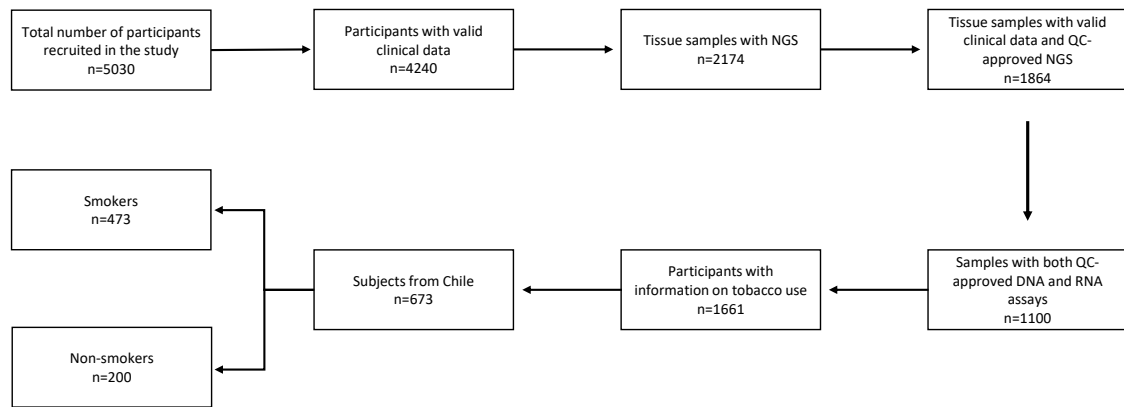


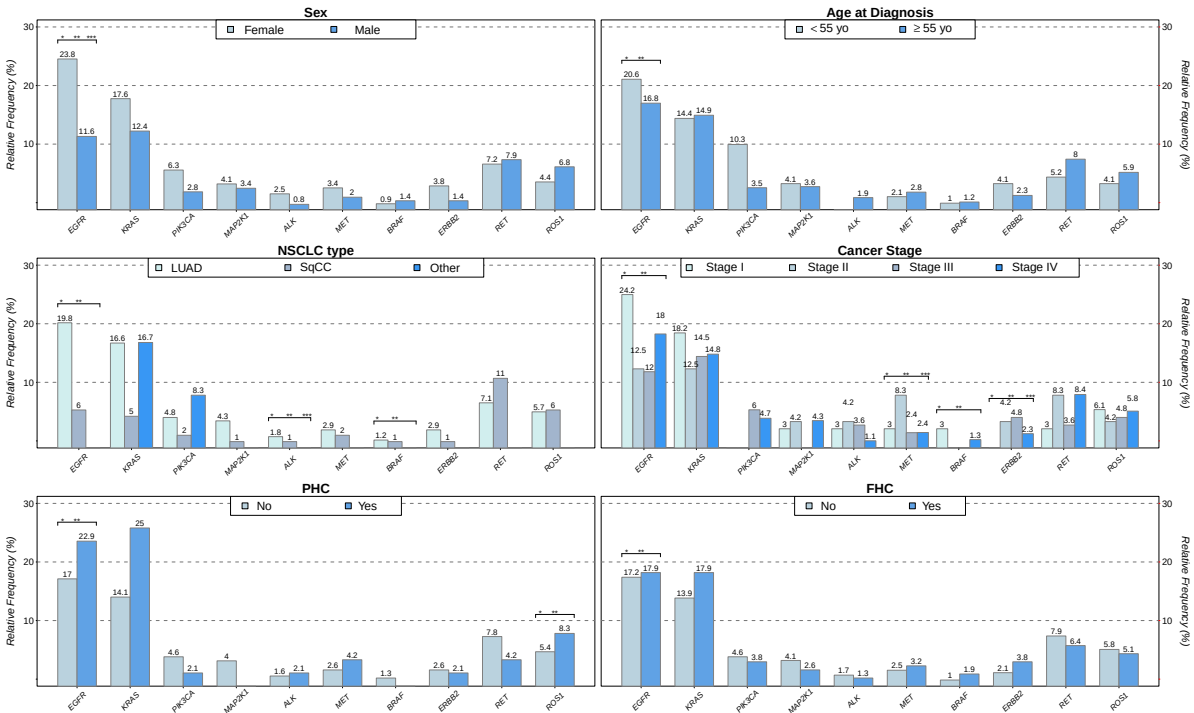
Additional file 1: Supplementary Figures

Figure S1: Study population flowchart.



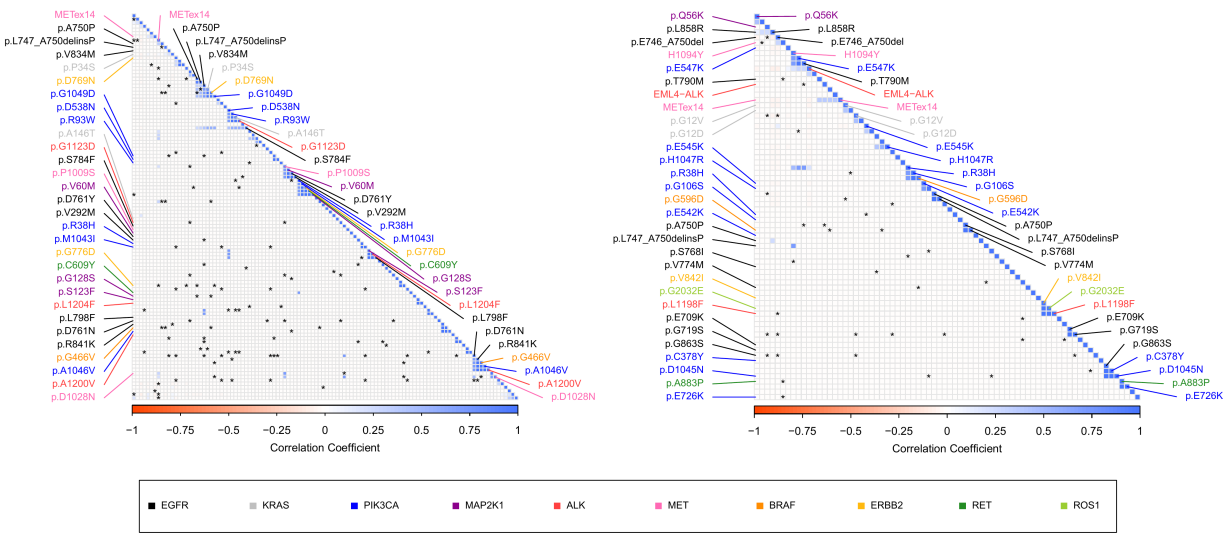
NGS: Next Generation Sequencing; QC: Quality Control.

Figure S2: Prevalence of GAs for the 10 driver genes by each of the main clinical variables.



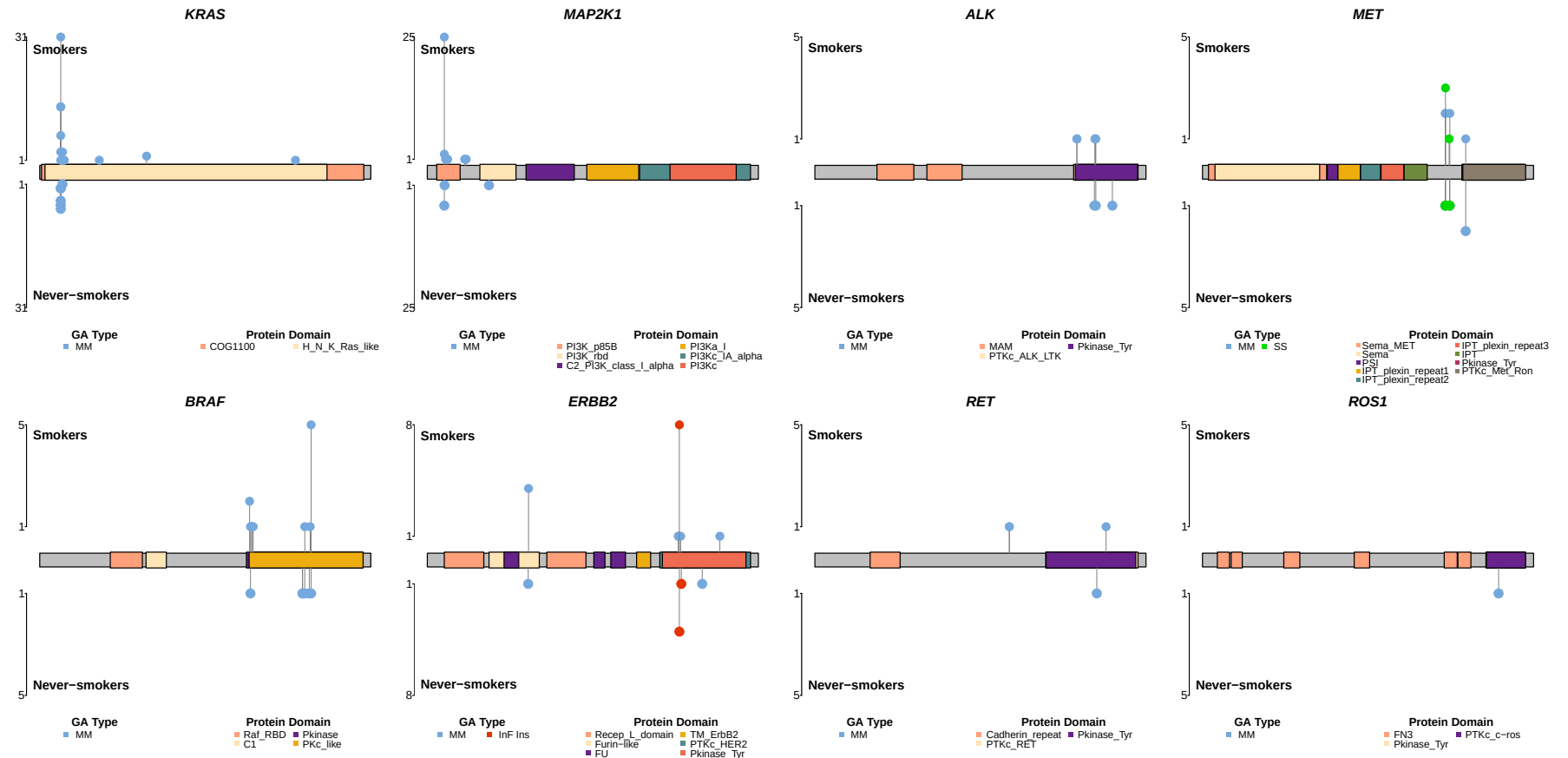
Genes are ordered by decreasing overall prevalence. Absolute and relative frequencies are specified for each group and gene. p-value for difference in proportions between groups was calculated using either the chi-squared test or the fisher exact test for small expected counts. (*) represents statistically significant difference (*: p-value<0.05, **: p-value<0.01, ***: p-value<0.001). NSCLC: Non-Small Cell Lung Cancer, LUAD: Lung Adenocarcinoma, SqCC: Squamous Cell Carcinoma, PHC: Personal History of Cancer, FHC: Family History of Cancer.

Figure S3: Matrices of co-occurrences and exclusions of the complete set of individual GAs.



The group of smokers are shown in the left and never-smokers in the right. The matrices display the pair-wise correlations between the complete set of GA in each group (smokers: $n=110$; never-smokers: $n=74$). GA are coloured by the genes to which they belong and are ordered according to the hierarchical cluster method. Negative correlations (exclusions) are coloured in red and positive correlation (co-occurrences) in blue. (*) represents statistically significant correlations at a 5% level.

Figure S4: Lollipop plots with the location and counts of individual GAs.



The group of smokers is shown at the top and never-smokers at the bottom. The genes *KRAS*, *MAP2K1*, *ALK*, *MET*, *BRAF*, *ERBB2*, *RET* and *ROS1* are displayed separately. GA: Genomic Alteration.