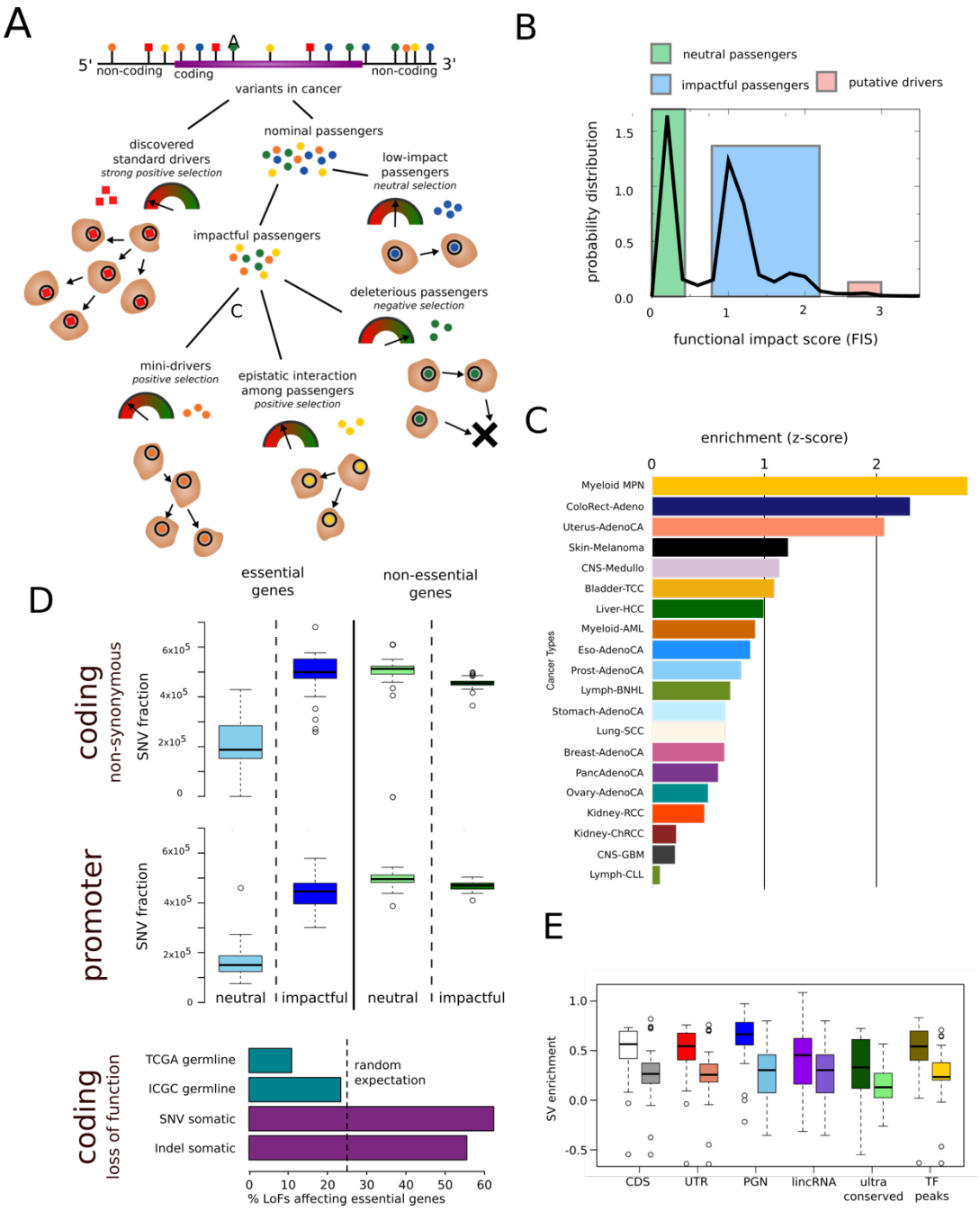


**Figure 1. A) Classification of somatic variants into different categories based on their selection characteristics:** deleterious passengers, epistatically interacting passengers and mini-drivers represent various categories of non-neutral passenger variants. B) Functional impact score distribution of non-coding SNVs present in the Myeloid cohort: we observe three peaks in impact score distribution corresponding to neutral passengers, non-neutral passengers and high impact variants, C) enrichment of non-neutral passengers SNVs in different cancer-subtypes in PCAWG: Multiple cancer-subtypes were enriched in non-neutral passenger variants , D) Fraction of pan-cancer neutral and non-neutral/impactful passenger SNVs influencing essential and non-essential genes in the coding(nonsynonymous & LOFs) and non-coding region (promoters), E) Fold enrichment score for somatic large deletions overlapping with different regions of the genome : pair of boxplot for each annotation correspond to enrichment score distribution for the engulfing(left) and partially overlapping (right) large deletions.

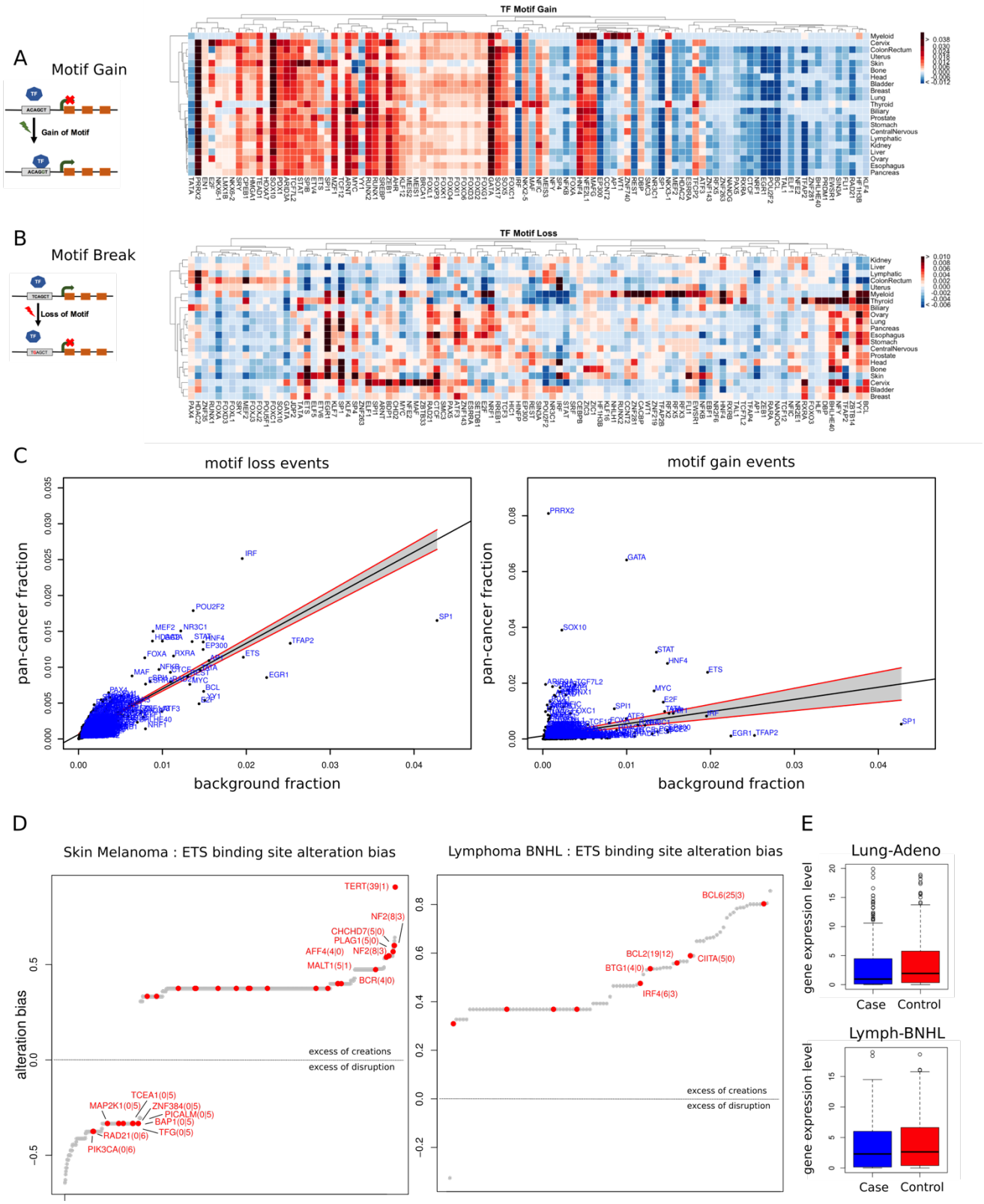
A
Motif Gain
Gain of Motif
TF Motif Gain
B
Motif Break
Loss of Motif
TF Motif Loss
C
motif loss events
motif gain events
pan-cancer fraction
background fraction
D
Skin Melanoma : ETS binding site alteration bias
Lymphoma BNHL : ETS binding site alteration bias
alteration bias
excess of creations
excess of disruption
TERT(39|1)
NF2(8|3)
CHCHD7(5|0)
PLAG1(5|0)
AFF4(4|0)
MALT1(5|1)
BCR(4|0)
TCEA1(0|5)
MAP2K1(0|5)
ZNF384(0|5)
PICALM(0|5)
BAP1(0|5)
TFG(0|5)
RAD21(0|6)
PIK3CA(0|6)
BCL6(25|3)
BCL2(19|12)
CIITA(5|0)
BTG1(4|0)
IRF4(6|3)
E
Lung-Adeno
Lymph-BNHL
gene expression level
Case
Control

Figure 2: **Overall functional burdening of TF motifs:** A) heat map presenting differential burdening of various TFs: SNVs leading to motif breaking events in different cohorts compared to the genomic background. B) heat map presenting differential burdening of various TFs: SNVs leading to motif gain events in different cohorts compared to the genomic background. C) Pan-cancer overview of TFs undergoing motif loss and gain events: scatter plots for motif loss inducing SNVs (left) and motif gain inducing SNVs (right). D) Target genes of TFBS undergoing alteration bias: target gene information for binding site of ETS transcription factor undergoing gain (positive bias score) or loss (negative bias score) of motif for the melanoma (left) and B-cell mature lymphoma (right). E) gene expression level changes due to motif breaking events: comparison of gene expression distribution for target genes in case and control.

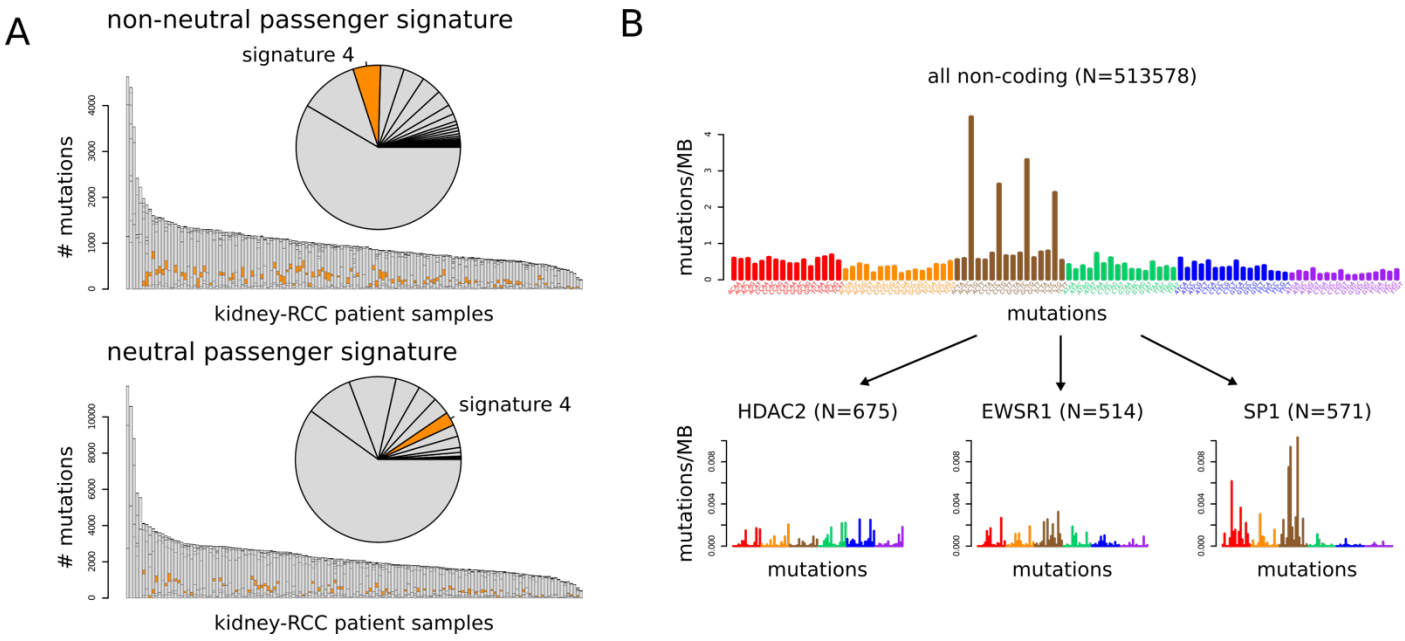


**Figure 3: Mutational signature analysis:** A) Distribution of canonical signatures in the Kidney-RCC cohort for non-neutral passenger (top) and neutral passenger SNVs (bottom). B) Mutation spectra associated with motif breaking events observed in HDAC2, EWSR1 and SP1 in the Kidney-RCC cohort.

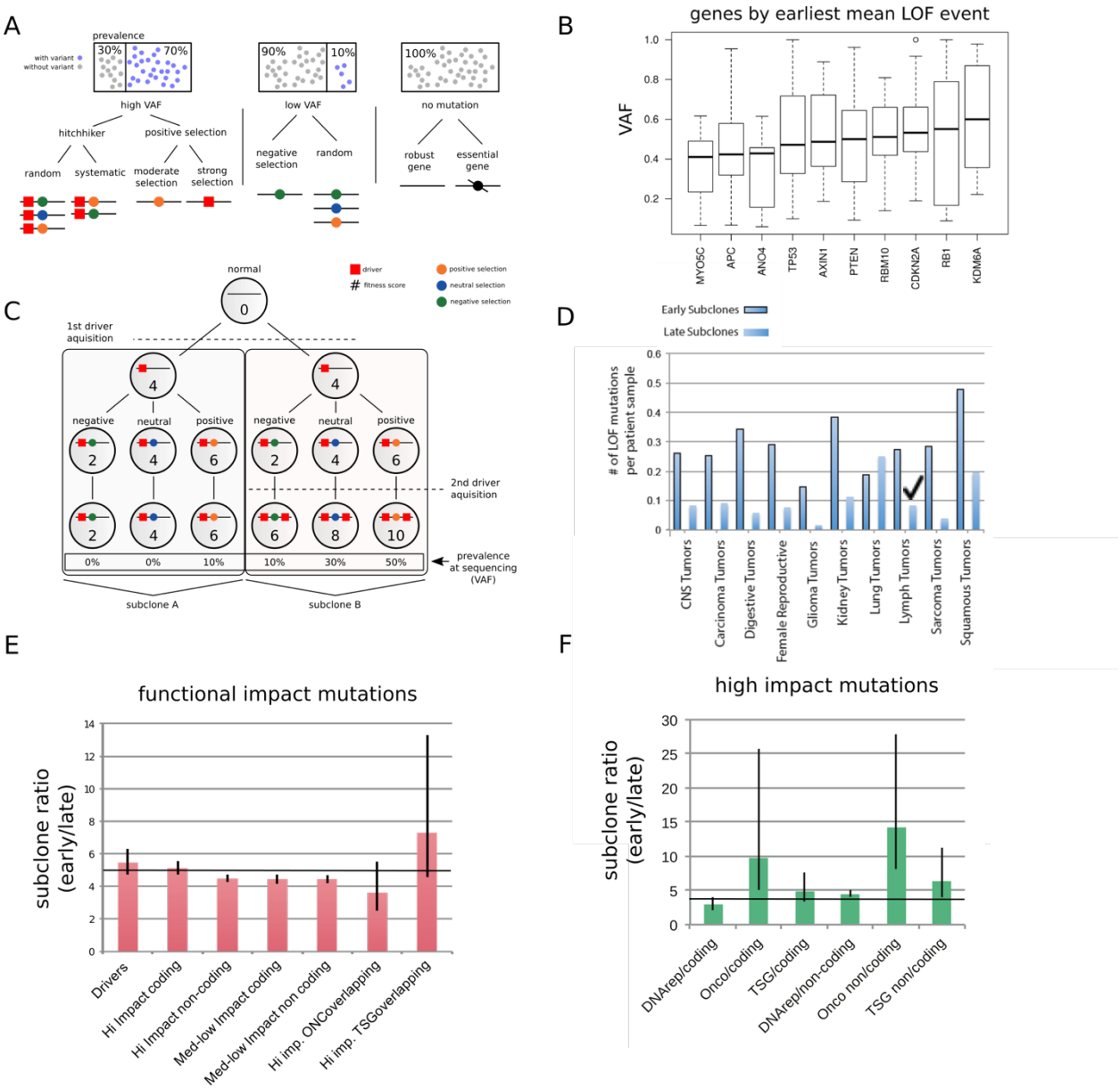


Figure 4: **Impact of coding and non-coding mutations in tumor progression**: A) distinct selection mode determines the observed VAFs for SNVs, B) distribution of VAFs associated with LOF events in distinct genes, C) selection workflow of passenger variants translates in their observed VAF values, D) Frequency of LOF SNVs in different cohorts for early (dark blue) and late subclones (light blue), E) Subclonal ratio (early/late frequencies) for different categories of SNVs (coding/non-coding) based on their impact score , F) Subclonal ratio (early/late frequencies) for high impact SNVs occupying distinct gene sets.

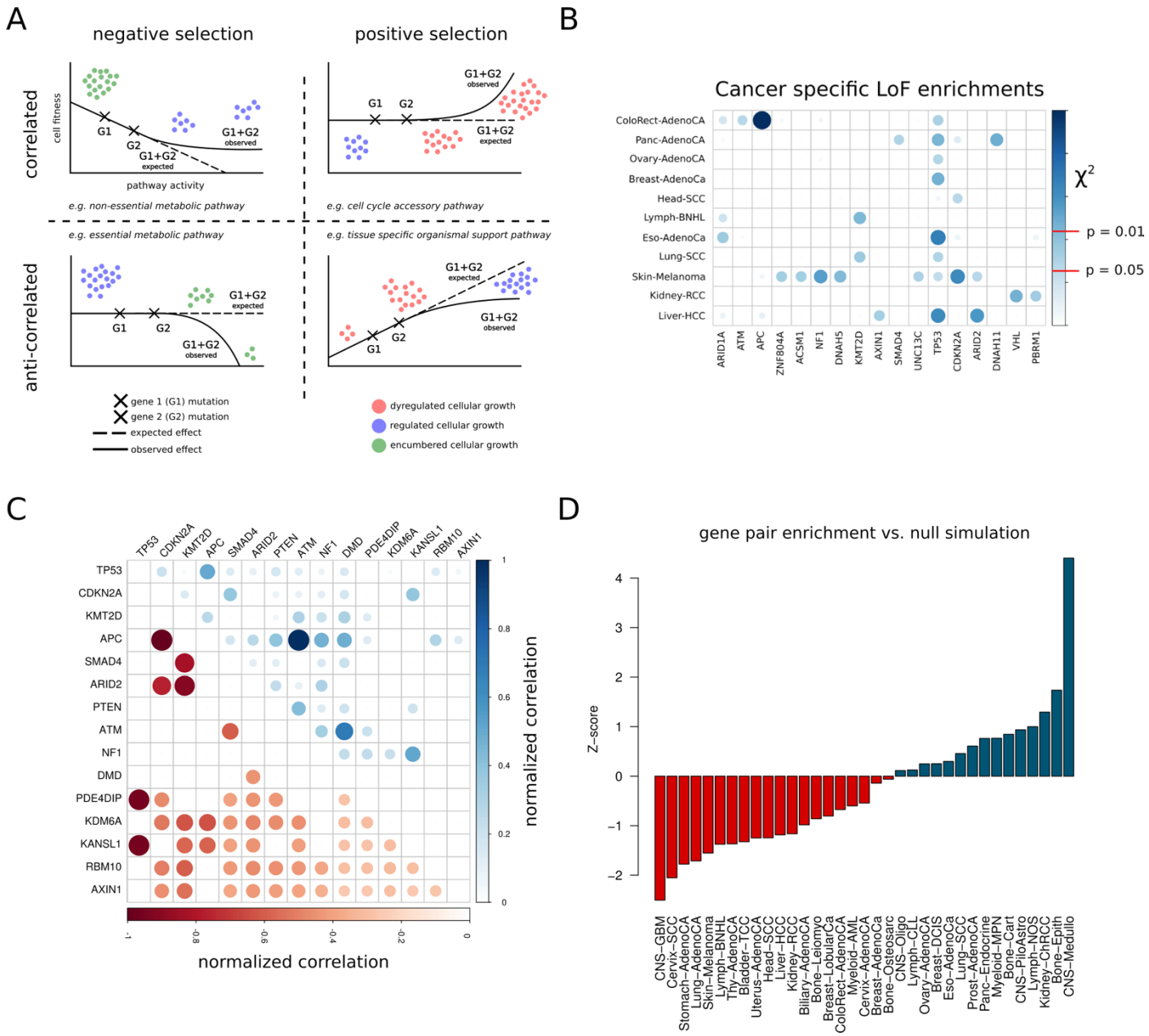


**Figure 5: Role of selection in driving gene pair interactions:** A) proposed mechanism explaining observed correlated and anti-correlated gene pairs, B) gene-level enrichment distribution of LOF SNVs in distinct cancer subtypes, C) Correlated and anti-correlated gene-pairs with LOF variants in the pan-cancer dataset, D) enrichment score of co-mutated gene pairs in different cancer-subtypes.

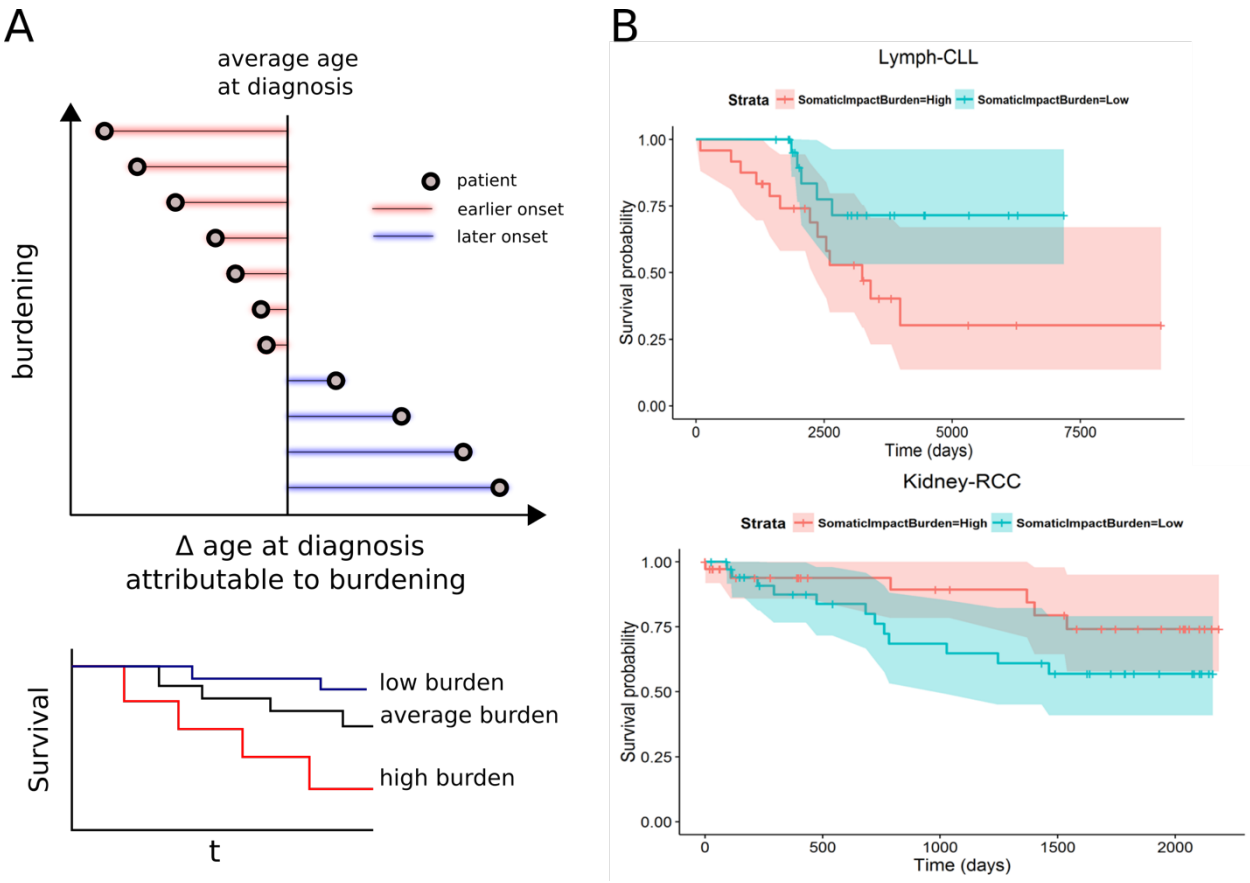


**Figure 6: Correlating functional burdening with patient age and patient survival:** A) Linear regression was applied for predicting age in years at diagnosis of cancer from somatic and germline mutation burden. B) Survival curves in CLL (*top panel*) and RCC (*bottom panel*) with 95% confidence intervals, stratified by normalized impact burden.