

FIGURE S1

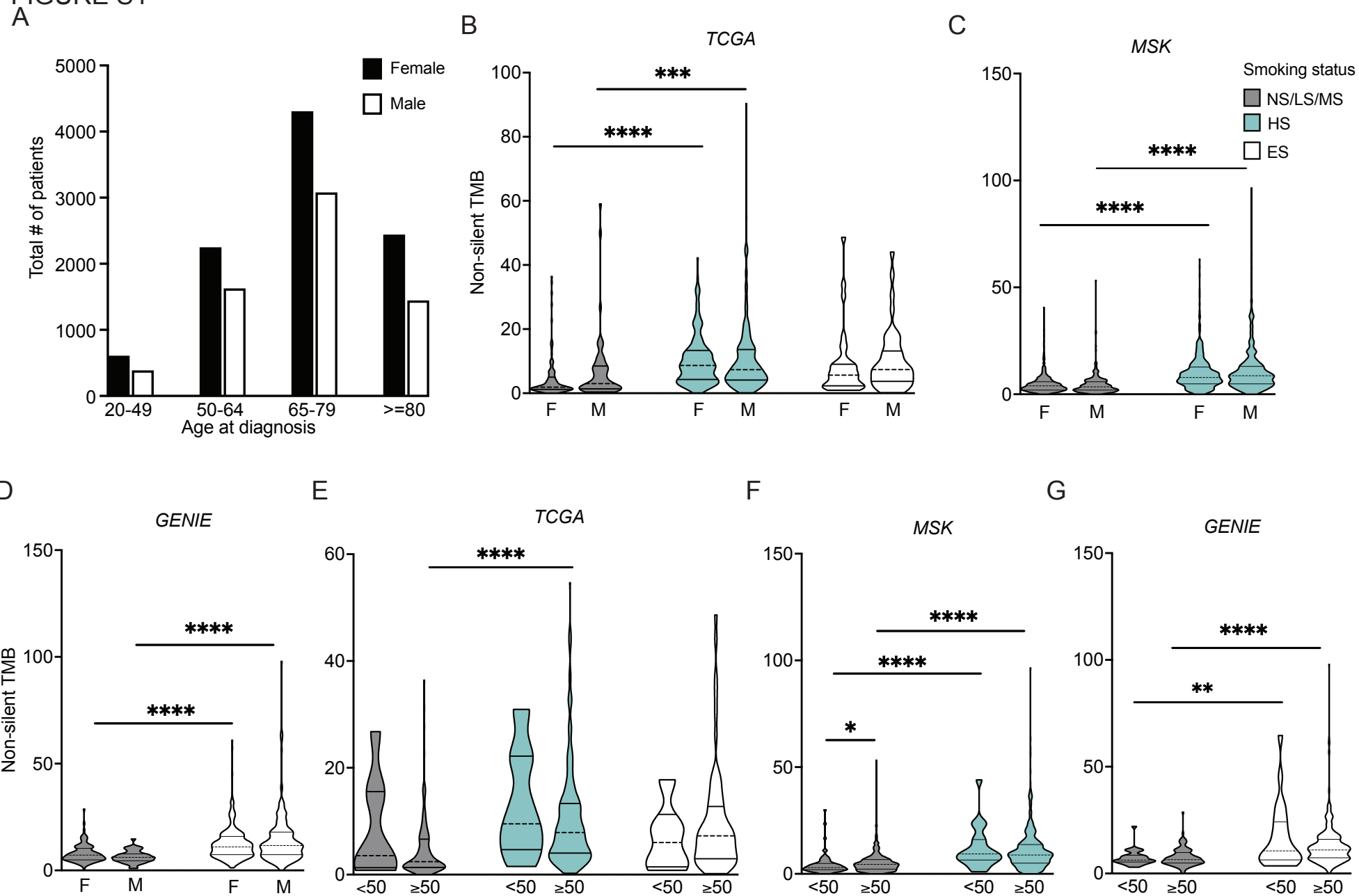


Figure S1. (A) Number of female and male never-smokers diagnosed with lung cancer stratified by age at diagnosis from Siegel *et al.*, 2021. **(B-D)** Non-silent TMB in males and females from the TCGA (B), MSK (C) and GENIE (D) cohorts. **(E-G)** Non-silent TMB in by age at diagnosis less than 50 years versus those diagnosed at 50 or more years from the TCGA (E), MSK (F) and GENIE (G) cohorts respectively. Kruskal-Wallis/Dunns test was used for comparison of TMB between smoking groups. p-values: * $p < 0.05$ ** $p < 0.01$. NS (Never-smoker), LS (Light smoker), MS (Moderate smoker), HS (Heavy smoker), ES (Ever-smoker), F (female), M (male).

FIGURE S2

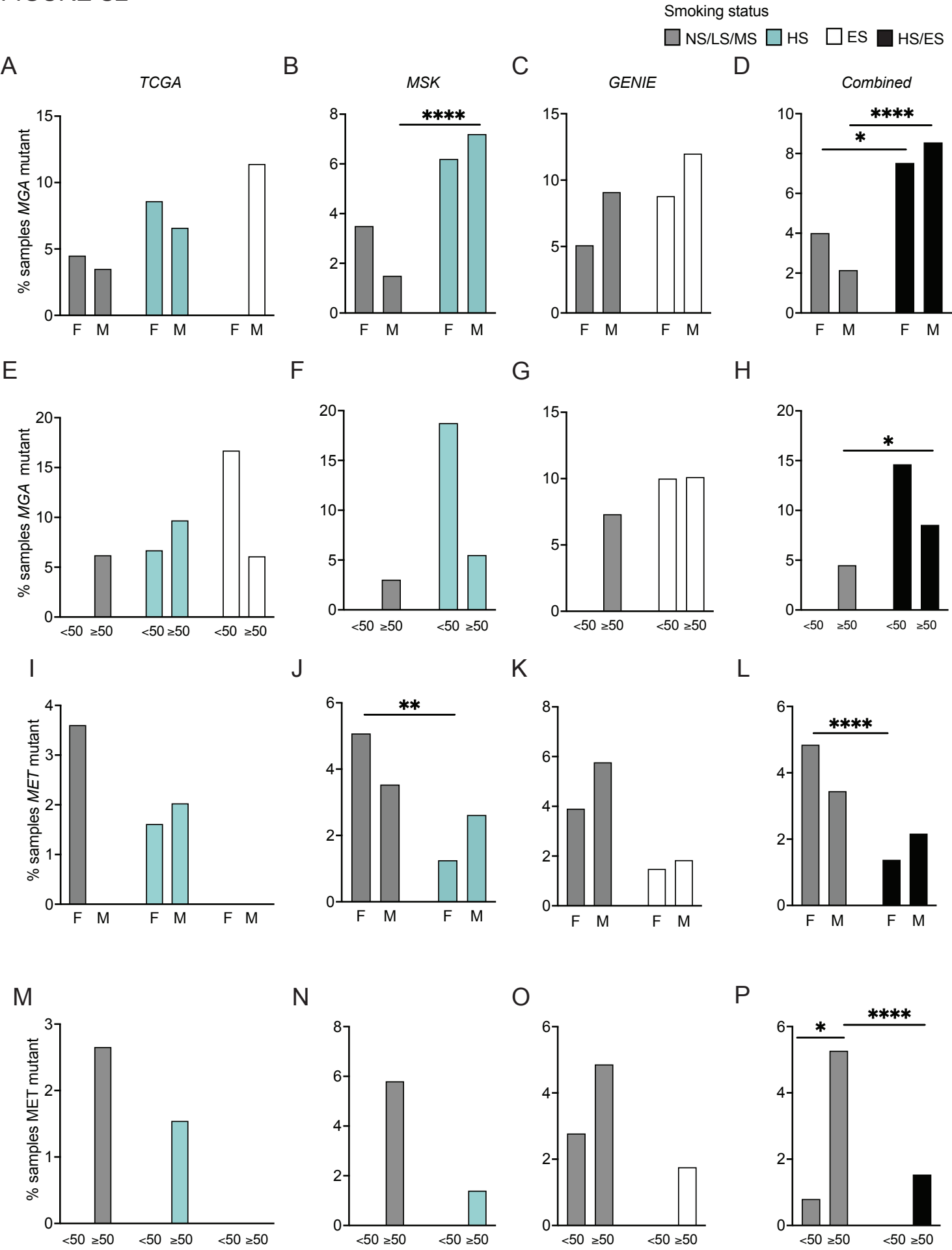
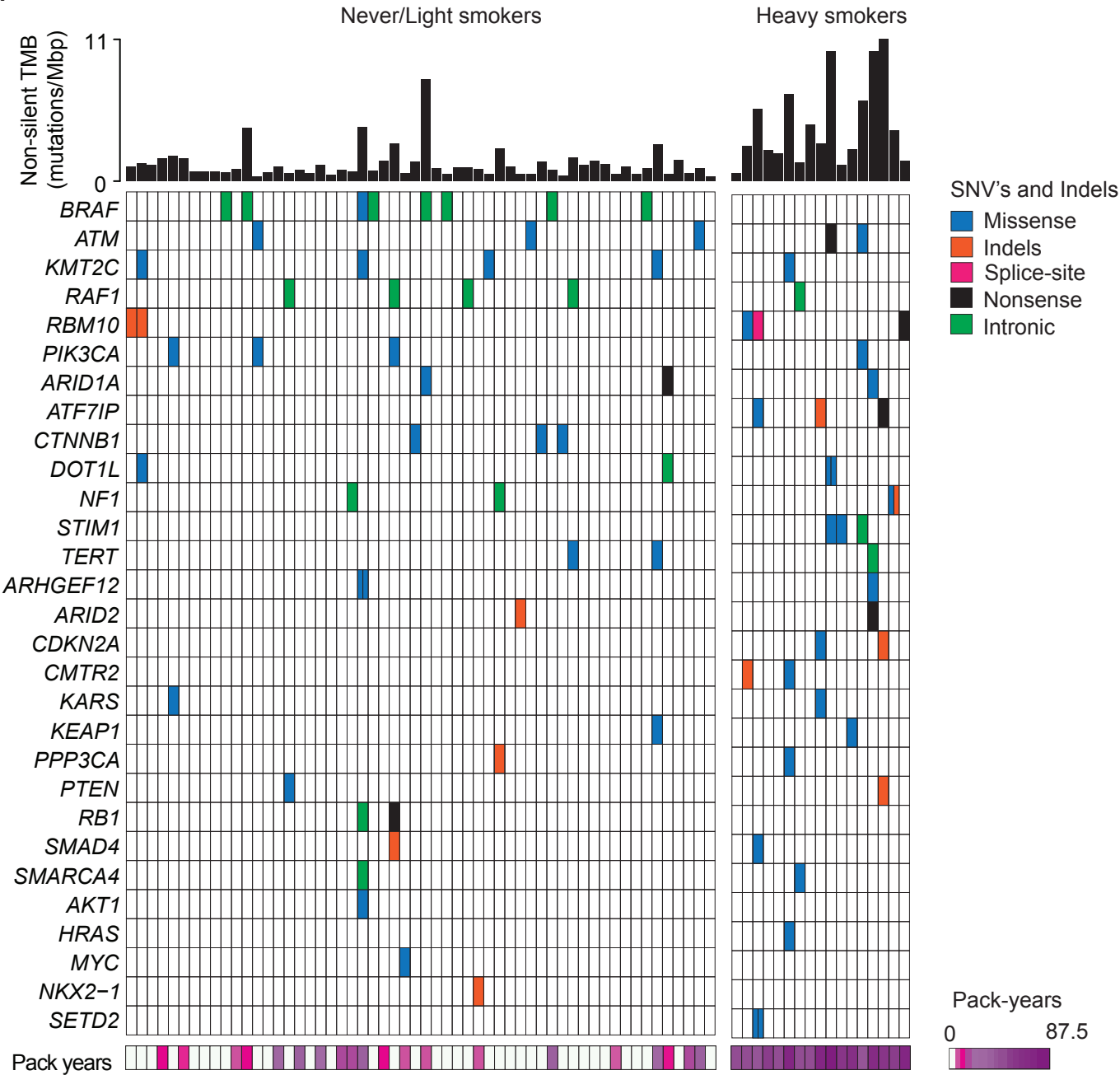


Figure S2. (A-D) Percent samples with *MGA* mutations split by smoking status and sex in the TCGA (A), MSK (B), GENIE (C) cohorts and the three cohorts combined (D). Fisher's exact test was used to compare the enrichment of *MGA* mutations between smoking groups. **(E-H)** Percent samples with *MGA* mutations split by age at diagnosis in the TCGA (E), MSK (F), GENIE (G) cohorts and the three cohorts combined (H). Fisher's exact test was used to compare the enrichment of *MGA* mutations between the age of diagnosis groups. **(I-L)** Percent samples with *MET* mutations split by smoking status in the TCGA (I), MSK (J), GENIE (K) cohorts and the three cohorts combined (L). Fisher's exact test was used to compare the enrichment of *MET* mutations between smoking groups. **(M-P)** Percent samples with *MET* mutations split by age at diagnosis in the TCGA (M), MSK (N), GENIE (O) cohorts and the three cohorts combined (P). Fisher's exact test was used to compare the enrichment of *MET* mutations between the age of diagnosis groups. Fisher's exact test p-values: * $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$. NS (Never-smoker), LS (Light smoker), MS (Moderate smoker), HS (Heavy smoker), ES (Ever-smoker).

FIGURE S3
A



B

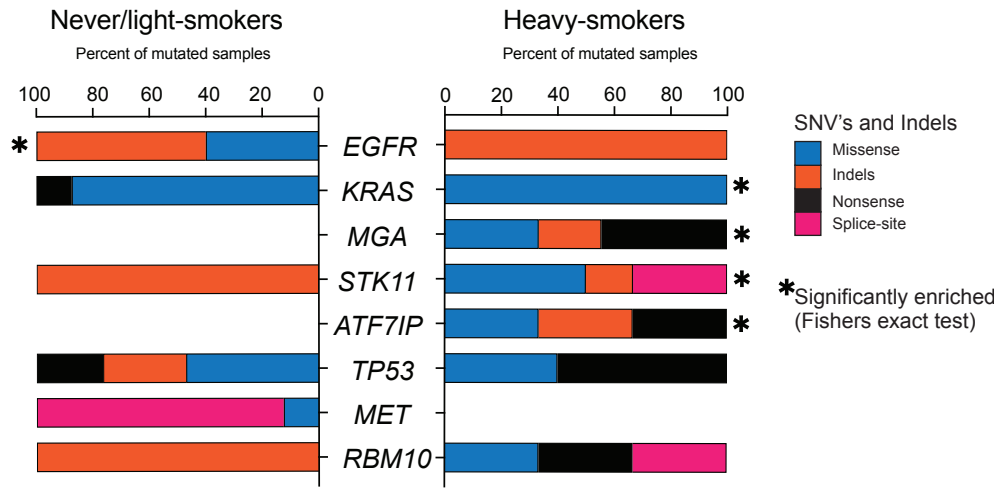
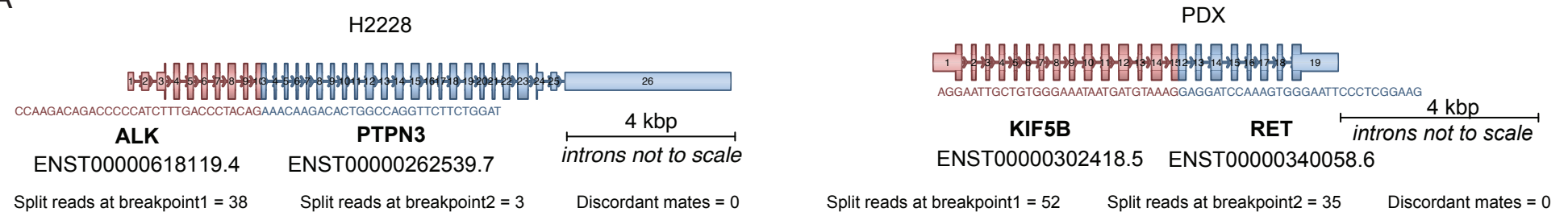


Figure S3. (A) Oncoplot of mutations belonging to the RTK/Ras/Raf pathway or genes known to be altered in lung adenocarcinoma from the WHI cohort. The top bar plot shows the non-silent TMB rate for each patient. **(B)** Percentage of types of mutations for key genes. Genes that were significantly enriched in one or the other smoking groups were marked by a star. Fisher's exact test was used to measure enrichment; the star represents enrichment in one smoking cohort over another ($p < 0.05$).

FIGURE S4

A



B

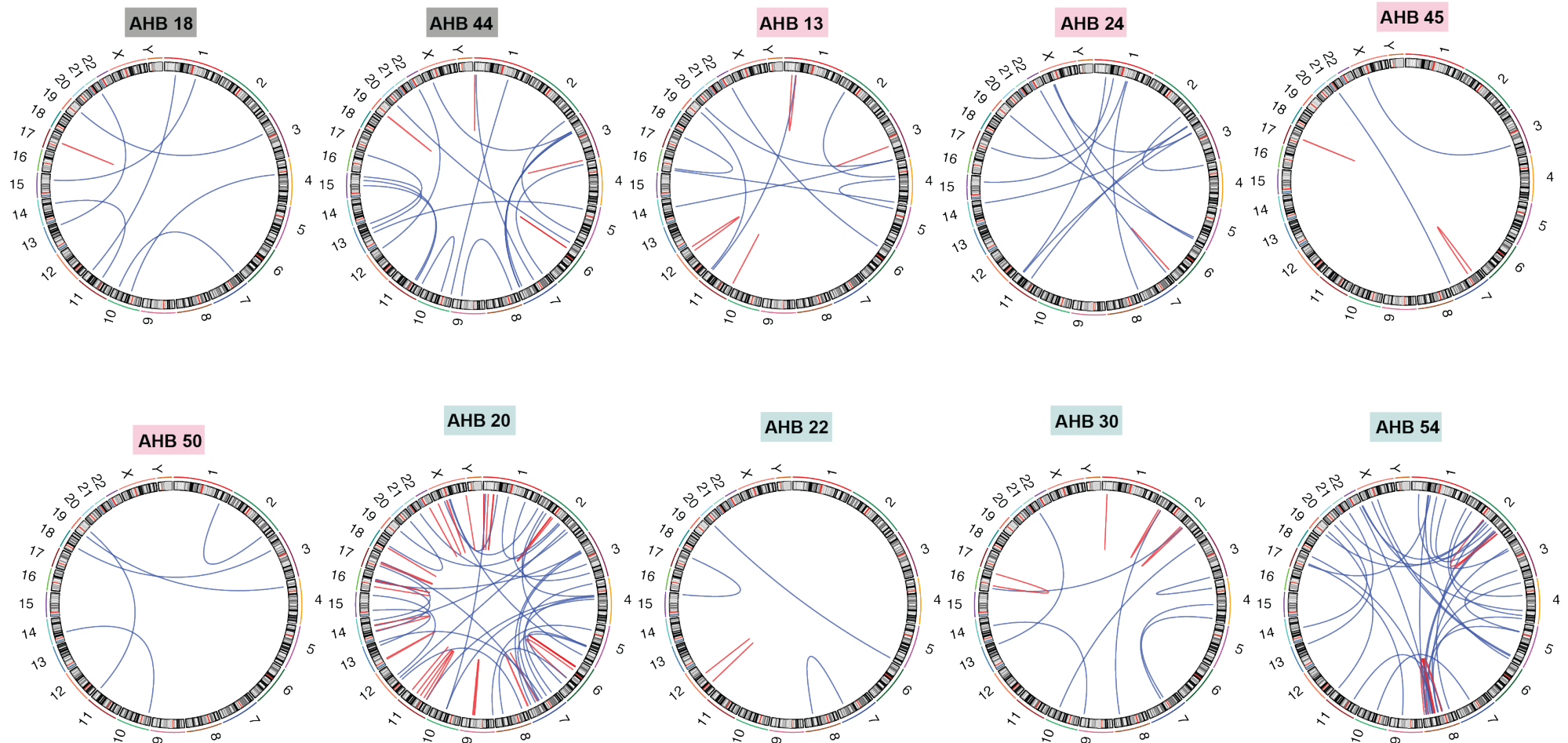
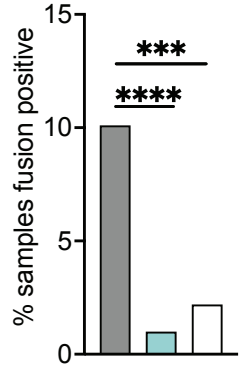


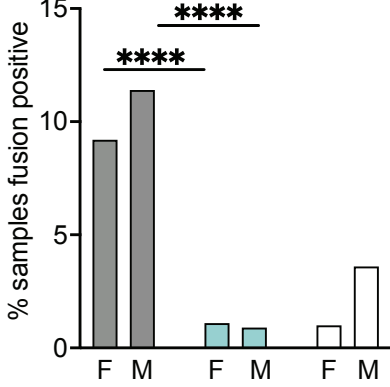
Figure S4. (A) Schematic representation of translocations found in control cell line *H2228* and lung adenocarcinoma metastatic PDX (AHB001) sample. Schematic shows the location of regions involved in the translocation on the origin chromosomes, coverage of different regions of the genes involved, and the final structure of the translocation. **(B)** Circos plots showing inter-chromosomal fusions (blue) and intra-chromosomal fusion (red) for 10 oncogene-negative samples that underwent whole-genome sequencing. SvABA was used to determine translocations spanning lengths >10kb and where partners involved in the translocation were in protein-coding regions. Sample IDs are colored to represent the smoking history of the patients never-smokers (gray), light smokers (pink) and heavy smokers (green).

FIGURE S5

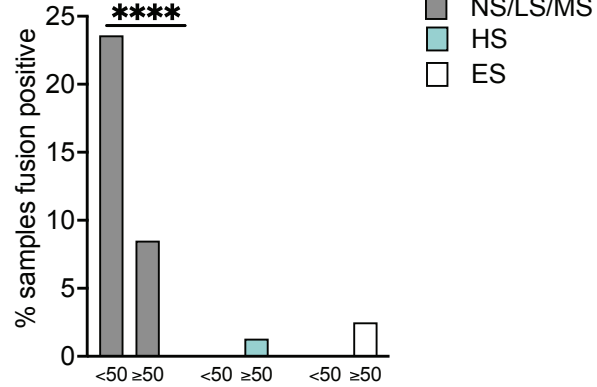
A



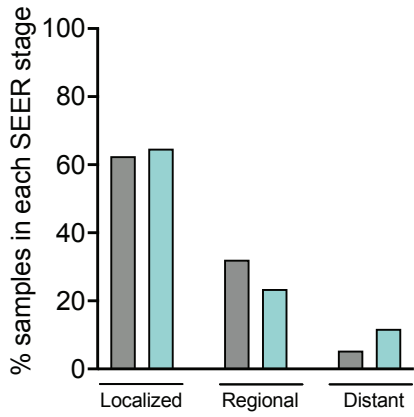
B



C



D



E

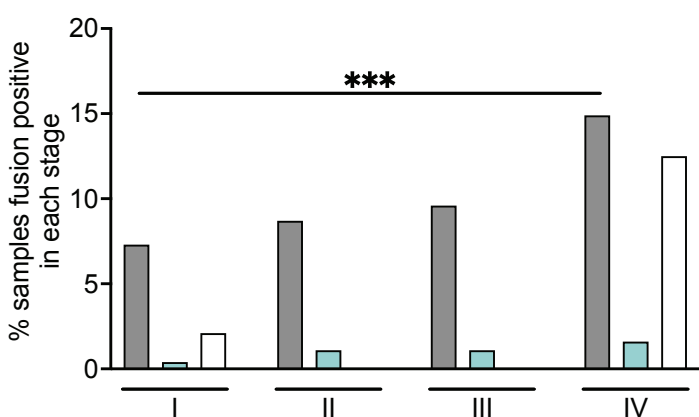


Figure S5. (A) Percent samples that are positive for *ALK*, *RET* and *ROS1* fusions from the analysis of external cohorts TCGA, MSK, OncoSG, Sherlock and Broad, by smoking status **(B)** Percent samples fusion-positive by sex and smoking status in the external cohorts. **(C)** Percent samples fusion-positive by age at diagnosis and smoking status in the external cohorts. **(D)** Percent samples in the WHI cohort that were diagnosed with tumors in the different SEER stage groups. **(E)** Percent samples fusion-positive by stage and smoking status in the external cohort. Fisher's exact test was used to calculate the enrichment between groups. Fisher's exact test p-values: *** $p < 0.001$, **** $p < 0.0001$. NS (Never-smoker), LS (Light smoker), MS (Moderate smoker), HS (Heavy smoker), ES (Ever-smoker).

FIGURE S6

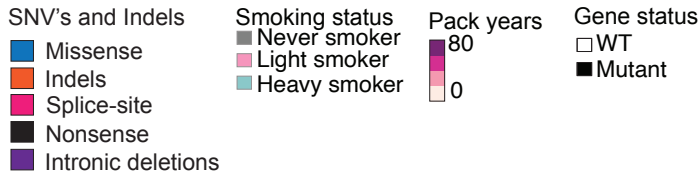
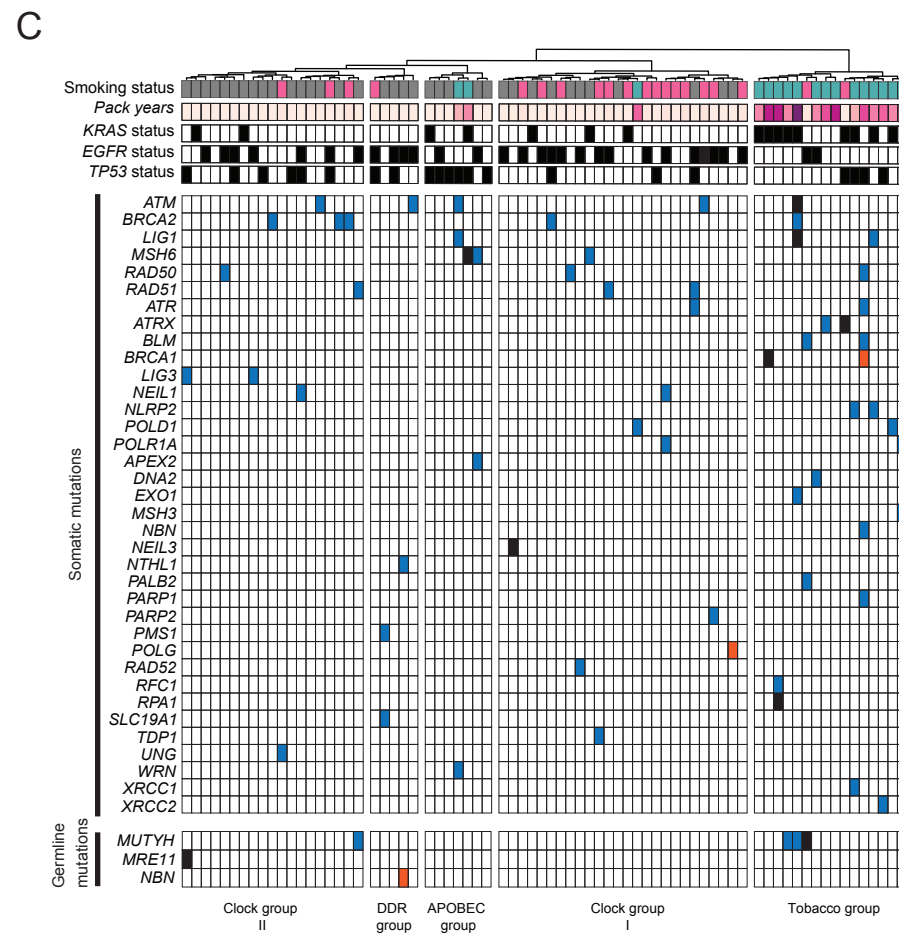
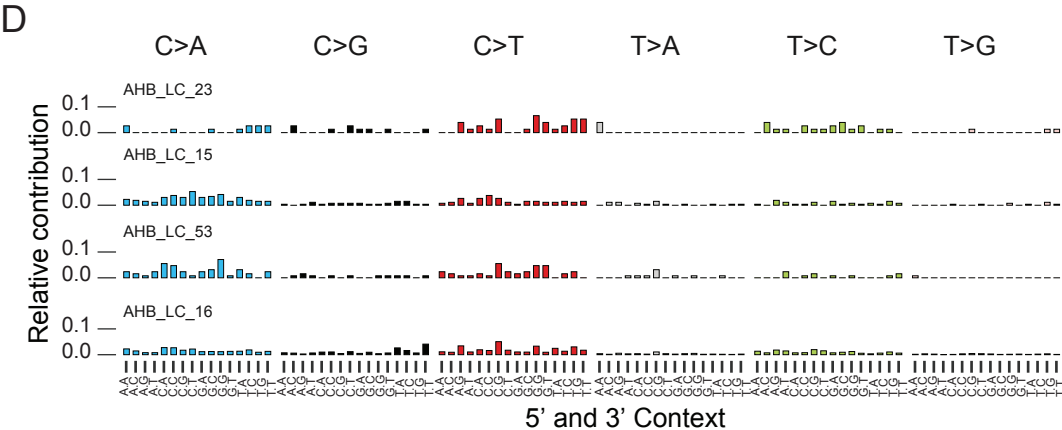
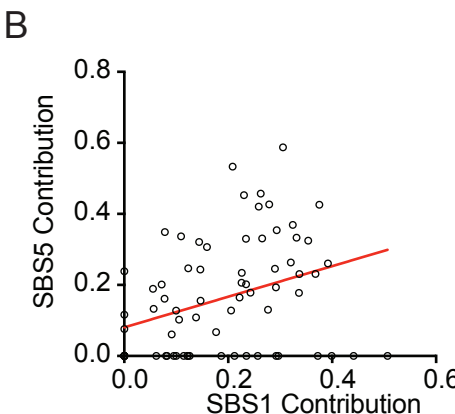
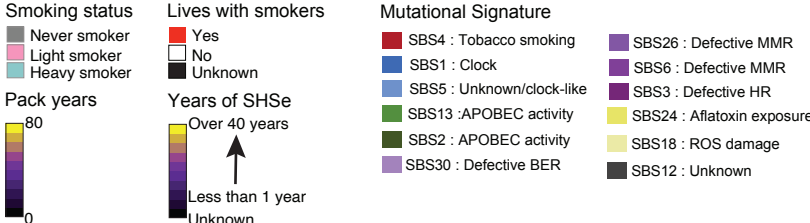
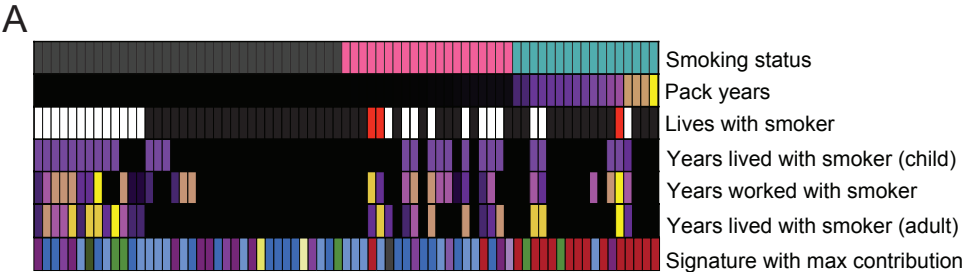


Figure S6. (A) Heatmap showing primary smoke exposure and second-hand smoke exposure (SHSe) for each patient. Samples are sorted based on increasing pack years. Also shown is the signature with the maximum contribution to the mutational signature burden. **(B)** Scatter plot for SBS1 versus SBS5 contributions per sample. Simple linear regression shows a significant association between SBS1 and SBS5 contributions in the WHI cohort; $p = 0.0026$. **(C)** Oncoplot for somatic and germline mutations in DNA damage and repair (DDR) genes. **(D)** Relative mutational profile in all possible 5' and 3' contexts in *MUTYH* mutated samples. Y-axis shows the relative contributions of each mutational context.

FIGURE S7

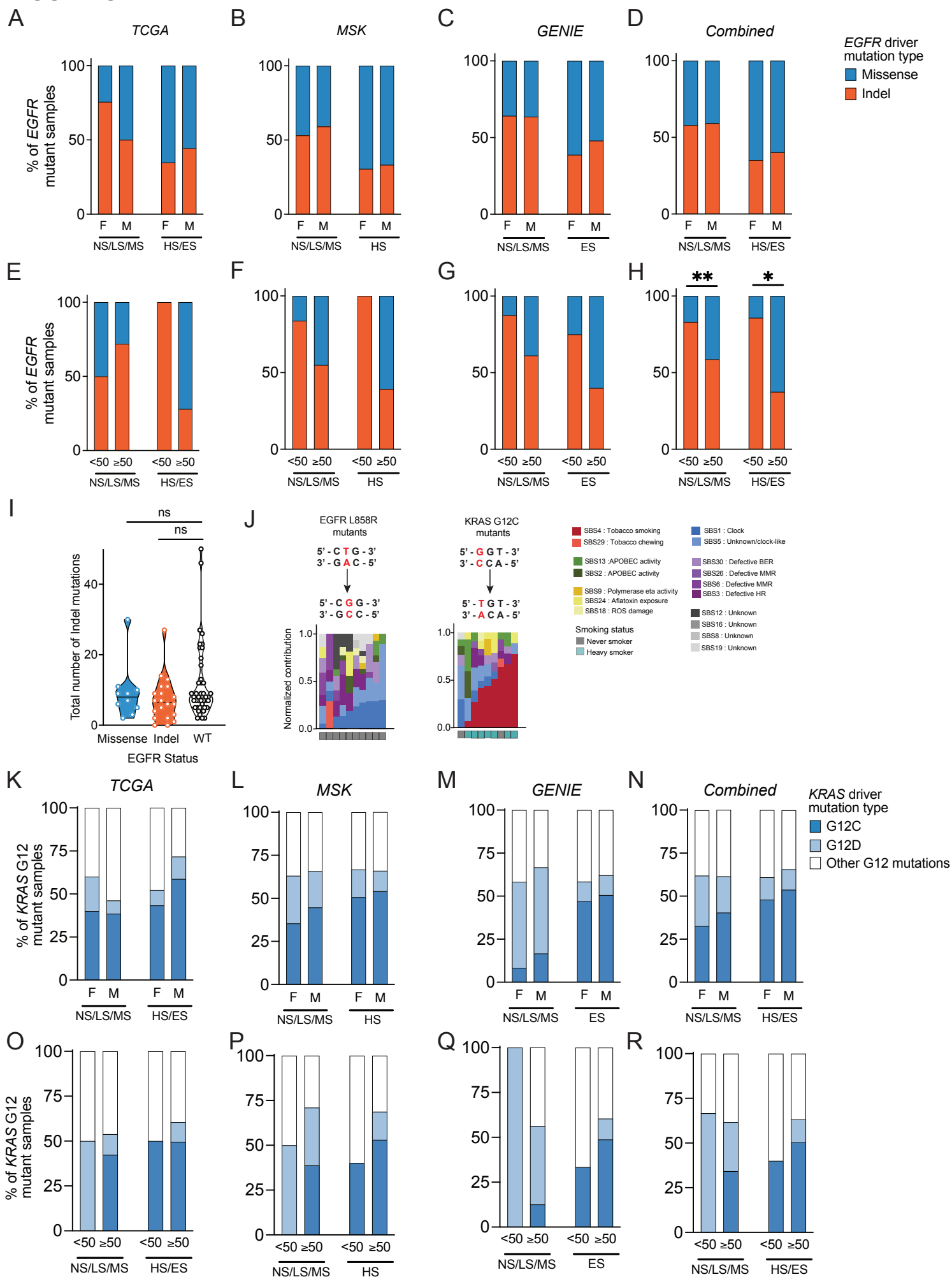
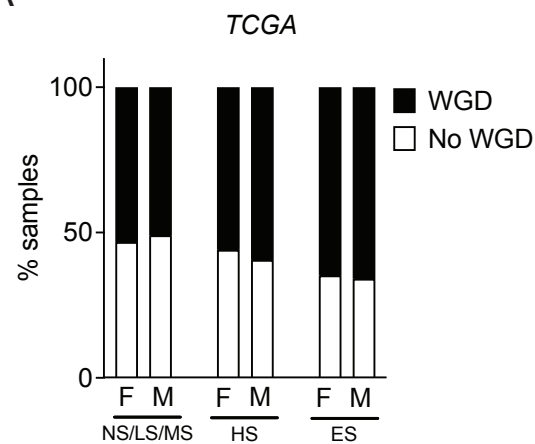


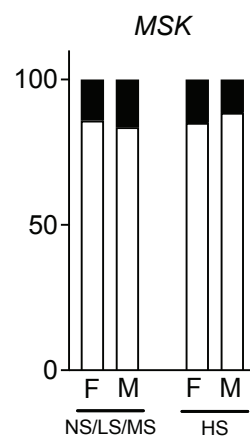
Figure S7. (A-D) The percent of samples with either *EGFR* indel (orange) or missense (blue) mutations split by sex and smoking status for MSK (A), TCGA (B) and GENIE (C) cohorts and combined for all three cohorts (D). Two-tailed Fisher's test was used to measure enrichment of indel mutations **(E-H)** The percent of samples with either *EGFR* indel (orange) or missense (blue) mutations split by age at diagnosis and smoking status for MSK (E), TCGA (F) and GENIE (G) cohorts and combined for all three cohorts (H). Two-tailed Fisher's test was used to measure enrichment of indel mutations **(I)** Total number of indel mutations in the WHI cohort split by if they belonged to *EGFR* missense, indel or wild-type groups. Mann-Whitney test was used to compare the number of indel mutations to other groups. **(J)** Schematic of *EGFR* L858R mutation and *KRAS* G12C mutation at the nucleotide level (top). Mutational spectrum of *EGFR* and *KRAS* mutant samples in the WHI cohort with either L858R or G12C mutations only (bottom). **(K-N)** The percent of samples with either *KRAS* G12C mutations (dark blue) compared to G12D (light blue) and other G12 mutations (white) split by sex and smoking status for MSK (K), TCGA (L), GENIE (M) cohorts and all cohorts combine (N) Two-tailed Fisher's-exact test was used to compare number of *KRAS* G12C mutations in each group. **(O-R)** The percent of samples with either *KRAS* G12C mutations (dark blue) compared to G12D (light blue) and other G12 mutations(white) split by age at diagnosis and smoking status for MSK (O), TCGA (P), GENIE (Q) cohorts and all cohorts combine (R) Two-tailed Fisher's-exact test was used to compare number of *KRAS* G12C mutations in each group.

FIGURE S8

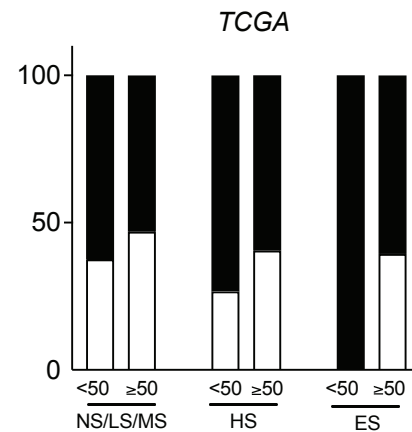
A



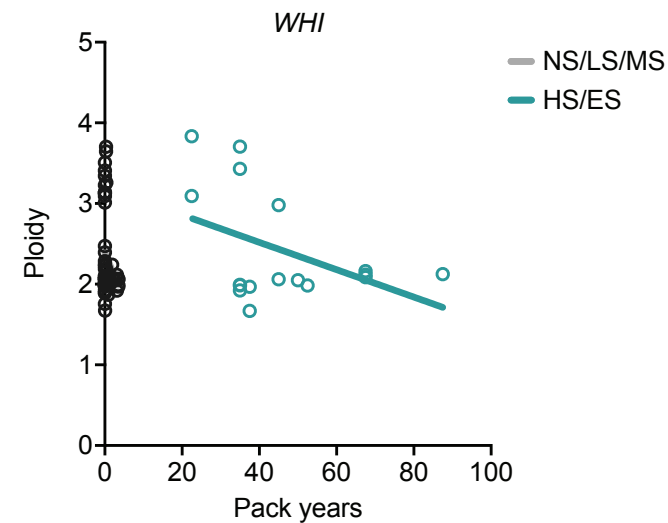
B



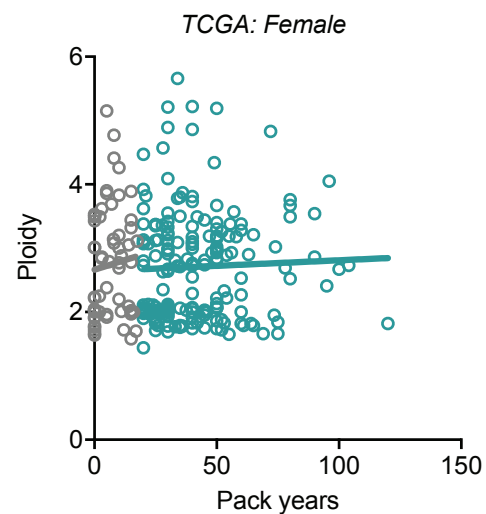
C



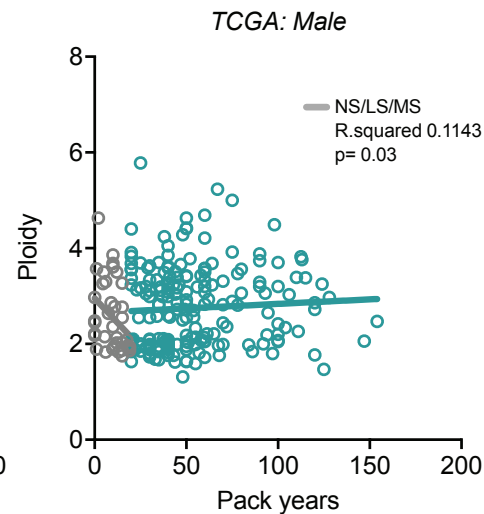
D



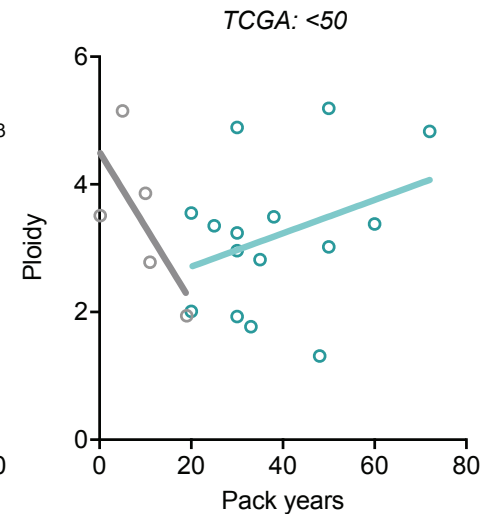
E



F



G



H

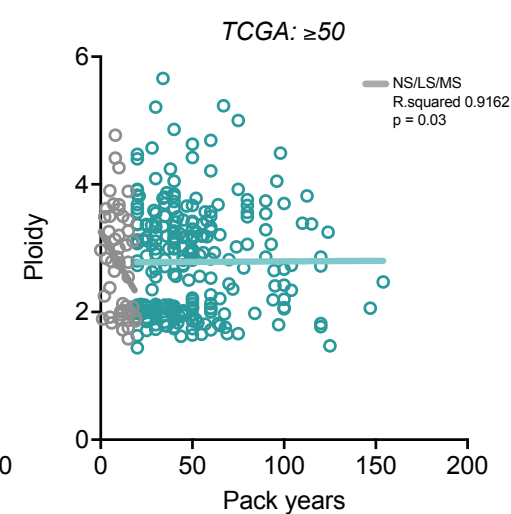


Figure S8. (A-B) Stacked bar graphs showing the percentage of samples with whole genome duplication (WGD) (black) or no WGD (white) in the TCGA (A) and MSK (B) cohorts. Enrichment of WGD was measured with a two-tailed Fisher's test. **(C)** Stacked bar graphs showing the percentage of samples with whole genome duplication (WGD) (black) or no WGD (white) in the TCGA cohort, split by age at diagnosis and grouped by smoking status. Enrichment of WGD was measured with a two-tailed Fisher's test **(D)** Scatterplot of ploidy versus pack-years cigarettes smoked in the WHI. **(E-H)** Scatterplot of ploidy versus pack-years cigarettes smoked in TCGA cohorts split by sex (E-F) or by age at diagnosis (G-H) Simple linear regression was used to identify any significant relationship between pack-years smoked and ploidy.

FIGURE S9

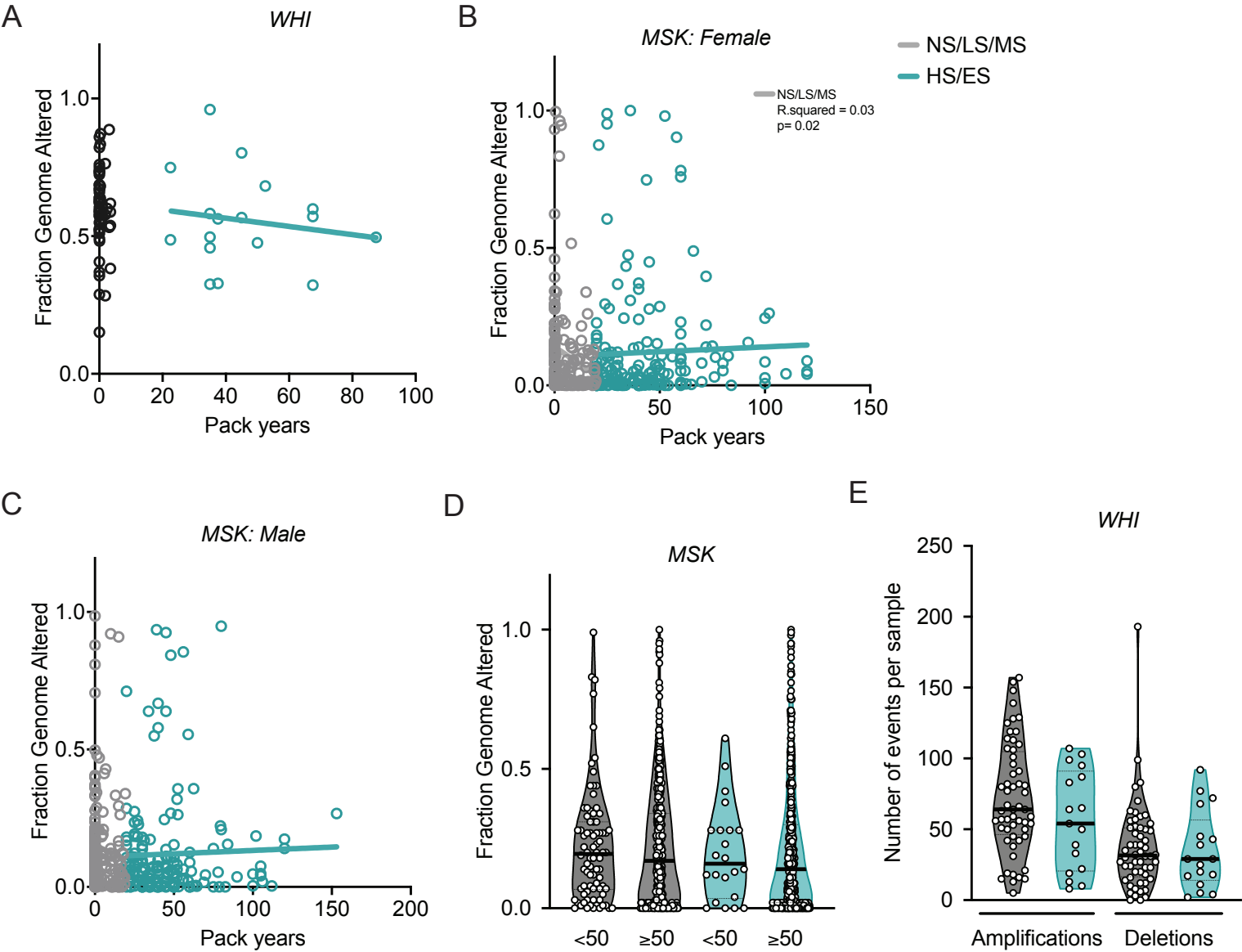


Figure S9. (A-C) Scatterplot of fractions genome altered (FGA) versus pack-years cigarettes smoked in the WHI and MSK cohorts. For the MSK cohorts scatter plots are split by sex (B-C). Simple linear regression was used to identify any significant relationship between pack-years smoked and FGA. **(D)** Fraction genome altered in the MSK cohort split by smoking status and grouped by age at diagnosis. **(E)** The total number of amplifications and deletions in the WHI cohort in never-/light smokers and heavy smokers. Kruskal-Wallis/Dunns Test was used to test significance (D-E).