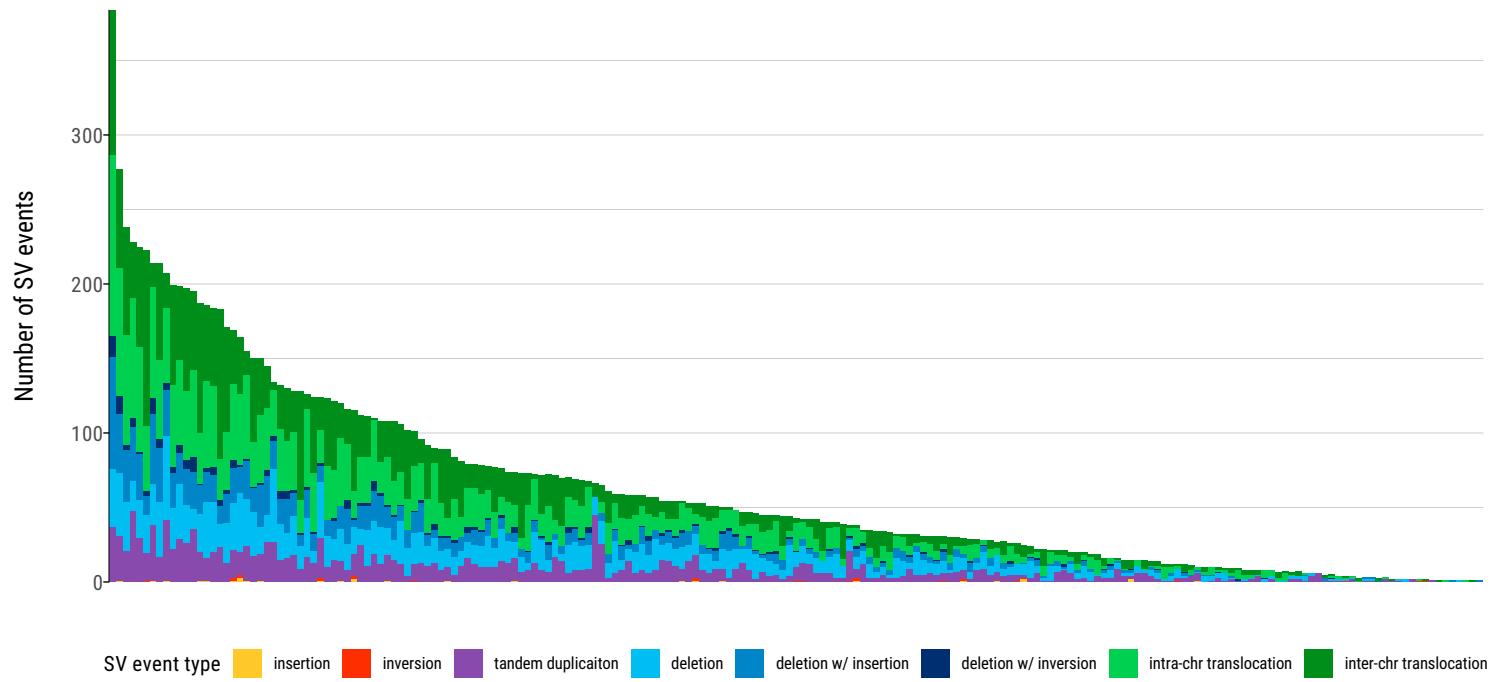
Supplementary Fig. 8



Supplementary Fig. 8 Identification of somatic mutations in the surfactant associated gene SPFTB.

Mutation functional classifications and types are highlighted in different colors and shapes, respectively. Individual color-coded by study centers.

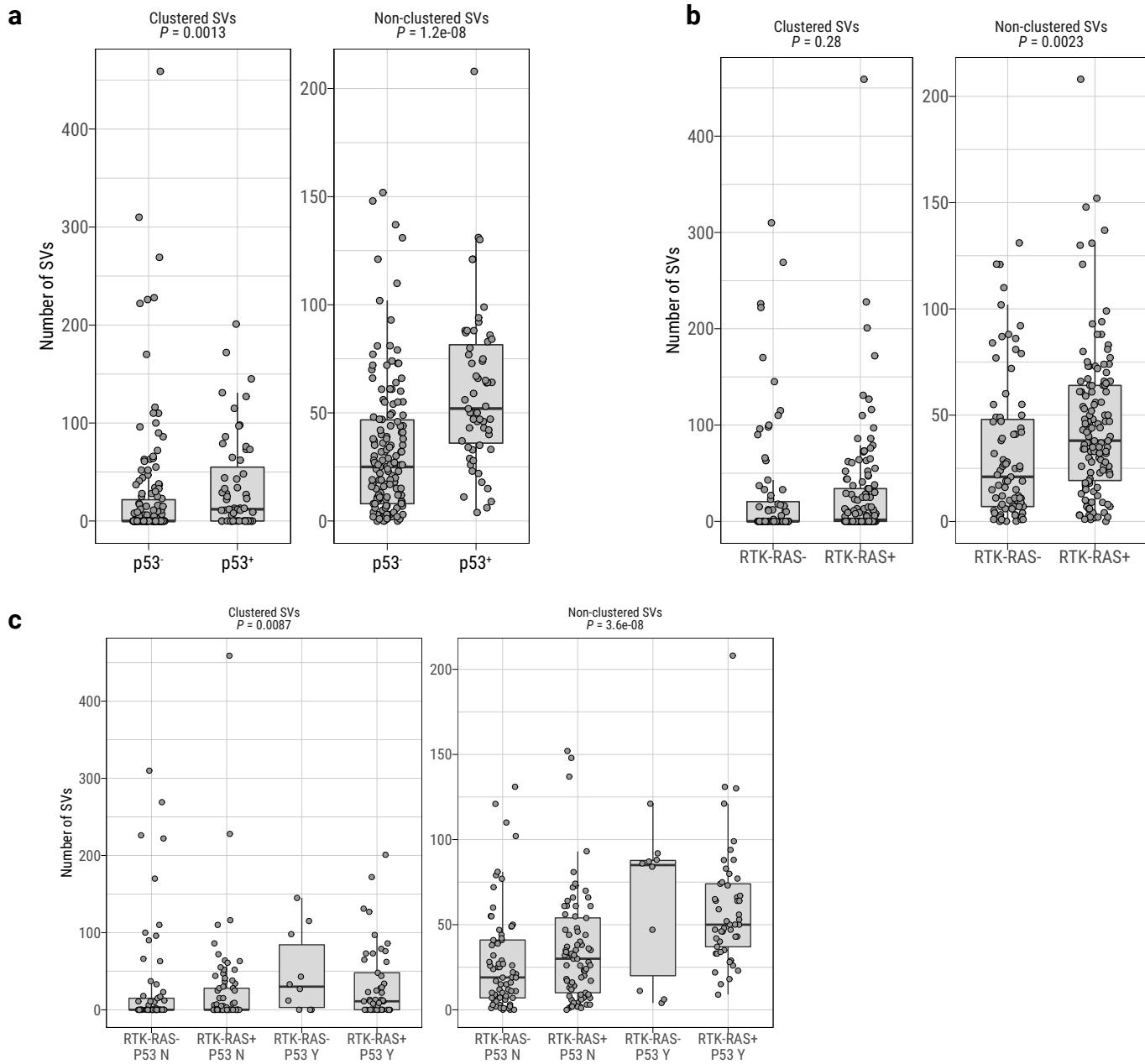
Supplementary Fig. 9



Supplementary Fig. 9 Frequencies of types of somatic SVs in each patient.

The y-axis shows the number of events. The colors indicate different SV types.

Supplementary Fig. 10



Supplementary Fig. 10 Enrichment of non-clustered SVs in tumors with somatic alteration in TP53 or RTK-RAS pathway.

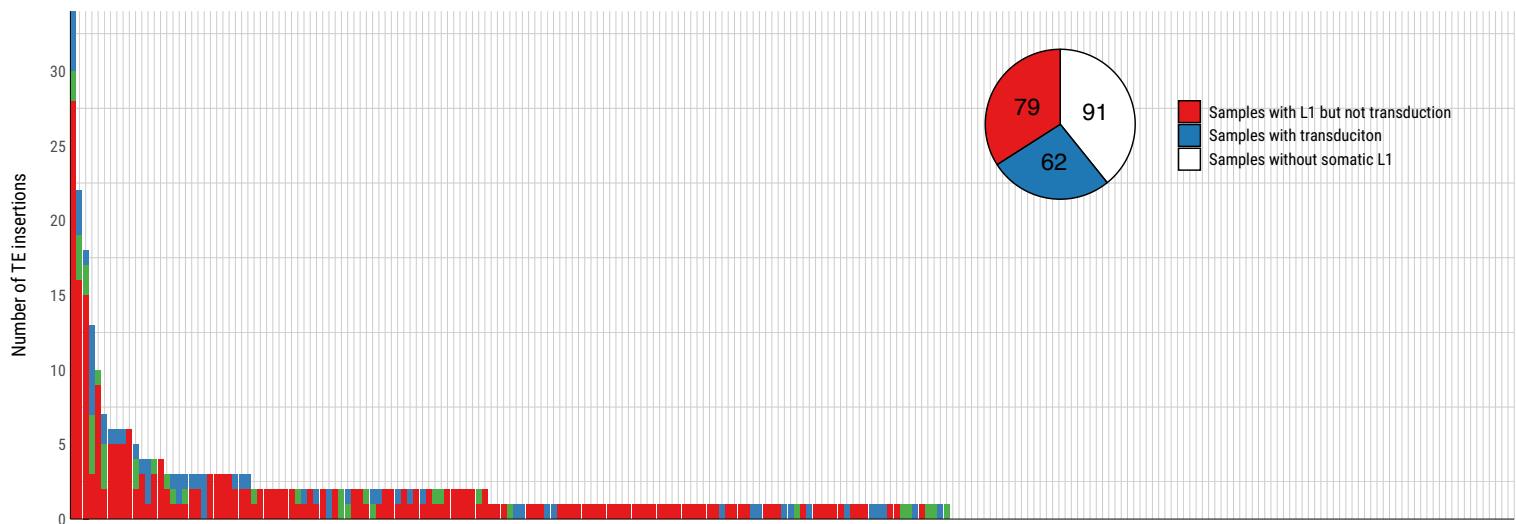
a, Comparison of the total number of clustered SVs (left) and non-clustered SVs (right) between TP53-proficient (group N) and TP53-deficient (group Y) tumors. **b**, Comparison of the total number of clustered SVs (left) and non-clustered SVs (right) between RTK-RAS negative and positive tumors. **c**, Comparison of the total number of clustered SVs (left) and non-clustered SVs (right) among subtypes combining TP53 and RTK-RAS status. P -values are calculated using the two-sided Mann-Whitney U test between “RTK-RAS- & TP53 N” and “RTK-RAS+ & TP53 Y”.

Supplementary Fig. 11

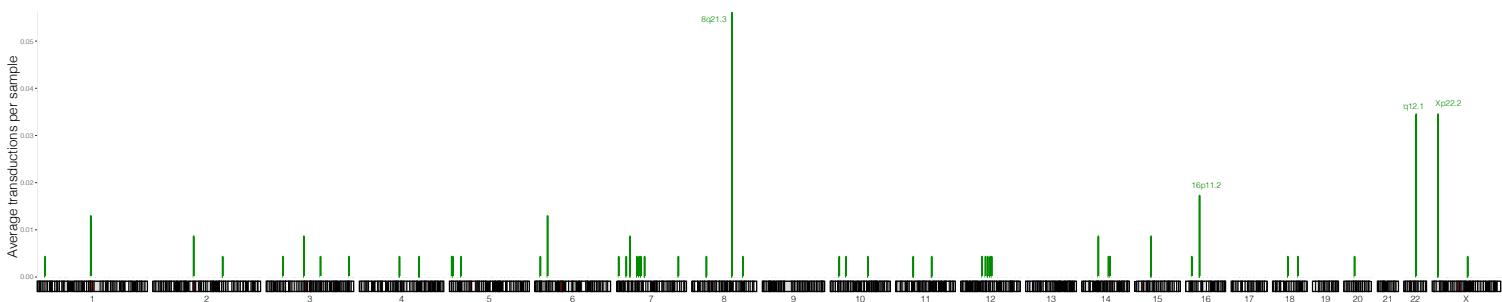
The somatic L1 retrotransposition activity

a

Type Solo-L1 Partnered-3' transduction Orphan-3' transduction



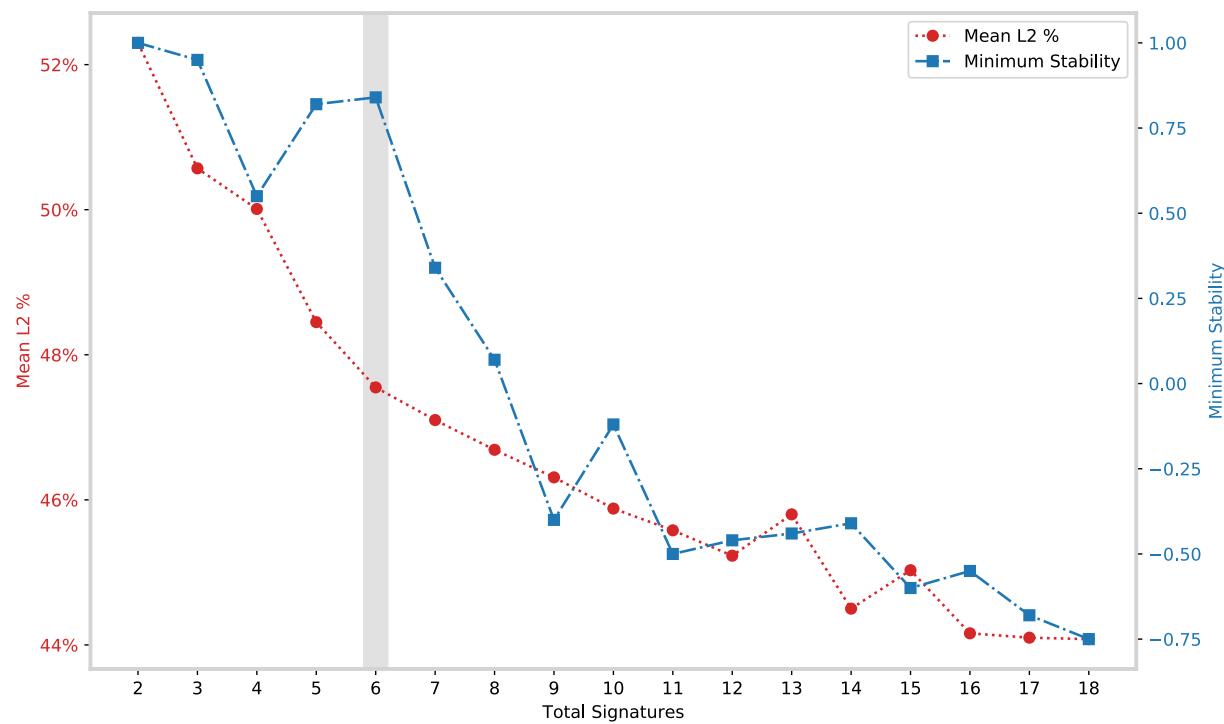
b



Supplementary Fig. 11 Activity of L1 retrotransposition insertion.

a, Frequencies of L1 retrotransposition activity. Pie chart represents the proportion of samples with transduction (blue), at least one solo L1 but no transduction (red) and no L1 retrotransposition (white). Bar plot represents the numbers of transposon insertions (y-axis) in each patient (x-axis). **b**, Summary of the rate of source element activity. Average number of transductions involving the given source element per sample (y-axis) are shown across the chromosomes (x-axis).

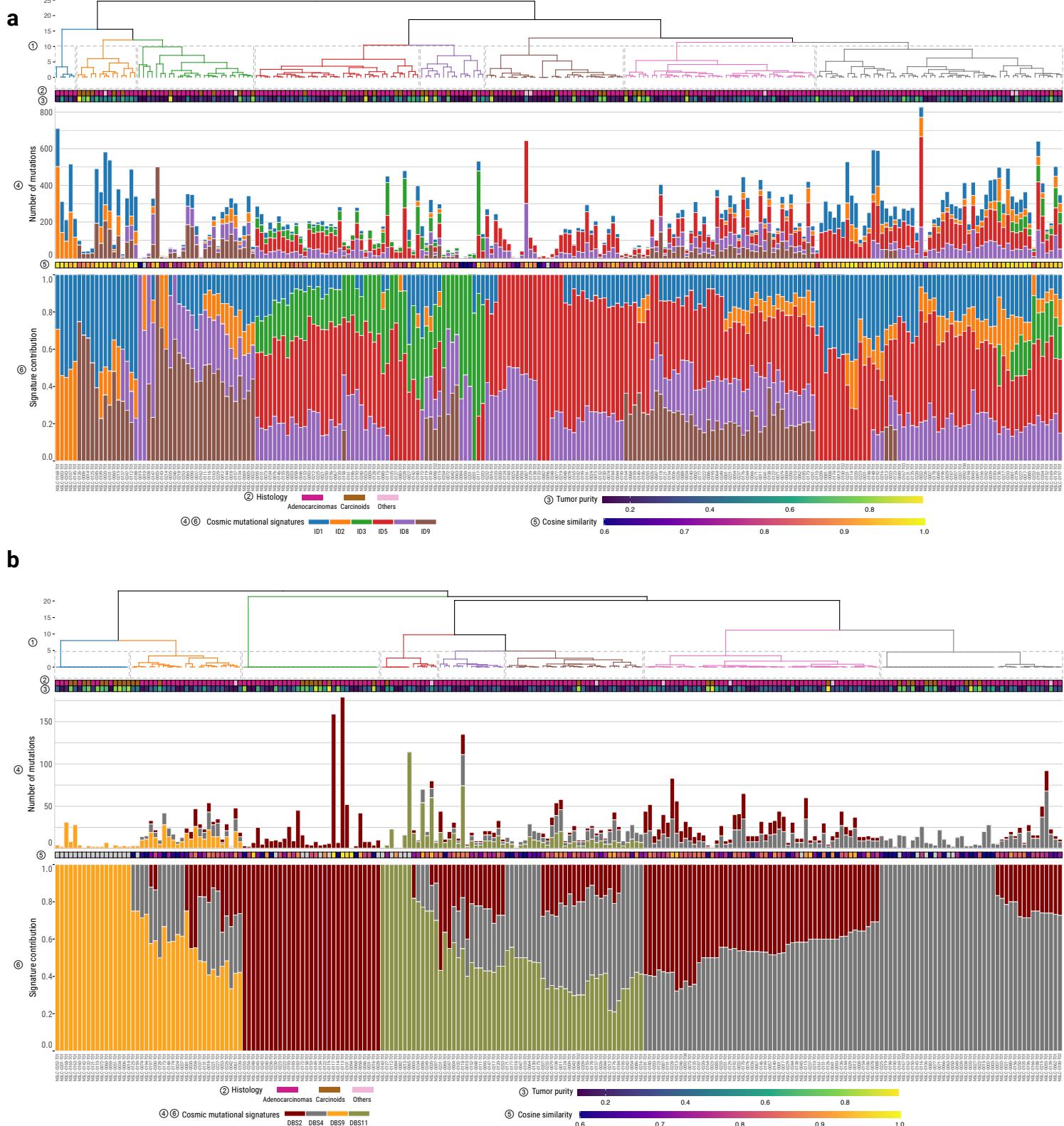
Supplementary Fig. 12



Supplementary Fig. 12 De-novo extraction of mutational signatures.

Assessment of signature decomposition stability during *de-novo* SBS signature extraction using SBS-1536 profile. Mean L2 (red) is calculated as the square root of the sum of the squared vector values, and minimum stabilities (blue) are calculated as average silhouette coefficient. The gray bar indicates the final selection for the total number of *de-novo* signatures.

Supplementary Fig. 13



Supplementary Fig. 13 Mutational signature profile of indels (ID) and doublet substitutions (DBS).

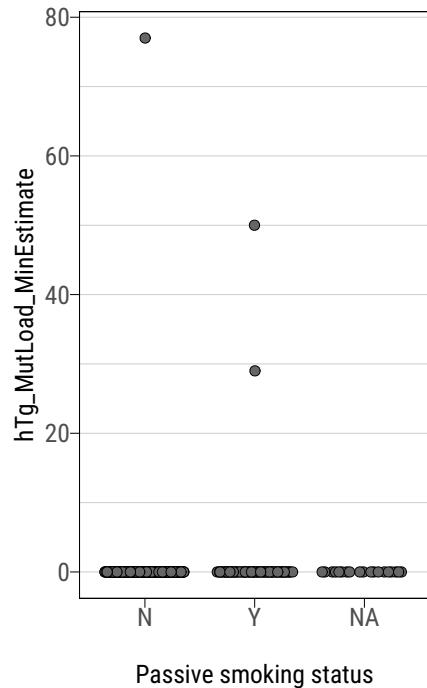
a, Mutational signature profile of indels across 232 Sherlock-Lung samples. Panels from top to bottom: 1) Unsupervised clustering based on the proportion of ID signatures decomposed by SigProfilerExtractor; 2) Histology; 3) Tumor purity; 4) Total number of ID assigned to each ID signature; 5) Cosine similarity between the original mutational profile and signature decomposition results; 6) Proportions of ID mutational signatures. **b**, Mutational signature profile of doublet substitutions across 232 Sherlock-Lung samples. Panels from top to bottom: 1) Unsupervised clustering based on the proportion of DBS signatures decomposed by SigProfilerExtractor; 2) Histology; 3) Tumor purity; 4) Total number of DBS assigned to each DBS signature; 5) Cosine similarity between the original mutational profile and signature decomposition results; 6) Proportions of DBS mutational signatures.

Supplementary Fig. 14

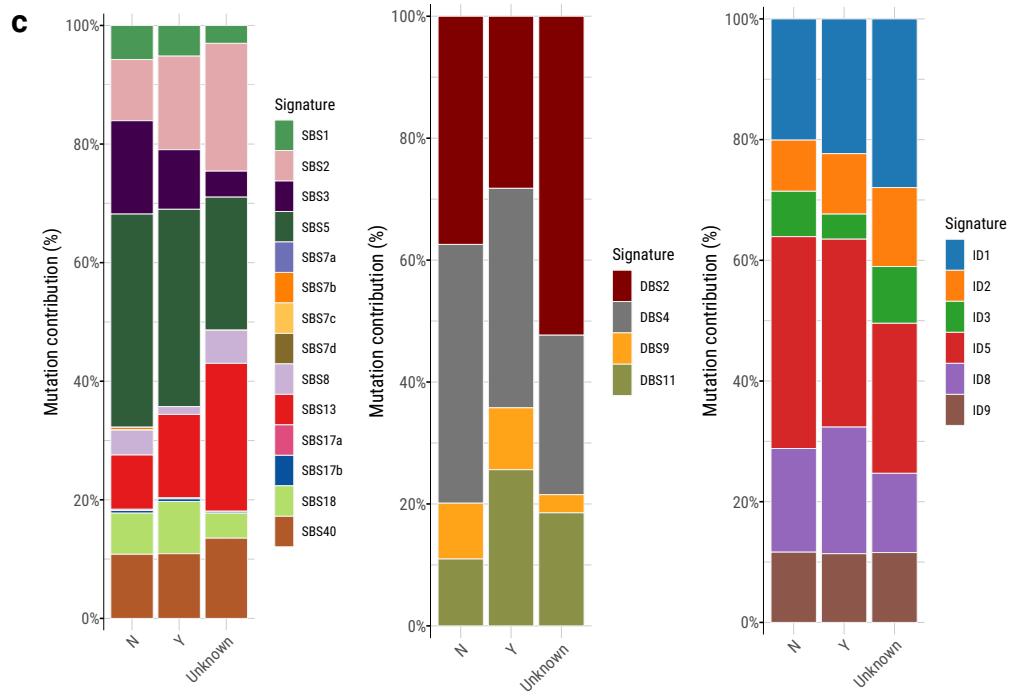
a



b



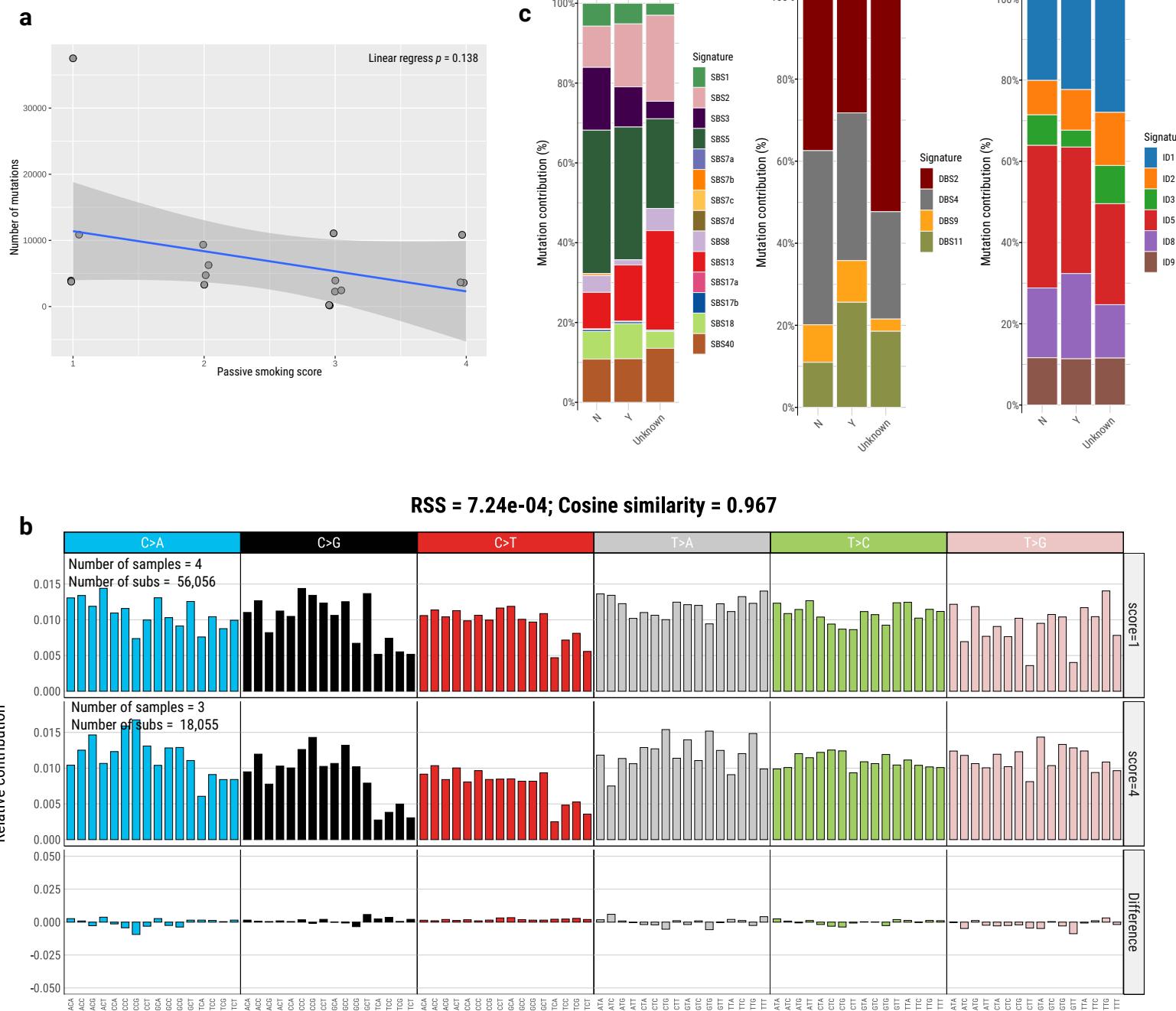
c



Supplementary Fig. 14 Mutational signature comparison between passive-smokers and non-passive smokers.

a, Spectra (left) and transcriptional strand bias analysis (right) of SBS mutations identified in passive-smokers (top) and non-passive smokers (bottom). **b**, Identification of alkylation-induced mutagenesis ($hTg \rightarrow hGg$ signature) among passive smokers (Y), non-passive smokers (N) and rest of patients (NA). Mutation loads identified by P-MACD are shown. **c**, Proportions of SBS, DBS and ID mutational signatures identified in samples from passive smokers (Y, n=64), non-passive smokers (N, n=149) and passive smoking status unknown (Unknown, n=19).

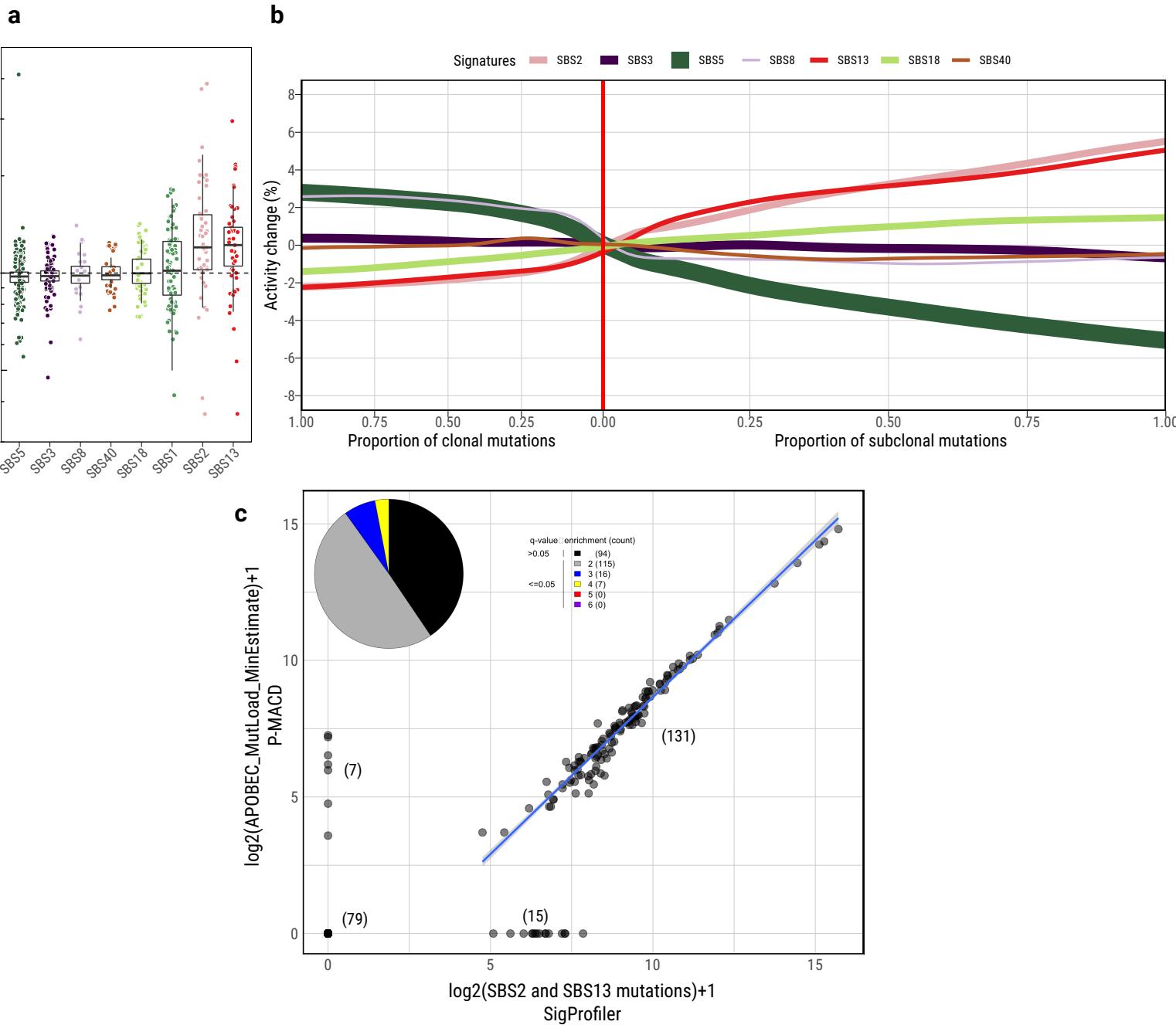
Supplementary Fig. x



Supplementary Fig. x Mutational signature comparison between passive-smokers and non-passive smokers.

a, Correlation between passive smoking score (1: life-time passive smoking to 4: period passive smoking) and number of mutations in EAGLE samples. **b**, Average mutational spectra comparison of SBS between patients with passive smoking score 1 and 4. **c**, Proportions of SBS, DBS and ID mutational signatures identified in samples from passive smokers (Y, n=64), non-passive smokers (N, n=149) and passive smoking status unknown (Unknown, n=19).

Supplementary Fig. 15



Supplementary Fig. 15 Dynamic mutational process during clonal and subclonal tumor evolution.

a, Fold changes between subclonal and clonal periods are shown in samples with measurable changes in mutation spectra based on mutational signatures. **b**, Reconstruction of the trajectory of signatures activity. Each horizontal line is a trajectory of changes of one SBS signature activity relative to its activity in the most recent common ancestor (MRCA) (y-axis) as a function of decreasing cancer cell fraction (x-axis). The width of the lines indicates the number of samples of estimated activity changes. The vertical red line depicts the time point of the MRCA. **c**, Comparison of the APOBEC mutation loads identified by P-MACD (APOBEC_mutLoad_MinEstimate) and SigProfilerExtractor (SBS2 and SBS13). The pie chart with legend indicates the fold-enrichment and significance of APOBEC mutagenesis signatures over the expected occurrence for random mutagenesis estimated by P-MACD. Numbers in brackets show the number of samples in each subset on the scatter plot.

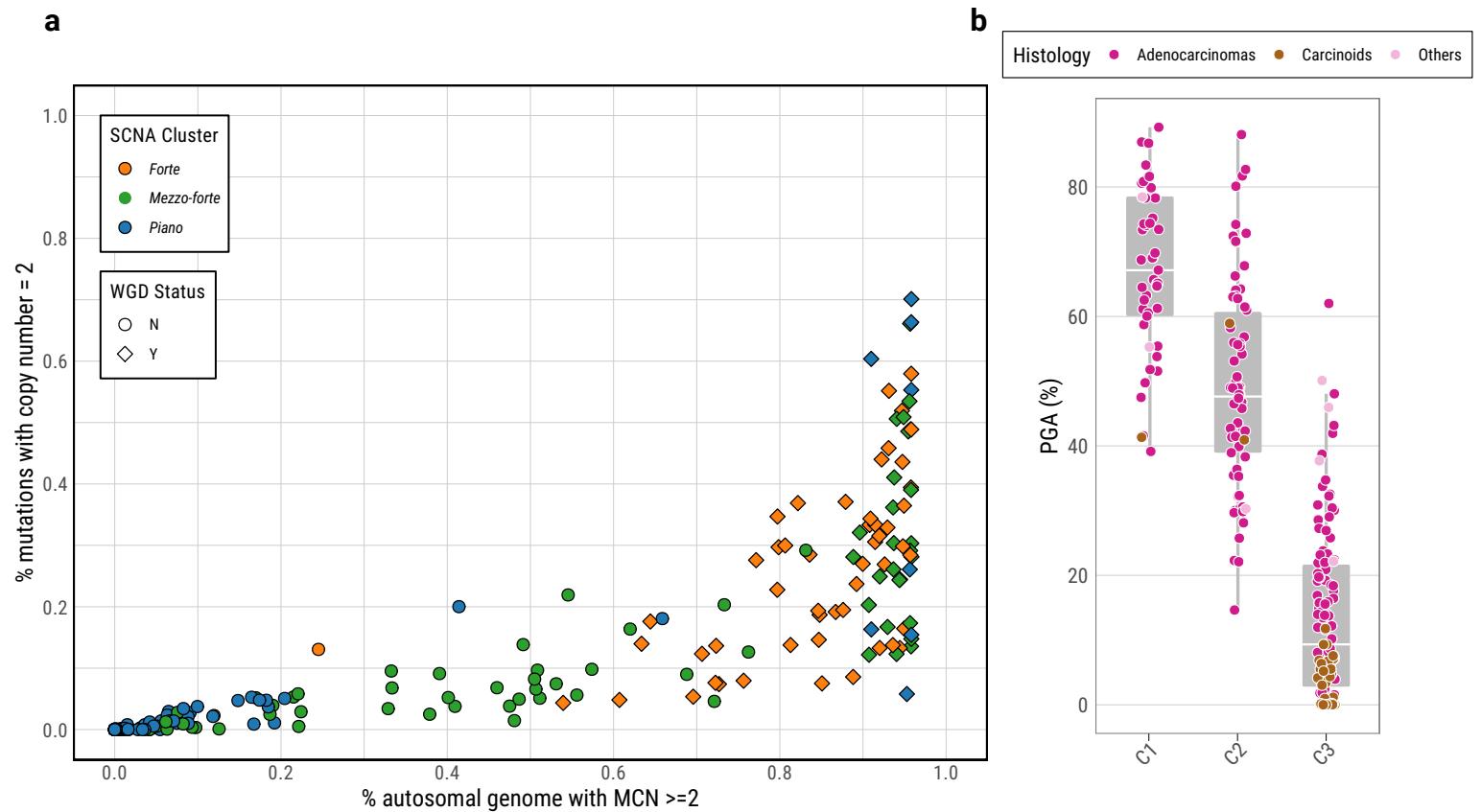
Supplementary Fig. 16



Supplementary Fig. 16 SBS mutational signature profile using combined endogenous and environmental signatures.

Panels from top to bottom: 1) Unsupervised clustering based on the proportion of SBS signatures decomposed by SigProfiler; 2) Histology; 3) Tumor purity; 4) Total number of SBS assigned to each signature; 5) Cosine similarity between the original mutational profile and signature decomposition results; 6) Proportions of SBS assigned to each mutational signatures.

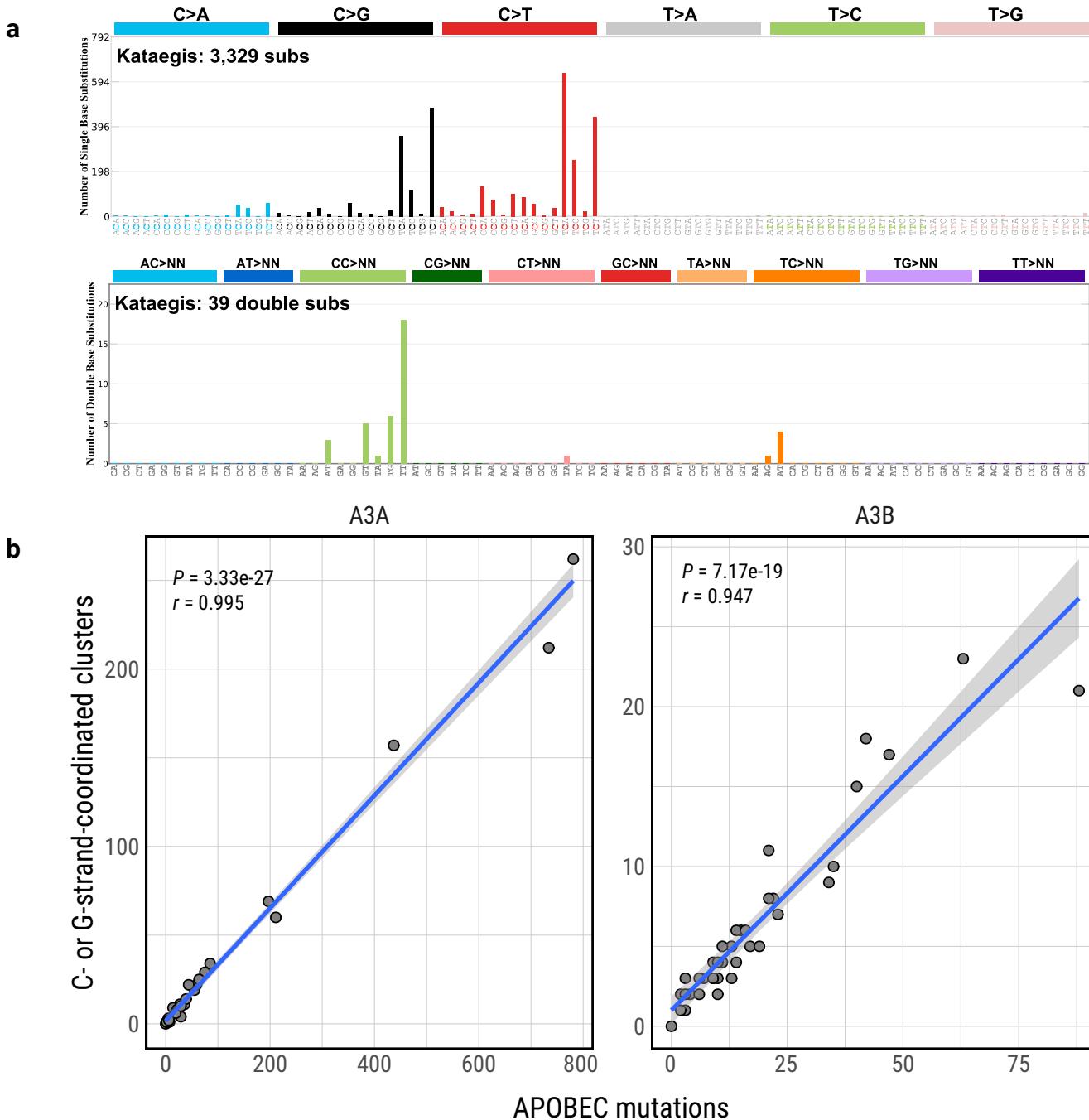
Supplementary Fig. 17



Supplementary Fig. 17 Genomic instability based on SCNAs.

a, Prevalence of whole genome doubling (WGD) in *Sherlock-Lung*. The x-axis shows the fraction of the autosomal genome with a major copy number (MCN) of two or greater. The y-axis shows the percentage of mutations with copy number=2. Each dot represents an individual (color-coded for the SCNA clusters; WGD status shown in shapes). **b**, Boxplot showing the percentage of genomic alterations (PGA) across different SCNA clusters.

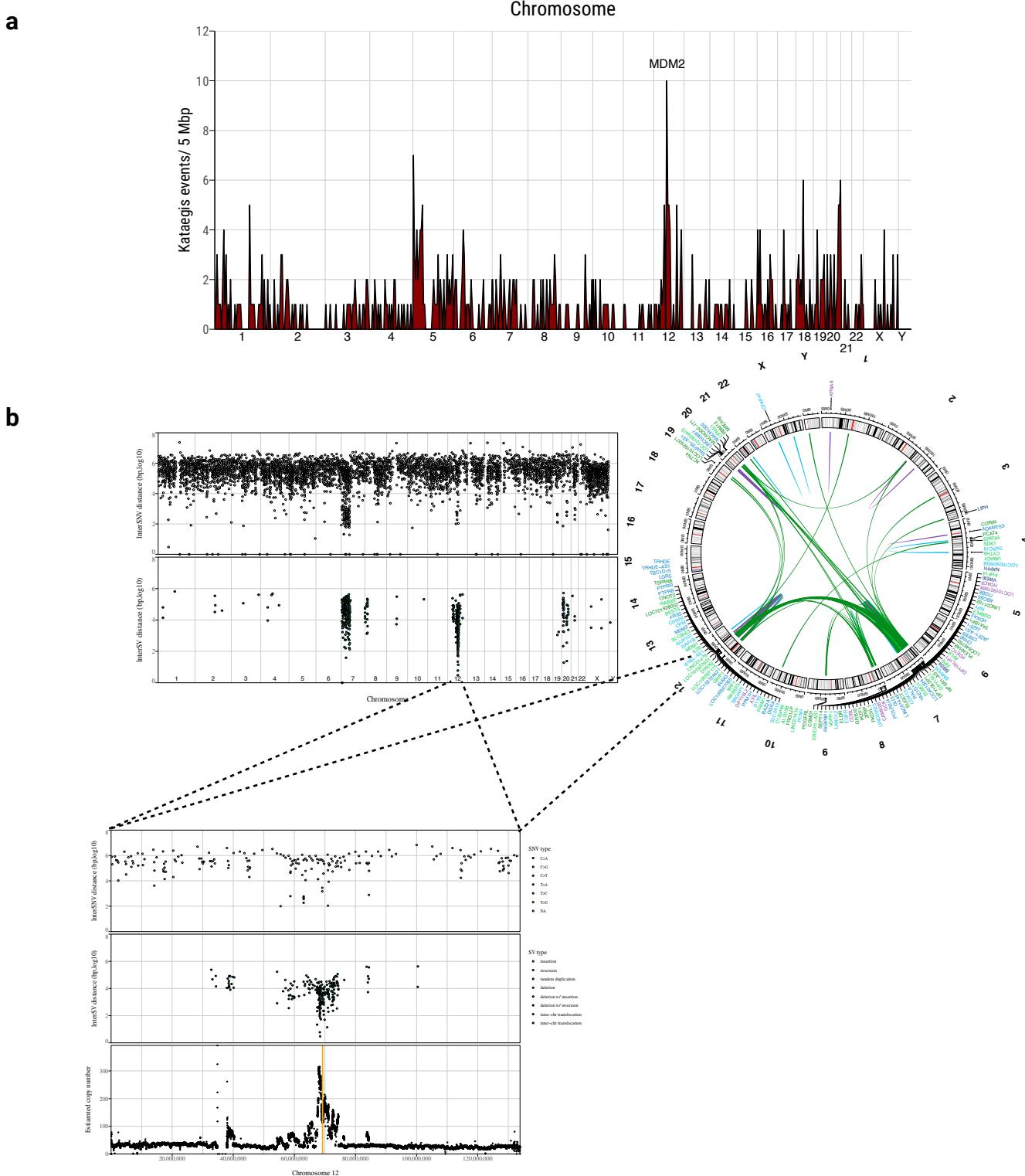
Supplementary Fig. 18



Supplementary Fig. 18 APOBEC emerged as the dominant mutational process in kataegis loci.

a, Spectra of single base substitutions and double base substitutions in samples within kataegis. **b**, Association between APOBEC mutation loads and the number of C- or G-strand-coordinated clusters related to kataegis in both A3A-like tumors and A3B-like tumors.

Supplementary Fig. 19

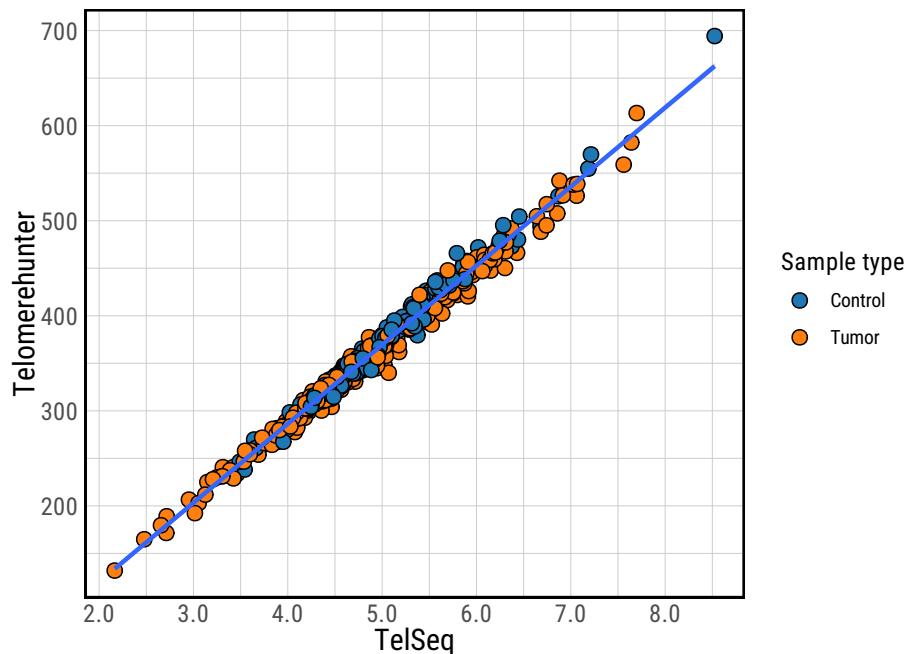


Supplementary Fig. 19 Colocalization between kataegis loci and *MDM2* amplification due to chromothripsis processes.

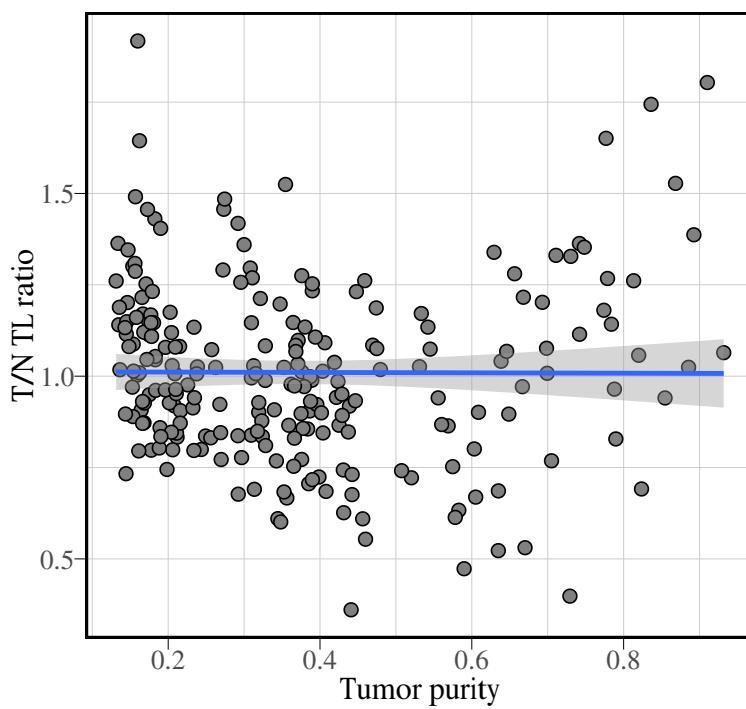
a, Frequency of kataegis events using a 5 Mbp window across the whole genome. **b**, Representative kataegis events in sample NSCL-0012-T01. Left top panels show rainfall plots of SNVs and SVs. Each dot represents an SNV or SV. SNVs and SVs are ordered by the genomic locations on the x-axis. The distances between two consecutive SNVs and SVs are plotted on the y-axis. The top right panels show a circos plot of SV rearrangements. The bottom panels show an enlargement of rainfall plots on chromosome 12 with the corresponding estimated copy numbers. The genomic location of *MDM2* is indicated by the orange line.

Supplementary Fig. 20

a



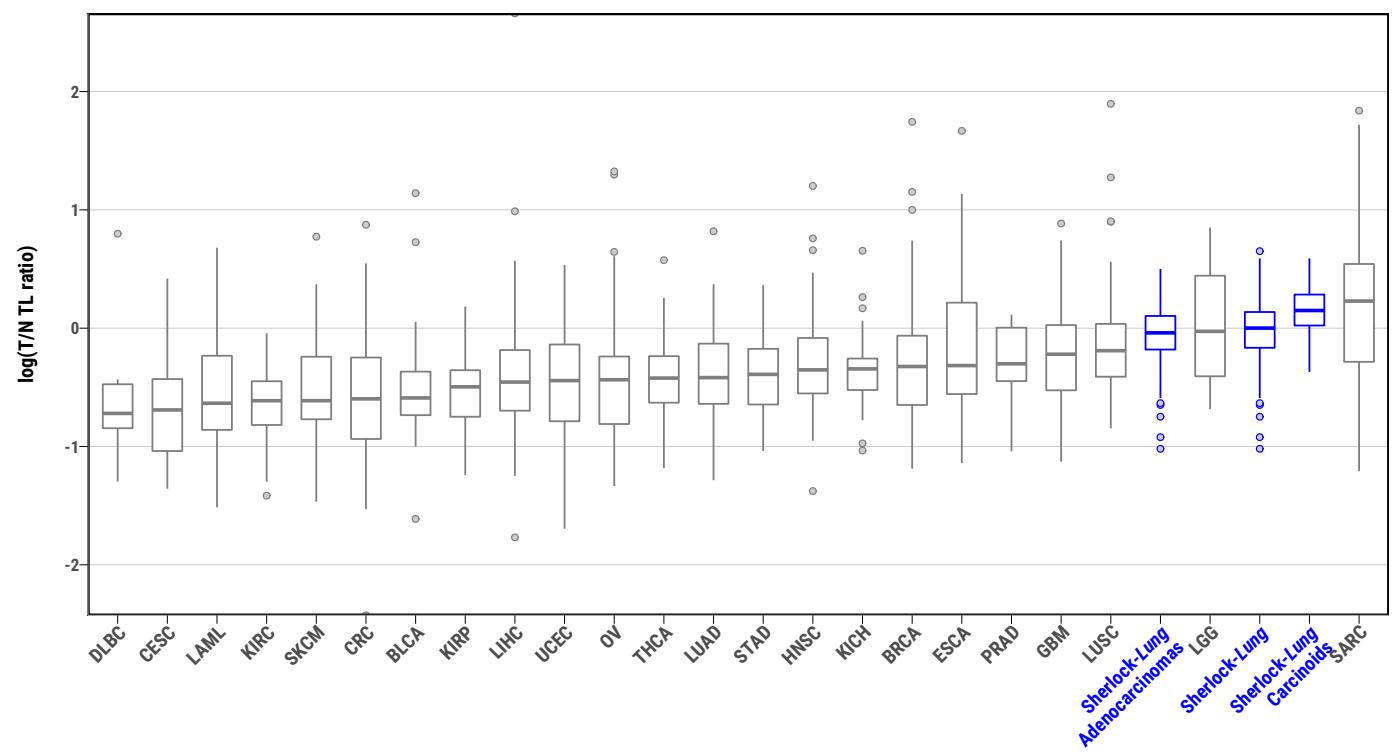
b



Supplementary Fig. 20 Telomere length estimation and association with tumor purity.

a, Correlation of telomere length estimations between TelSeq and TelomereHunter by sample type. **b**, Correlation between tumor purity and Tumor/Normal telomere length (T/N TL) ratio.

Supplementary Fig. 21

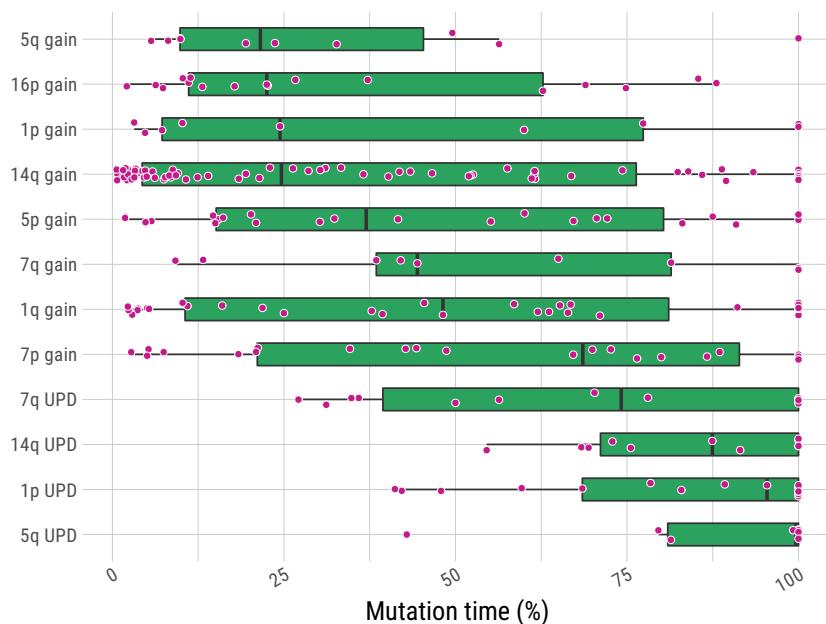


Supplementary Fig. 21 The comparison of T/N TL ratio between Sherlock-Lung and other cancer types in the TCGA study.

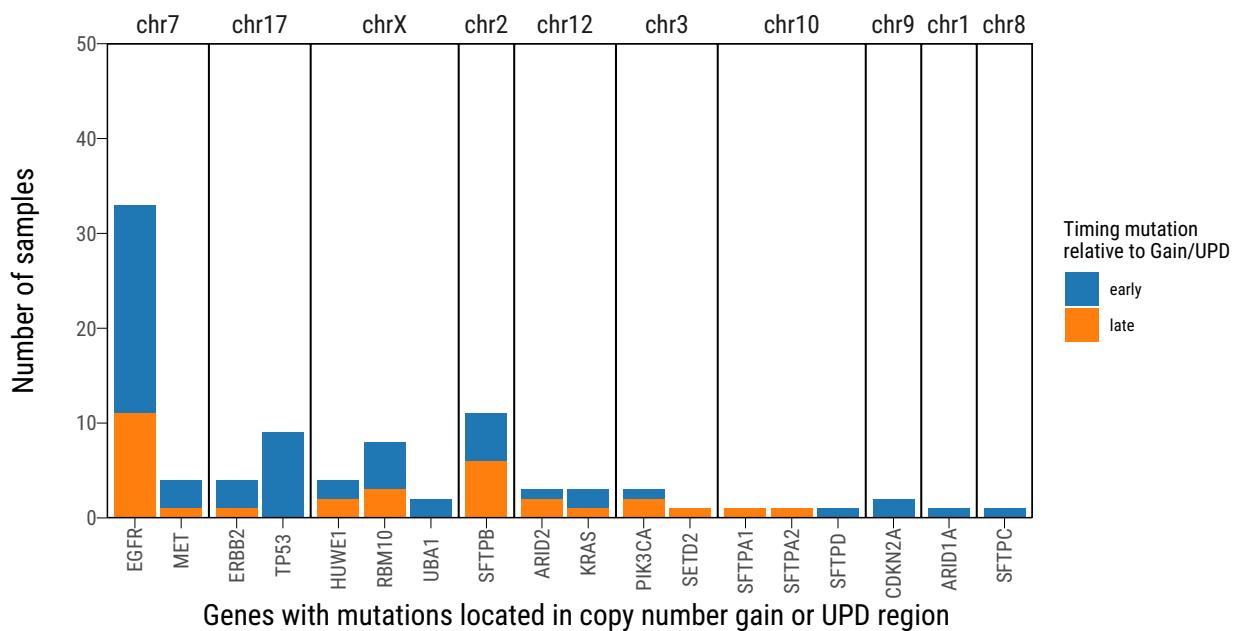
Boxplot of the T/N TL ratio in *Sherlock-lung* (blue) and TCGA Cancer Types (gray). The name abbreviation of the TCGA cancer types can be found in Supplementary Methods.

Supplementary Fig. 22

a



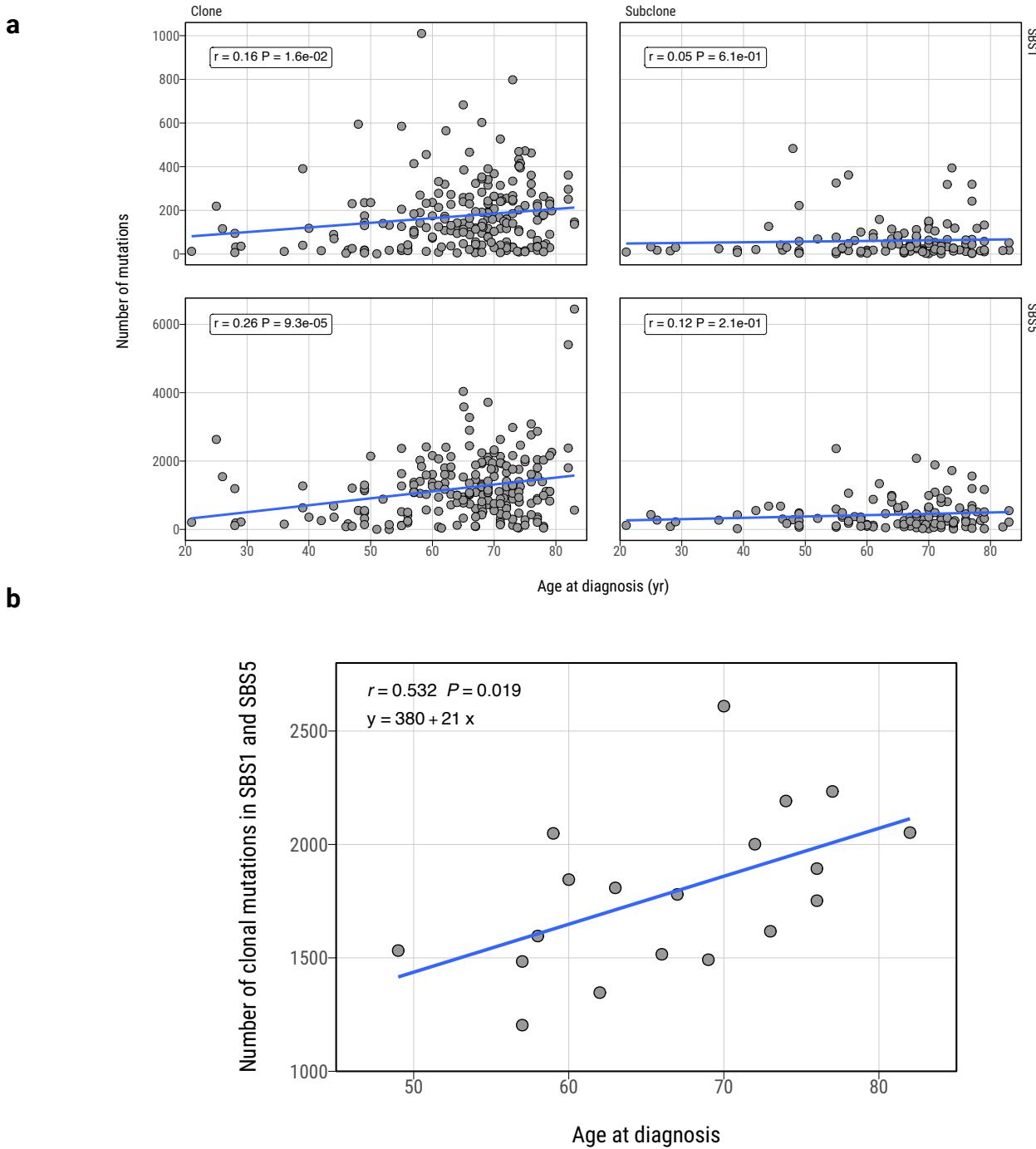
b



Supplementary Fig. 22 Estimation of the timing of duplications and driver mutations

a, Molecular timing estimation of the arm-level chromosome duplications. Each dot represents a tumor sample harboring arm-level chromosome duplication. The molecular time is estimated based on co-amplified point mutations in individual samples. UPD: uniparental disomy (UPD=CN-LOH). **b**, Timing mutations in driver genes relative to copy number gain or uniparental disomy (UPD) events of the corresponding chromosomal regions. The bar plots indicate the number of samples with mutations in driver genes occurring earlier (blue) or later (orange) than the SCNA events at the corresponding chromosomal regions.

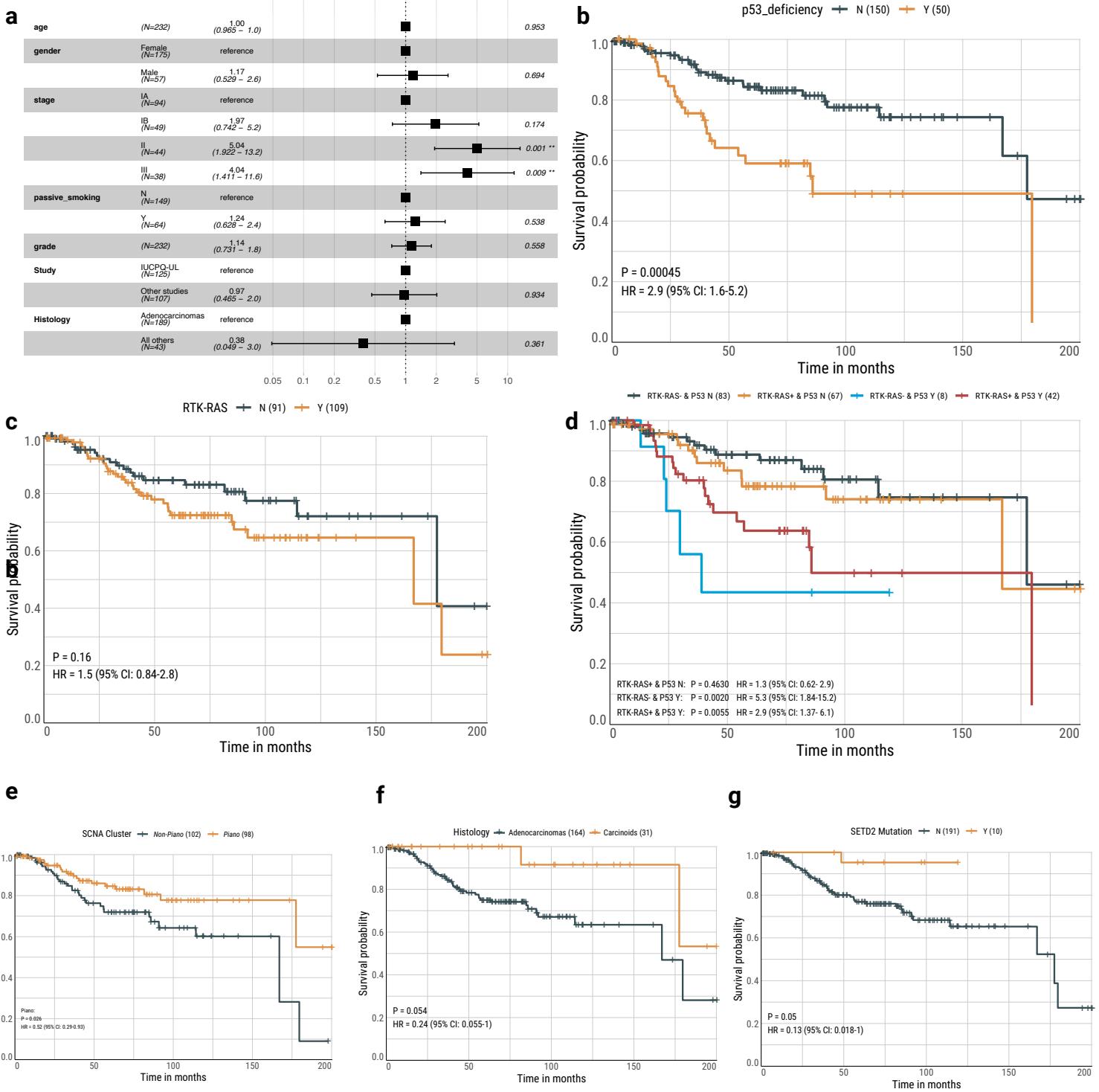
Supplementary Fig. 23



Supplementary Fig. 23 Association between clock-like clonal mutations and age at diagnosis

a, Scatter plots showing the correlations between age at diagnosis and number of clonal mutations assigned to SBS1 and SBS5 signatures. **b**, Linear regression between age at diagnosis and the number of clonal mutations in SBS1 and SBS5 in 19 diploid tumors (all LUAD) with observed adequate clonal SBS1 and SBS5 contributions.

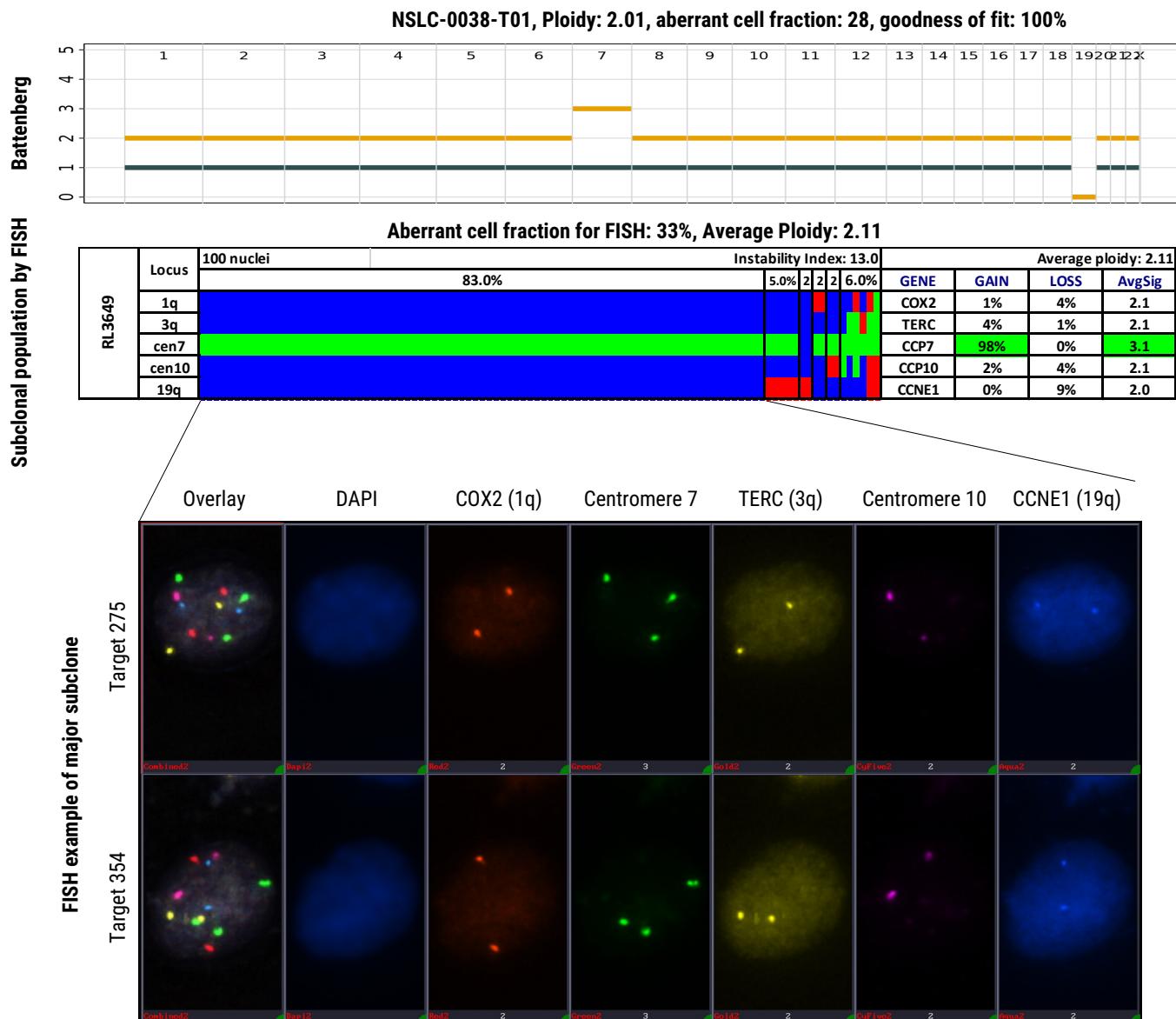
Supplementary Fig. 24 Hazard ratio



Supplementary Fig. 24 Additional associations between genomic aberrations and clinical outcomes.

a, Forest plot showing hazard ratios of multivariate analysis of clinical and pathological features for overall survival probability, including age at diagnosis, gender, tumor stage, tumor grade, passive smoking status, study sites, and tumor histological types. Kaplan-Meier survival analyses for overall survival probability stratified by overall TP53 deficiency status (**b**), overall RTK-RAS pathway activation status (**c**), combining TP53 deficiency and RTK-RAS pathway activation status (**d**), piano tumor (**e**), tumor histology type (**f**), and SETD2 mutation status (**g**). P-values for significance and hazard ratios (HR) of difference are calculated using the log-rank test with adjustment for age, gender and tumor stage.

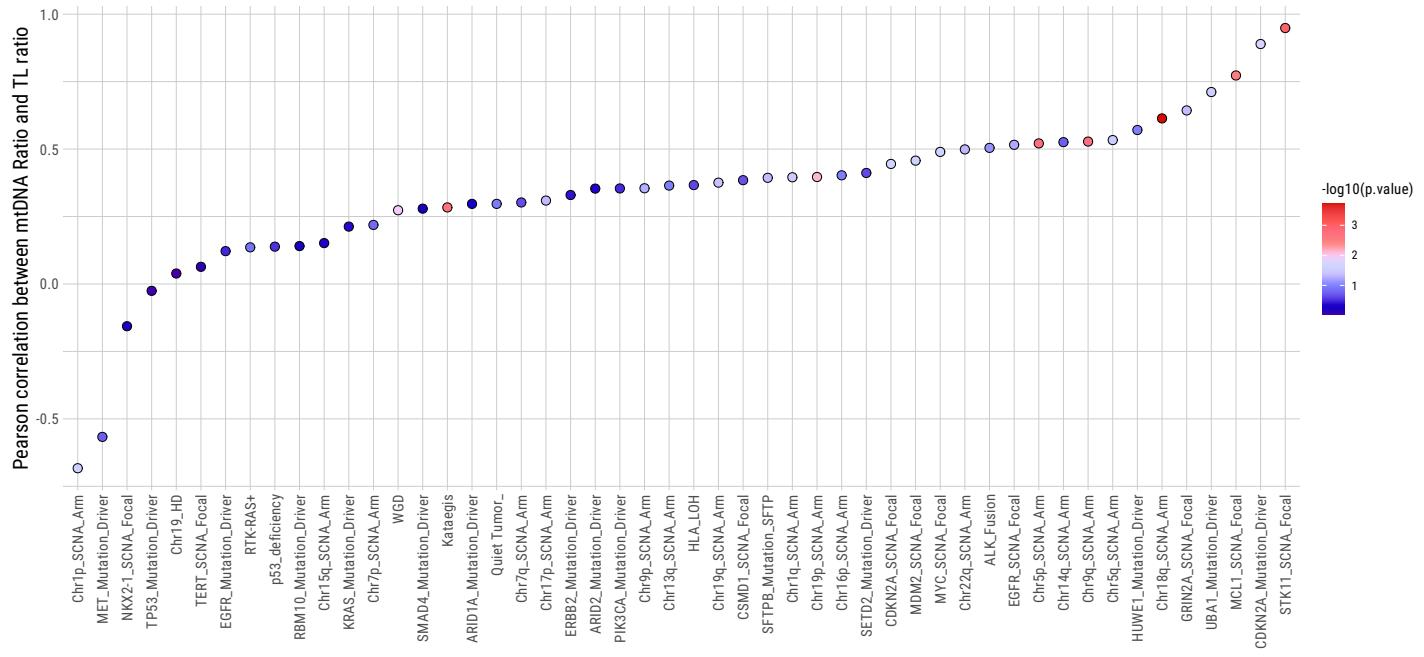
Supplementary Fig. 25



Supplementary Fig. 25 Example of Chr.19 homozygous deletion validation originally identified by Battenberg algorithm.

Multicolor Fluorescence *in situ* hybridization (mFISH) of samples with apparent Chr.19 homozygous deletion (HD) from Battenberg analysis. One representative tumor (NSLC-0038-T01) where Battenberg showed a gain of Chr.7 and HD loss of Chr.19 (top panel) but the mFISH only confirmed the gain of Chr.7 (lower panels). 100 aberrant nuclei were counted. In the mFISH color chart, blue=copy number neutral, green=copy number gain, red=copy number loss. The mFISH panel consisting of three gene-specific probes: COX2 (1q33.2; red), TERC (3q26.2; gold), and CCNE1 (19q12;aqua), and two centromere probes: CEP7 (green) and CEP10 (far red), revealed a gain in Chr.7 in ~98% of aberrant nuclei and no HD of Chr.19.

Supplementary Fig. 26



Supplementary Fig. 26 Pearson correlation between T/N mtDNA ratio and T/N telomere length (TL)TL ratio in a subset of tumors with the same genomic alterations (N>5).