# Tobacco smoke alters the trajectory of lung adenocarcinoma evolution via effects on somatic selection and epistasis

Krishna Dasari,[1,2] Jorge A. Alfaro-Murillo,[1] Jeffrey P. Townsend[1,3,4,5*]

[1]*Department of Biostatistics, Yale School of Public Health, New Haven, CT.*

[2]*Medical Scientist Training Program, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.*

[3]*Yale Cancer Center, Yale University, New Haven, CT.*

[4]*Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT.*

[5]*Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT.*

*Corresponding author. Email: jeffrey.townsend@yale.edu

## Abstract

### Background

Tobacco smoke is a known mutagen. However, its physiological effects on the lung may also influence the somatic selective pressures acting on mutations, further shaping cancer evolution. The relative contributions of these mutagenic and physiological effects to oncogenesis have not been quantified, despite their importance in predicting the differential therapeutic effect of targeting variants in lung cancers of smokers and nonsmokers.

### Methods

We classified 1,722 lung adenocarcinoma (LUAD) sample genomes from The Cancer Genome Atlas and other projects as ever-smoker (ES) or never-smoker (NS) LUAD based on smoking-associated mutational signature attribution or clinical annotation. We then independently calculated background oncogenic mutation rates in the ES- and NS-LUAD groups. Comparing these background rates to observed variant prevalences enabled us to estimate and compare the selective advantages conferred by each mutation in ES- and NS-LUAD. Finally, we quantified pairwise and higher-order epistatic effects by estimating selection for each mutation in specific somatic genotypes.

### Results

As expected, background oncogenic mutation rates were gene-specifically elevated in ES-LUAD. However, differences in oncogenic mutation rates between ES- and NS- LUAD were insufficient to explain differences in the prevalence of some mutated genes, implying that such mutations must have further experienced differential somatic selection. In particular, *KRAS*, *KEAP1*, and *STK11* mutations experienced substantially stronger selection in ES-LUAD, whereas mutations of *EGFR*, *PIK3CA*, *SMAD4*, and other genes were more strongly selected in NS-LUAD. Mechanistically, *EGFR* mutations were associated with upregulation of genes

involved in the epithelial-mesenchymal transition in NS-LUAD, but not in ES-LUAD. Epistasis was pervasive and distinct between the subtypes: ES-LUAD featured more frequent synergy and substantially less antagonism. These patterns entail divergent evolutionary trajectories, with NS-LUAD constrained to fewer, narrower paths and ES-LUAD exploring a broader, more permissive adaptive landscape. Furthermore, we identified higher-order epistasis—systematically examined here for the first time in cancer—manifesting as sub-additive and emergent synergistic interactions that selectively promote trajectories of oncogenesis.

**Conclusions**

This disambiguation of the mutagenic and selective effects of tobacco smoke reveals how environmental insult can reshape the evolutionary trajectories of LUAD and enables quantitative prediction of treatment vulnerabilities based on smoking status and tumor somatic genotype.

**Keywords**

Lung adenocarcinoma, tobacco smoke, cancer genetic evolution, somatic selection, epistasis, microenvironmental interactions

## Background

Tobacco smoke is well known to cause cancer, in part through its mutagenic effects [1]. Carcinogen exposure from tobacco generates distinct patterns of mutations [2], a subset of which then contribute to oncogenesis [3–9]. In addition to this mutagenic effect, tobacco smoke induces physiological changes in the lung—inflammation, oxidative stress, immune dysregulation, and tissue damage [1,10–12]—that can alter the lung microenvironment and thereby the selective pressures acting on specific somatic mutations. These microenvironmental disruptions could modulate the fitness advantages conferred by specific mutations, thereby influencing their likelihood of contributing to cancer development. Indeed, several studies have suggested that, in principle, endogenous or exogenous alterations to the microenvironment caused by factors such as aging or tobacco exposure likely shape the trajectories of cancer evolution [13–17]. However, the extent to which such physiological effects alter the adaptive landscape of cancer has not previously been quantified. Therefore, quantification of the adaptive impacts of a highly disruptive exogenous insult such as tobacco smoke is crucial to understanding how environmental exposures shape the evolution of individual cancer evolutionary trajectories.

Among lung cancer subtypes, LUAD is the most common and accounts for the highest proportion of cases in never-smokers [18]. As of 2002, nearly 23% of LUADs in the United States occurred in never-smokers [19], a proportion that has been rising as cigarette usage declines [20]. Numerous studies have reported that LUAD exhibits distinct somatic variant profiles in never-smokers compared to ever-smokers, which has been taken as evidence of their divergent genetic vulnerabilities [21–25]. However, differences in variant prevalence alone are confounded as to whether they are a consequence of effects of tobacco-induced mutagenesis or a consequence of tobacco-altered selection dynamics. Therefore, quantification of the relative oncogenic roles of the mutagenic and physiological effects of smoking in LUAD will be

increasingly crucial to determination of optimal use of targeted therapies for patients with and without a prior smoking history.

Computational models now enable inference of baseline somatic mutation rates from synonymous mutations observed in tumor genomes—implicitly [6] or explicitly [26] accounting for carcinogen-induced mutagenesis, such as that caused by tobacco smoke. Comparing these baseline mutation rates to the observed prevalence of nonsynonymous mutations enables estimation of somatic selection acting on specific substitutions [8,27,28]. This framework provides a foundation for disentangling the oncogenic contributions of the mutagenic and selective effects of smoking. Moreover, selection on each mutation is likely influenced not only by the microenvironmental effects of tobacco smoke, but also by pairwise [28–30] and higher-order [28] selective epistasis between mutations. In selective epistasis, mutations affect the selection operating on other mutations. These interactions between the selective effects of mutations shape the adaptive landscape of cancer evolution [5,28,31,32]. Understanding how tobacco smoke influences not just mutation and selection, but also epistasis, is essential for a complete picture of LUAD evolution.

Because knowledge of epistasis is crucial to inference of cancer-driving mutation modules and synthetic vulnerabilities, a number of algorithms have been developed that attempt to detect epistasis between mutations or sets of mutations by quantifying over- and under-representation [33,34] in the context of heterogeneous mutation rates [30,35–37]. However, with the exception of Coselens [30], these methods do not enable meaningful quantitative comparisons of the effects of mutations in distinct somatic genotypes. This quantification is a crucial step for prioritization and translation of knowledge of epistasis into effective therapeutic strategies. Moreover, all of these approaches are restricted to pairwise inferences [30,34] or to inference of discrete modules [33,35–37], and are not well suited to the quantitative inference of higher-order epistasis (involving more than two mutations). When

present, higher-order epistasis substantially influences the adaptive landscape and shapes evolutionary trajectories [38–43], including the degree to which selective epistasis diminishes along the adaptive trajectory [44–47]. Systematic characterization of the presence and biological significance of higher-order epistasis to cancer evolution has been challenging. Recent theory facilitates understanding of higher-order epistasis in cancer by providing the selection strength of mutations for each somatic genotype [28]. This novel approach has not yet been applied in the context of environmental risk factors that could also affect selection, such as tobacco smoke. Accordingly, analysis of the effect of smoke on the adaptive landscape of LUAD will benefit from consideration of pairwise and higher-order epistatic effects and potential differences in epistasis between microenvironments.

To address these gaps, we quantified the mutagenic, selective, and epistatic effects of tobacco smoke on single-nucleotide variants (SNVs) in LUAD. We assembled a large dataset of lung adenocarcinoma tumor genomes and classified them by smoking status into ever-smoker (ES) and never-smoker (NS) groups. We estimated background mutation rates in each group and inferred the strength of somatic selection acting on driver mutations in each gene. We then investigated how selection pressures differed between ES- and NS-LUAD and associated these differences with differential gene expression. Finally, we extended our analysis to inference of selection in specific somatic genotypes, detecting and quantifying the extent of pairwise and higher-order epistasis. The degree of divergence in the mutational, selective, epistatic, and transcriptional landscapes of LUAD between ever-smokers and never-smokers enables novel insight into how environmental exposure shapes cancer evolution—identifying genotype- and environment-specific vulnerabilities that can inform personalized therapeutic strategies.

# Methods

## Design overview

The aim of this study was to test for differences in background mutation rates, subsequent somatic selection, and gene-by-gene epistatic effects between ES- and NS-LUAD. To do so, we aggregated mutation data of 2,097 LUAD tumor samples from multiple institutional sources. We classified tumors as either ES- or NS-LUAD either by detection of the smoking-associated mutational signature SBS4 in exome or genome sequences or, in the case of targeted panel sequence data, by relying on the patient's self-reported smoking history. We applied established approaches to estimate background mutation rates, scaled selection coefficients, and pairwise and higher-order selective epistasis within the ES- and NS-LUAD cohorts. To assess the differential transcriptional effects of driver mutations, we compared ES- and NS-LUAD tumors with and without specific mutations. Differences in selection were conservatively evaluated based on non-overlapping 95% confidence intervals.

## Data sources

We sourced clinicogenomic data from (i) eight published datasets hosted by cBioPortal [48–56], (ii) The Cancer Genome Atlas dataset hosted on Genomic Data Commons [57], and (iii) 108 samples sequenced at Yale University [58]. All datasets reported sample sizes, sequencing methods, and clinical covariates, and were incorporated into our analysis accordingly (**Supplementary Table 1**). We only included samples from primary LUAD tumors, and we did not include any datasets from studies that preselected patients based on their mutational profiles. When multiple samples from a single patient were available, one was selected at random. In cases where patients appeared in more than one dataset, we retained the sample with the highest-quality sequencing and most complete clinical annotations.

For consistency across datasets, we standardized all variant genomic coordinates to the hg19 reference genome using the liftOver function of rtracklayer v1.62.0 [59]. We used cancereffectsizeR v2.10.1 to remove variants that failed liftover, were incorrectly classified as single-nucleotide variants (SNVs), or did not meet quality criteria for downstream analysis. For panel-sequenced samples, we inferred the genomic regions covered by each panel from the gene list provided for each panel using the corresponding hg19 coordinates.

## Mutational signature attributions and smoking-status classification

We performed mutational signature deconvolution for tumor samples with whole-exome or whole-genome sequence data, but not for panel-sequenced samples, due to the low reliability of signature attribution when few silent somatic variants are present. We conducted signature attribution using cancereffectsizeR with MutationalPatterns v3.12.0 [60], applying COSMIC v3.2 single-base substitution signatures [26,61]. We excluded COSMIC signatures that were unlikely to have been observed in lung adenocarcinomas [61], and we excluded treatment-associated signatures when deconvolving samples known to be treatment-naive. Samples in which the smoking-associated SBS4 signature contributed more than the MutationalPatterns default threshold of 5% of the total signature burden were classified as ever-smoker ES-LUAD, and otherwise were classified as never-smoker NS-LUAD. Zhang et al. detected no signature of SBS4 in the samples used in their study; therefore we directly classified these samples as NS-LUAD. As in Rosenthal et al. [62], we excluded samples with ≤50 single-nucleotide variants in whole-exome data from classification due to insufficient mutation burden for reliable signature deconvolution. We compared signature-based smoking assignments to available clinical annotation of smoking status across datasets, validating that there was 67–90% concordance. Discrepancies were likely due to underreporting of smoking, second-hand smoke exposure, minimal mutational consequences in light smokers, or misclassification in clinical records. The

selection and epistasis results produced from either classification method were 99.4% consistent with each other. We classified panel-sequenced samples according to clinical annotations of smoking status when available; we excluded samples annotated as exposed to "former light" smoking from all smoking-stratified analyses due to ambiguity in exposure and effect.

## Estimation of oncogenic mutation rates

We estimated gene-specific oncogenic mutation rates separately for the ES- and NS-LUAD cohorts following the approach of Alfaro-Murillo and Townsend [28]. The oncogenic mutation rate for a gene was defined as the neutral rate at which oncogenic variants occur within the gene. We defined oncogenic variants as variants that were observed at least once in any tumor sample in the dataset. The background oncogenic mutation rate of each gene was computed as the sum of the background rates of all oncogenic variants within the gene. For a specific variant derived from a trinucleotide mutation type $m$ (*e.g.* CCT → CAT) at a site $i$ within gene $g$ in a tumor sample cohort $c$, the somatic background mutation rate

$$\mu_{i,g,c}^{m} = \mu_{g,c} \cdot \mathrm{Prob}\left\{T_m|c\right\} \cdot \mathrm{Prob}\left\{S_{i,m_{\mathrm{WT}},g}\right\},$$

where $\mu_{g,c}$ is the background mutation rate of the entire gene $g$ (the rate at which variants occur in $g$) within cohort $c$, $T_m$ is the event that an arising variant is of trinucleotide mutation type $m$, and $S_{i,m_{\mathrm{WT}},g}$ is the event that the variant occurs at site $i$ among all other trinucleotide sequences that match the wild-type (WT) allele of $m$ ($m_{WT}$, *e.g.* CCT), within $g$.

The probability of the variant being of type $m$ given a cohort $c$, $\mathrm{Prob}\left\{T_m|c\right\}$, is computed as the proportional frequency of $m$ within cohort $c$, equivalent to the ratio of the number of variants of type $m$ in cohort $c$, $n_m^c$, to the total number of variants in $c$, $N_c$. Assuming an equal probability of such a variant occurring at all possible sites within $g$, the probability of such a variant occurring at site $i$, $\mathrm{Prob}\left\{S_{i,m_{\mathrm{WT}},g}\right\}$, is the inverse of the number of sites within $g$ with the same trinucleotide sequence as $m_{\mathrm{WT}}$, $S_{m_{\mathrm{WT}}}^{g}$. Therefore

$$\mu_{i,g,c}^{m} = \frac{\mu_{g,c} \cdot n_{m}^{c}}{S_{m_{\mathrm{WT}}}^{g} \cdot N_{c}}.$$

Finally, the background oncogenic mutation rate of gene $g$ is the sum of the background rates of all oncogenic variants within $g$:

$$\hat{\mu}_{g,c} = \sum_{(i,m) \in \mathcal{S}_g} \mu_{i,g,c}^{m}$$

where $\mathcal{S}_g$ denotes the set of pairs of site indices and mutation types identifying the oncogenic variants in $g$. This approach enables estimation of mutation rates that are both gene-specific and context-dependent, accounting for trinucleotide mutational biases across the genome. The oncogenic mutation rates of the genes were consistently lower than the mutation rates of the genes.

### Estimation of scaled selection coefficients and selective epistasis

To quantify the growth and survival advantages conferred by somatic SNVs in LUAD, we calculated scaled selection coefficients for recurrent mutations in a curated set of 1,200 genes [57,63–66]. These coefficients roughly equate to the ratio of the fixation rate for the oncogenic mutant (the rate at which the mutant is generated and subsequently selected to a high enough frequency to be observed as a somatic variant within the cancer cell population) to its background mutation rate. Precisely, we computed the fixation rate by analytical maximization of the binomial likelihood of observing at least one non-silent SNV in the gene given the number of samples with and without a non-silent SNV in the gene. Scaled selection coefficients for each mutated gene were estimated separately for ES- and NS-LUAD, as well as for a combined cohort. Within each group, we included samples with whole-exome, whole-genome, or targeted gene panel sequences (TGS). Results were consistent when excluding TGS samples, indicating robustness to sequencing strategy. Inclusion of TGS samples provided greater statistical power to detect significant differences in selection between cohorts. Results were also robust to

cohort-specific normalization: applying median-scaling of coefficients prior to between-cohort comparison yielded consistent outcomes.

From the curated set of 1,200 genes, we selected 21 for detailed analysis of selective epistasis based on strong prior evidence as LUAD drivers [57,64] and their elevated average strength of selection across somatic genotypes. We quantified epistasis by calculating scaled selection coefficients conditional on the somatic genotype as in Alfaro-Murillo and Townsend [28]: for each pair or triplet of genes, we calculated the fixation rate of mutations in one gene conditional on the presence or absence of somatic variants in the others. We performed this calculation by maximizing the multinomial likelihood for the number of samples with each somatic genotype using PyMC [67]. The probability of a tumor being in a specific somatic genotype during the somatic evolutionary trajectory from organogenesis to biopsy was computed via a continuous-time Markov chain model [28]. We computed asymmetric 95% confidence intervals using Wilks' theorem [68]. To obtain scaled selection coefficients for each mutation occurring within each somatic genotype, we divided the respective Poisson-corrected fixation rate [69] by the background oncogenic mutation rate of the gene.

To systematically assess selective epistasis in both the ever-smoker and the never-smoker groups, we evaluated 3,080 models representing all possible two- and three-gene combinations from the 21 LUAD driver genes. For each model, somatic genotypes were defined exclusively by the variant status (mutant or wildtype) of the genes included in each combination. To ensure complete genotype classification, we only included panel-sequenced samples in a given model if the sequencing panel covered all of the genes under analysis.

## Classification of synergistic and antagonistic selective epistasis

To classify a mutant somatic genotype as interacting synergistically or antagonistically with a certain mutation, we compared the scaled selection coefficients for a mutation when it

occurred in the mutant somatic genotype to the scaled selection coefficients obtained for the same mutation in a somatic genotype lacking mutations in the genes included in the model. If the scaled selection coefficient for a mutation in gene *A* when it occurred in a wild-type genotype ($\varnothing \rightarrow A$) was less than the scaled selection coefficient for a mutation in gene *A* when it occurred in a somatic genotype with a mutation in gene *B* ($B \rightarrow A + B$), then we deemed the epistatic effect of mutant gene *B* on mutant gene *A* to be synergistic. If the scaled selection coefficient for $\varnothing \rightarrow A$ was greater than the scaled selection coefficient for $B \rightarrow A + B$, then we deemed the epistatic effect of mutant gene *B* on mutant gene *A* to be antagonistic. We further distinguished between two scales of antagonism: antagonistic sign epistasis when there was a reversal of the direction of selection—from positive (coefficient $\gamma > 1$) to negative selection (coefficient $\gamma < 1$)—in the antagonistic genotype, and antagonistic magnitude epistasis when there was a reduction in strength of selection but selection remained positive in the antagonistic genotype.

### Differential expression analysis

To evaluate the transcriptional impact of key driver mutations across smoking-stratified LUAD tumors, we performed differential expression analysis using RNA-seq data from the TCGA-LUAD project. We obtained unstranded raw read counts for 59 normal lung and 539 primary LUAD tumor samples using TCGAbiolinks v2.30.0. We excluded formalin-fixed paraffin-embedded samples and tumors which received neoadjuvant treatment to minimize technical and biological confounders. We retained solid-tissue normal samples as controls. We excluded genes with fewer than 10 total counts across all samples. For each sample, we assigned smoking status based on SBS4 signature weight from the matched MAF. The genes *EGFR*, *KRAS*, and *KEAP1* were considered mutated if at least one non-silent SNV, insertion, or deletion was present in the gene (excluding intronic regions). After this filtering and annotation,

we validated that a principal-component analysis revealed no evidence of batch effects, confirming the suitability of the dataset for differential expression analysis.

We performed differential expression analysis with DESeq2 v1.42.1 [70] using two designs. Differential transcriptional effects of *EGFR* mutations in ES- and NS-LUAD were quantified as

$$K_i = \beta_0 + \beta_1 x_t + \beta_2 x_s + \beta_3 x_e + \beta_4 x_r + \beta_5 (x_s * x_e) + \beta_6 (x_s * x_r),$$

where $K_i$ is the count for gene $i$, and $x_t$, $x_s$, $x_e$, $x_r$ are indicator variables for the sample type, smoking status, *EGFR* mutation status, and *KRAS* mutation status. The differential effect in question was measured by the interaction term between smoking status and *EGFR* allele status. The *EGFR*-focused-design accounts for the fact that most *EGFR*-wildtype LUADs are *KRAS*-mutant. Differential transcriptional effects of *KEAP1* mutations in ES- and NS-LUAD, were quantified as

$$K_i = \beta_0 + \beta_1 x_t + \beta_2 x_s + \beta_3 x_k + \beta_4 (x_s * x_k),$$

where $x_k$ is the *KEAP1* mutation status.

We performed gene-set enrichment analysis with fgsea v1.28.0 using the MSigDB Hallmark gene sets. We ranked transcripts by signed *t*-statistics to capture both degree of significance and magnitude of fold change. We performed gene set variation analysis with GSVA v1.50.5 using the MSigDB canonical gene sets.

## Statistical Analysis

Differences in selection coefficients were assessed conservatively by non-overlapping 95% confidence intervals for each estimate. Differential gene expression was analyzed using Wald tests, and statistical significance was determined based on Benjamini-Hochberg–adjusted $P < 0.05$. Gene set enrichment analyses were likewise considered significant at a

Benjamini-Hochberg–adjusted $P < 0.05$. Analysis was performed in Python (3.9.5) and R (4.3.0).

## Results

### Mutation rates and somatic selection pressures jointly shape the prevalence of driver variants in lung adenocarcinoma

To disentangle the relative contributions of mutation and selection to observed mutation frequencies in LUAD, we deconvolved the prevalence of somatic variants in our curated set of 21 well-supported driver genes into estimates of the background oncogenic mutation rates and scaled selection coefficients. Across all smoking-status-classified LUAD samples in our dataset ($n$ = 1,722), *TP53* and *KRAS* were the most frequently mutated genes, followed distantly by *EGFR*, *KEAP1*, and *STK11* (**Fig. 1A**). Background oncogenic mutation rates for the majority of these genes spanned slightly more than an order of magnitude—between $7.6 \times 10^{-8}$ and $8.7 \times 10^{-7}$ (**Fig. 1B**). When compared to the observed variant prevalences, this variation in background oncogenic mutation rate frequently implies substantial and variable impacts of somatic selection on the prevalence of each variant.

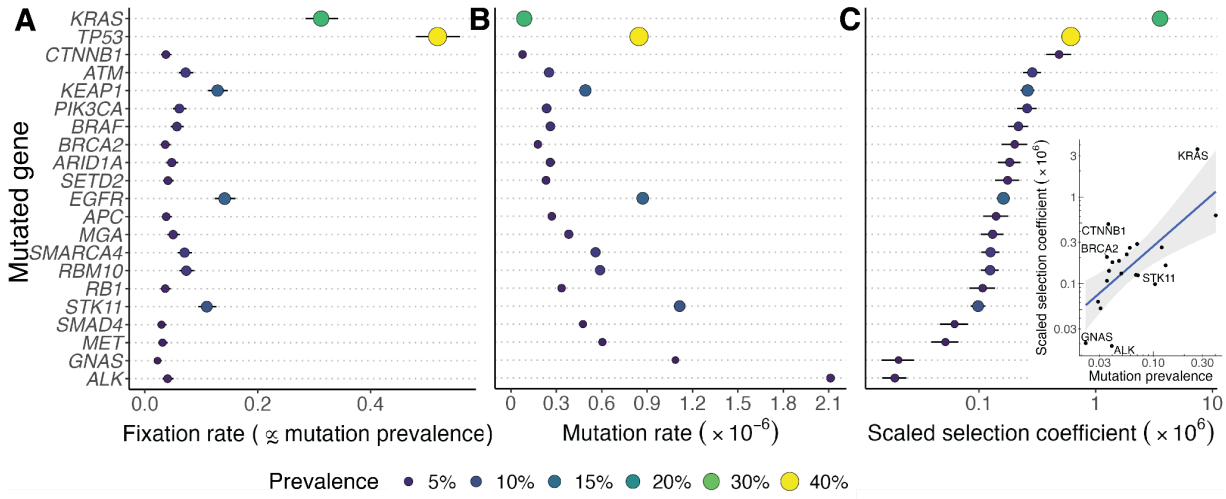


**Figure 1. Fixation rates and scaled somatic selection coefficients for 21 known and potential driver mutations among 1,722 lung adenocarcinomas.** (**A**) Poisson-corrected fixation rates (rates at which each mutation occurs and becomes fixed within the population). (**B**) Background rate at which oncogenic gene mutations occur. (**C**) Strength of somatic selection (log-scaled) experienced by oncogenic gene mutations. Inset: Correlation between the prevalence of an oncogenic mutation and the strength of selection acting on it ($P < 0.001$).

Scaled selection coefficients for oncogenic mutations in each gene exhibited a long-tailed distribution (**Fig. 1C**), with mutations in *KRAS* experiencing the strongest selection, followed distantly by mutations of *TP53* and *CTNNB1*, then mutations in *ATM*, *KEAP1*, and *PIK3CA*, and then other genes that experienced weaker but still substantial selection (**Fig. 1C**). Notably, *CTNNB1* mutations were present in only 3.7% of tumors. Nevertheless, knowledge of the low underlying oncogenic mutation rate of *CTNNB1* indicates that these infrequent mutations experienced stronger selection than did the more frequent mutations in *STK11, RBM10*, and *SMARCA4*. Generally, selection and mutation prevalence were correlated ($r = 0.68$, $P < 0.001$), but several discrepancies were present between the prevalence of a mutation and the relative intensity of selection it experienced (**Fig. 1C,** inset). These discrepancies highlight the importance of reporting selection as the effect size of mutations, rather than reporting solely prevalence, which is only moderately associated with strength of selection.

The oncogenic mutation rates and selection coefficients presented above were quantified based on all tumors, regardless of smoking status. However, both the mutagenic processes and selective advantages of mutations were plausibly influenced by the genotoxic and physiological effects of tobacco smoke. To assess these distinct effects, we stratified tumors by smoking status and performed separate analyses of somatic selection in ES- and NS-LUAD tumors.

## Mutation and selection differ substantially between ever-smoker and never-smoker LUAD

Stratification of LUAD tumors by smoking status based on mutational signature attributions and clinical annotation yielded 1,066 ES-LUADs and 656 NS-LUADs. In each group, we independently estimated oncogenic mutation rates and selection coefficients for the 21 well-established driver genes (**Fig. 2**). Both mutation rates and selection pressures differed markedly between ES- and NS-LUAD. Oncogenic mutation rates were elevated in ES-LUAD across all 21 genes. Increases ranged from modest (39%–43% for *RB1 and CTNNB1*) to extreme (571–1474% for *EGFR*, *SMAD4*, *GNAS*, and *ALK*; **Fig. 2A inset**). These elevated mutation rates demonstrate the intense mutagenic effect of tobacco smoking on the peripheral tissue of the lung.

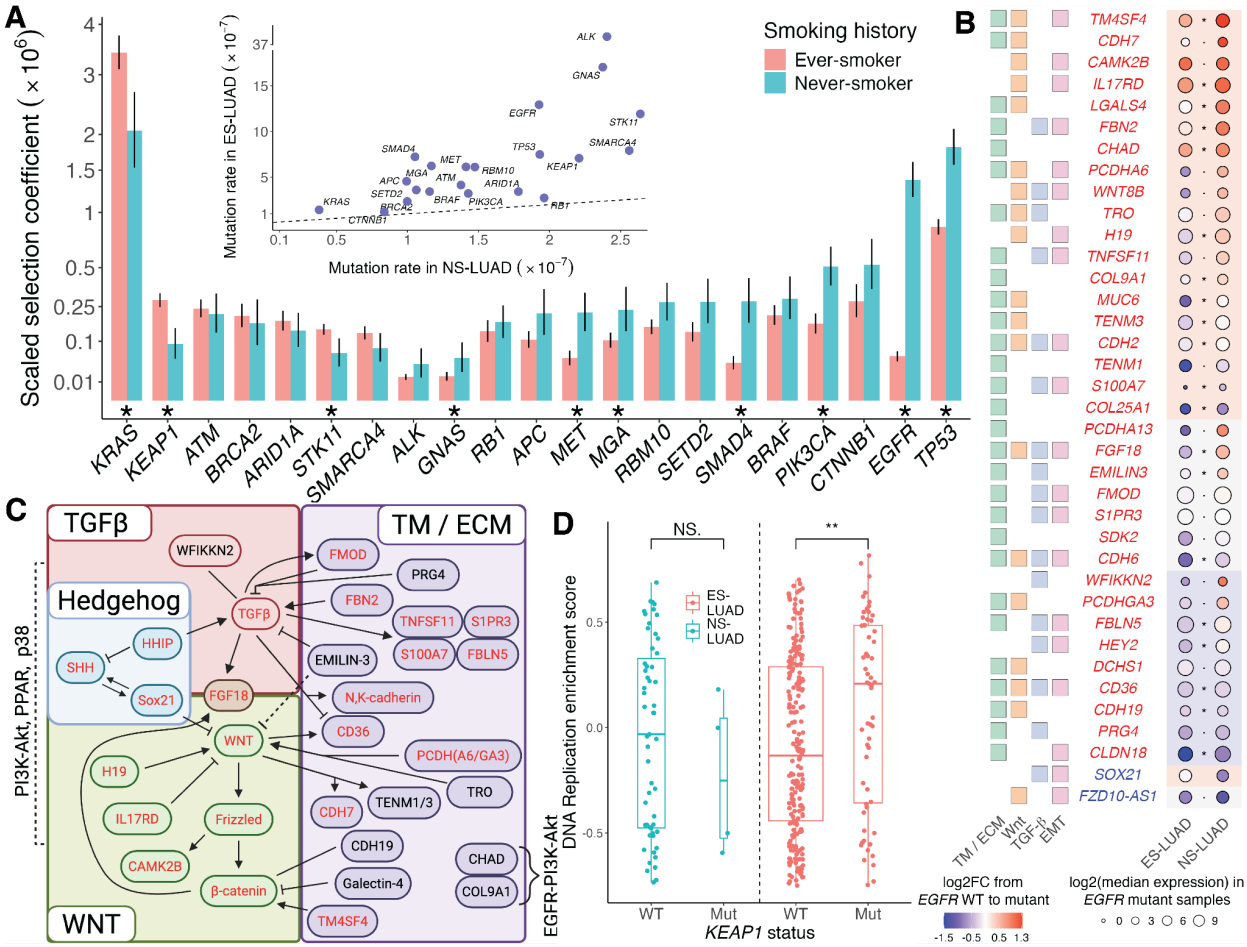


**Figure 2. Mutational and adaptive landscapes of ever-smoker lung adenocarcinoma (ES-LUAD) and never-smoker lung adenocarcinoma (NS-LUAD).** (**A**) Strengths of somatic selection (*y* axis is square-root scaled; *: non-overlapping 95% confidence intervals) for mutations of 21 known and potential driver genes in NS-LUAD (teal bars; $n$ = 656) and ES-LUAD (salmon bars; $n$ = 1,066), averaged across somatic genotypes. Inset: Background oncogenic mutation rates for these genes in ES-LUAD and NS-LUAD (dashed line: equality). (**B**) Differential expression of genes between *EGFR* wild-type (WT) and *EGFR* mutant samples in ES-LUAD ($n$ = 273) and NS-LUAD ($n$ = 59; log2FC: log2 of fold change in expression) for genes deemed to have a significant gene expression interaction between *EGFR* allelic status and smoking status. Color of each gene name denotes whether the change in expression was more positive (red) or more negative (blue) in NS-LUAD compared to ES-LUAD (*: $P < 0.05$, •: $P < 0.1$, *SOX21*: $P = 0.116$; background color: change in gene expression from normal ($n$ = 33) to LUAD ($n$ = 332) is increased [pink], decreased [light blue], or insignificant [gray]). Leftward squares designate whether the gene produces a transmembrane (TM) or extracellular matrix (ECM) protein as well as whether it facilitates or interacts with the Wnt, and TGF-β and the epithelial-mesenchymal transition (EMT) signaling pathways. (**C**) Interactions between proteins (red: implicated in EMT) whose *EGFR*-mutation-associated expression changes differ between ES-LUAD and NS-LUAD. (**D**) Sample-specific DNA replication expression scores for *KEAP1* WT and mutant ES- and NS-LUAD (**: $P < 0.01$).

We found that reported differences between ES- and NS-LUAD in the prevalences of key driver mutations [21,22,71] were often driven by differential selection, rather than solely by differential mutagenesis (**Fig. 2A**). Among the 21 driver genes analyzed, *TP53*, *EGFR*, *PIK3CA*, *SMAD4*, *MET*, *MGA*, and *GNAS* mutations experienced stronger positive selection in NS-LUAD. *STK11*, *KEAP1*, and *KRAS* mutations experienced stronger positive selection in ES-LUAD. *RB1*, *BRAF*, and *BRCA2* mutations experienced similar selection in both cancer types.

Most mutations that experienced significantly different selection between ES- and NS-LUAD were strongly selected in one context and weakly selected in the other, demonstrating pronounced context-dependence in LUAD evolution. In ES-LUAD, *KEAP1* mutations experienced the third strongest selection—after *TP53* and *KRAS*—and experienced over three times stronger selection than in NS-LUAD. Likewise, *STK11* mutations experienced over twice as much selection in ES-LUAD. Indeed, both *KEAP1* and *STK11* mutations are enriched in ever-smoker populations [23,72,73]. *EGFR* mutations exhibited the most dramatic difference:

they were more than twenty times more strongly selected in NS- than ES-LUAD, and were also the third most strongly selected mutations in NS-LUAD after *TP53* and *KRAS*. *EGFR* mutations are much more common in NS-LUAD than in ES-LUAD [21,74]. Similarly, *PIK3CA*, *SMAD4*, and *MET* mutations experienced over three times stronger selection in NS- relative to ES-LUAD.

To investigate mechanisms underlying the differential selection of *EGFR* mutations, we assessed their differential transcriptional effects in ES- and NS-LUAD using TCGA expression data. In fewer than one percent of genes did the expression fold change from *EGFR* wild type samples compared to *EGFR* mutant samples differ significantly between ES- and NS-LUAD (**Fig. 2B**). However, several of these genes exhibited markedly stronger upregulation or downregulation in EGFR-mutant NS-LUAD than in ES-LUAD. Genes that were upregulated were frequently components of Wnt, TGF-β, and Hedgehog signaling pathways or of the extracellular matrix (**Fig. 2B–C**). Indeed, in an analysis of gene sets upregulated in NS- versus ES-LUAD, the Hallmark gene set associated with epithelial-mesenchymal transition was the most enriched. On the other hand, *KEAP1* mutations were associated with a substantially higher DNA replication enrichment score in ES-LUAD but not in NS-LUAD (**Fig. 2D**). These examples relate microenvironmental context to modifications of the downstream effects of driver mutations and their accompanying selective advantage.

Extending our analysis from 21 to 1,200 LUAD-associated genes revealed a striking asymmetry in selection by smoking status: only six genes—*KRAS*, *KEAP1*, *STK11*, *PBRM1*, *DNMT3B,* and *CDH1*—were significantly more strongly selected in ES- than NS-LUAD), whereas 87 experienced stronger selection in NS-LUAD (**Fig. S1**). Among the mutations most strongly favored in NS-LUAD were those in *EGFR*, *SMAD4*, *SMAD2*, *PTEN*, *PIK3R1, PAK1*, *FGF3*, *GATA3*, *JUN*, *CTNNA2*, and *TLR4*. Mutations often experienced weaker selection in ES-LUAD, which is more heavily mutated and frequently possesses multiple known drivers, than in NS-LUAD, which often features a single known driver (**Fig. S1**).

Overall, somatic mutations often experienced substantially altered selection in the tobacco-smoke-altered ES-LUAD compared to NS-LUAD, contributing to differential evolution of the two cancers. Along with the somatic environment, the somatic genotype can also induce differential selection on specific mutations: effects of each mutation can be analyzed across somatic genotypes to reveal selective epistasis that shapes the evolutionary trajectories of tumors in ever-smokers and never-smokers. Therefore, we assessed the influence of pairwise selective epistasis.

## Pairwise epistasis shapes the adaptive landscape of ES-LUAD and NS-LUAD

To assess how frequently and how strongly the evolutionary trajectory of LUAD is affected by pairwise epistasis, we quantified the effects of driver mutations both in isolation (in tumors that were wild type at a paired gene) and in the presence of each paired gene mutation. For ES- and NS-LUAD, we evaluated all 420 possible gene pairs among the 21 driver genes: pairwise interactions were identified that modulate the selective advantage of mutations, and did so in a smoking-dependent manner.

### *Mutation of* TP53 *followed by* RB1 *is the most likely trajectory for ES-LUAD and NS-LUAD*

To investigate directional epistasis between two common LUAD drivers, *TP53* and *RB1*, we first stratified tumors by smoking status and somatic genotype at these two loci, and then estimated the fixation rates of each mutation from each somatic genotype (**Fig. 3A**). In ES-LUAD, *TP53* mutations exhibited a high fixation rate when *RB1* was wild-type but were inferred to exhibit a negligible fixation rate when *RB1* was mutated ($\lambda$**, Fig. 3B**). Conversely, *RB1* mutations exhibited a low fixation rate when *TP53* was wild-type but exhibited a high fixation rate when *TP53* was mutated. These asymmetries indicate the most likely evolutionary

sequence in ES-LUAD: mutation of *TP53* followed by mutation of *RB1*. The mutation rate underlying *TP53* variants observed in this dataset ($\mu = 7.5 \times 10^{-7}$) was over double that of *RB1* variants ($\mu = 2.7 \times 10^{-7}$; **Fig. 3B**). Accordingly, *RB1* variants in a *TP53*-mutant genotype must have experienced comparable selection to *TP53*-variants in a wild-type genotype ($\gamma$, **Fig. 3B**), despite the substantially lower fixation rate of the *RB1* variants. This pattern reflects synergistic epistasis: prior mutation of *TP53* increases selection for *RB1*. In contrast, the reverse order exhibited antagonistic epistasis—*TP53* was less strongly selected in the presence of *RB1* mutation—though this last effect did not achieve statistical significance.

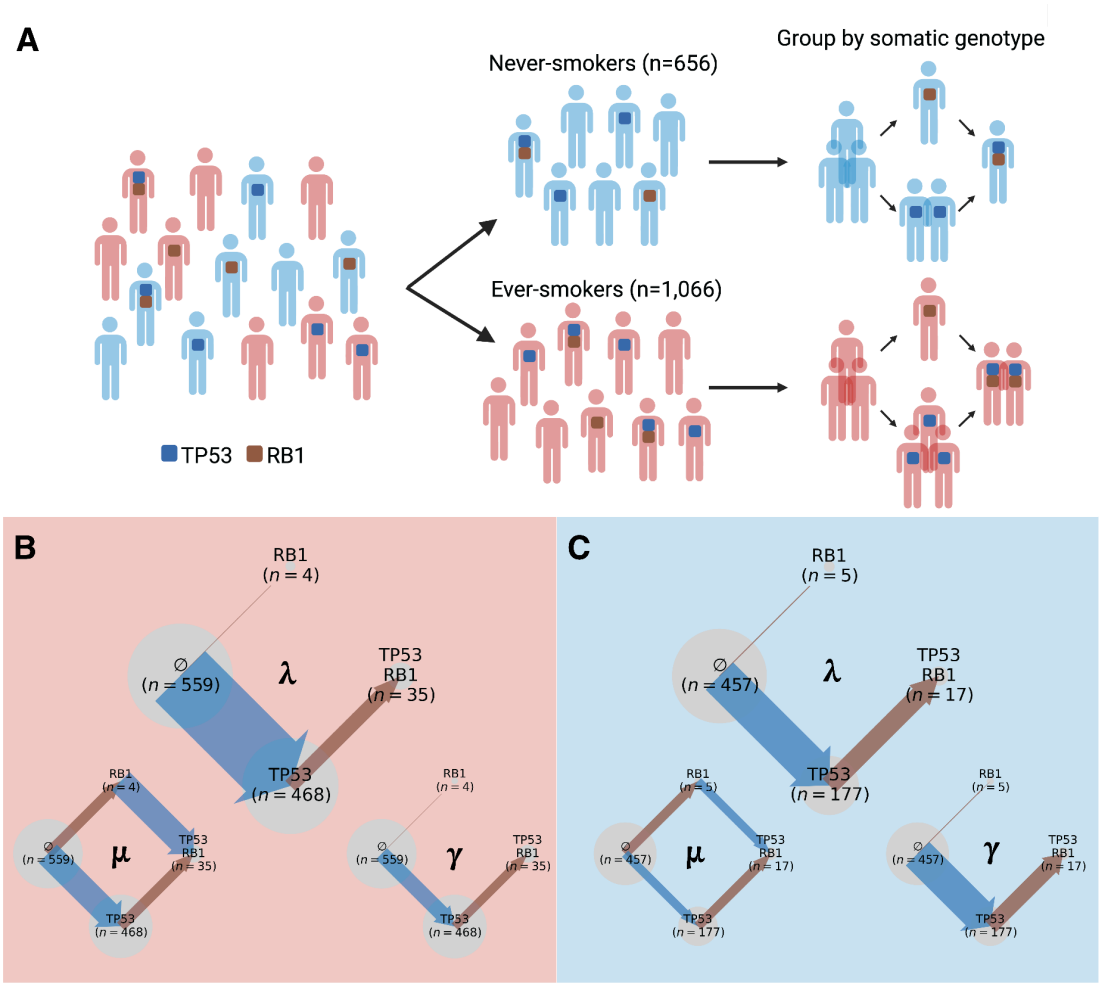


**Figure 3. Pairwise selective epistasis inferred within the evolutionary trajectories of lung adenocarcinoma (LUAD), depicted for *TP53* and *RB1*.** (**A**) Genomic sequence data from lung adenocarcinomas can be stratified by clinically- and genetically-determined patient smoking status (blue: never-smokers; red: ever-smokers) and somatic genotype. Analysis of (**B**) ever-smoker LUAD and (**C**) never-smoker LUAD deconvolves somatic genotype prevalences (gray circle radius) into genotype-specific fixation rates (λ; arrow width), oncogenic mutation rates (μ; arrow width), and strengths of somatic selection (γ; arrow width) on mutations of *TP53* (dark blue arrows) and *RB1* (brown arrows).

In NS-LUAD, the fixation rate for *TP53* mutations from a wild-type genotype was lower than in ES-LUAD, and the fixation rate for *RB1* mutations in a *TP53*-mutant background was higher ($\lambda$**, Fig. 3C**). The underlying mutation rates of *TP53* variants ($1.9 \times 10^{-7}$) and *RB1* variants ($2.0 \times 10^{-7}$) were nearly equal in NS-LUAD, in contrast to their disparity in ES-LUAD. Despite these differences, the inferred evolutionary trajectory mirrored that of ES-LUAD: the most likely ($\lambda$) and most adaptive ($\gamma$) path involved mutation of *TP53* followed by mutation of *RB1*. As in ES-LUAD, prior *TP53* mutation synergistically enhanced selection for *RB1* mutation, and prior *RB1* mutation antagonized selection for *TP53* mutation, without reaching statistical significance ($\gamma$, **Fig. 3C**).

### *Synergistic epistasis is widespread and often substantial in ES-LUAD*

Analysis of all pairwise combinations of the 21 reputed LUAD driver genes revealed frequent synergistic and antagonistic epistasis. In ES-LUAD, synergistic interactions were especially pronounced: for genes whose mutations were under strong positive selection (with the notable exceptions of *TP53* and *KRAS*) synergistic epistasis led to two- to six-fold increases in selection. For example, selection on mutations in *KEAP1, ATM*, *CTNNB1*, *PIK3CA*, and *ARID1A* was markedly enhanced in the presence of prior *TP53* or *KRAS* mutations (**Fig. 4A**). Additional strong synergistic interactions included those affecting mutants of *ARID1A* in the context of somatic genotypes with *ALK*, *ATM*, or *RBM10* mutations; *STK11* in the context of

mutated *KEAP1*; mutants of *BRAF* in the context of mutated *EGFR*; and mutants of *ATM* in the context of mutated *STK11*. Mutations in the non-*TP53* tumor suppressors *BRCA2*, *SETD2*, *RBM10*, *RB1*, *MGA*, and *APC*, as well as in the chromatin-remodeling genes *ARID1A* and *SMARCA4*, experienced moderate baseline selection that was frequently enhanced by the presence of diverse prior mutations. These genes never experienced antagonistic epistasis with each other and rarely with other mutations.

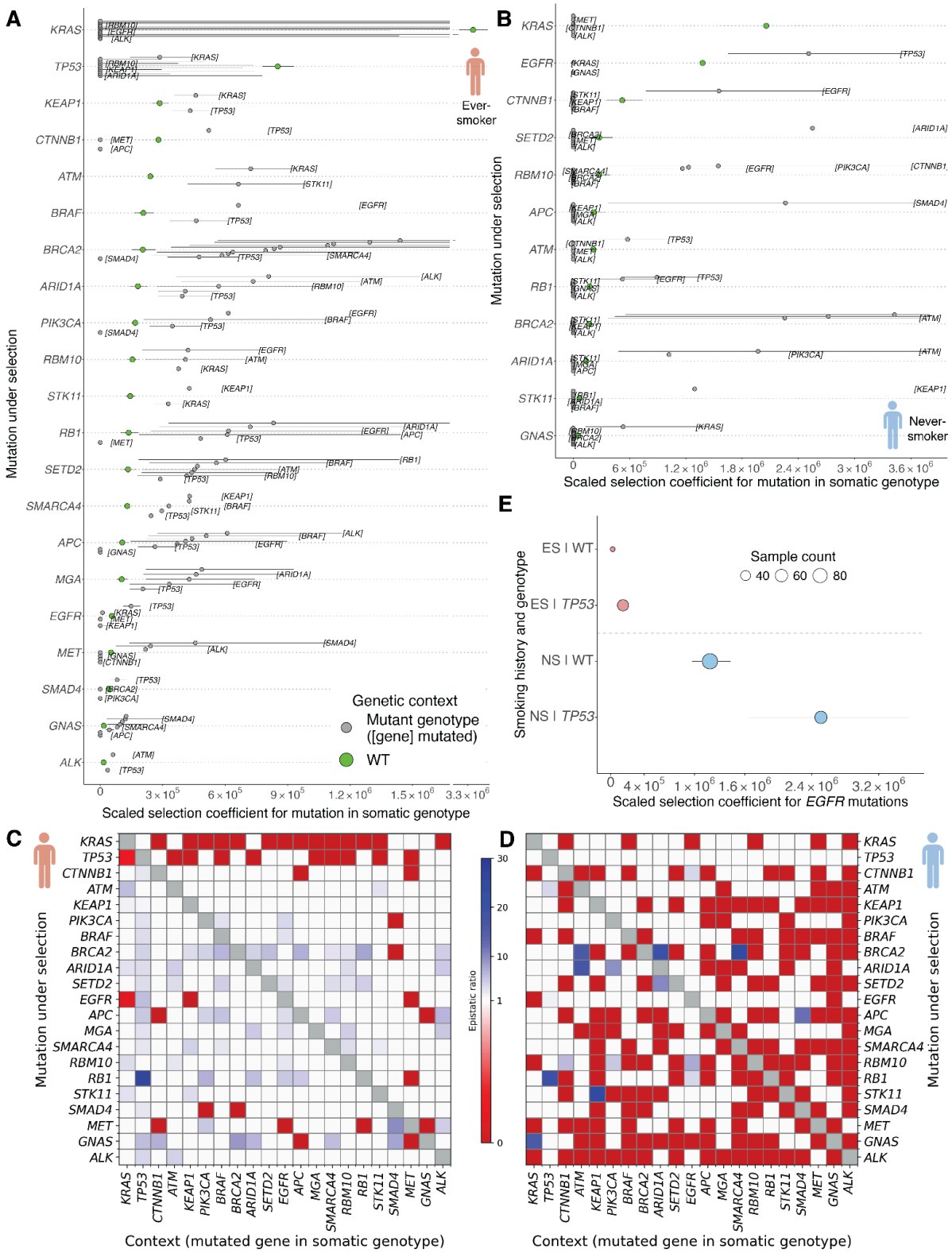

A
Mutation under selection
Ever-smoker
Genetic context
Mutant genotype ([gene] mutated)
WT
Scaled selection coefficient for mutation in somatic genotype
B
Never-smoker
Scaled selection coefficient for mutation in somatic genotype
E
Smoking history and genotype
ES | WT
ES | TP53
NS | WT
NS | TP53
Sample count
40 60 80
Scaled selection coefficient for EGFR mutations
C
Epistatic ratio
Context (mutated gene in somatic genotype)
D
Context (mutated gene in somatic genotype)

**Figure 4. Pairwise epistasis in ever-smoker (ES) and never-smoker (NS) lung adenocarcinoma (LUAD).** Strengths of somatic selection for mutations in wild-type tissue (green) and in previously mutated tissue (gray, labeled with previous mutation in brackets) in (**A**) ES-LUAD and (**B**) NS-LUAD. Significant pairwise epistatic interactions depicted in heatmap form for (**C**) ES-LUAD and (**D**) NS-LUAD based on non-overlapping confidence intervals for epistatic strengths of selection. Heatmap values are ratios of selection coefficients for mutations (in each row) from a somatic genotype containing a mutation in another gene (in each column) to the selection coefficient from a somatic genotype lacking the mutation. (**E**) Strengths of somatic selection for mutation of *EGFR* in *TP53*-mutant and *TP53*-wild-type somatic genotypes in ES-LUAD (red) and NS-LUAD (blue).

Patterns of epistasis involving oncogenes were heterogeneous. Notably, *KRAS* mutations showed no elevation of selection in the presence of any prior mutations in the 21-gene set. Conversely, selection for *GNAS* was enhanced by six prior mutations (**Fig. 4A**). For many oncogenes, selection was enhanced by prior *TP53* mutation, but was antagonized by mutations in other oncogenes such as *MET* and *EGFR*. Across all tested genes, the median effect size of statistically significant synergistic epistasis was a 4.3-fold increase in selection. The magnitude of this epistatic enhancement did not correlate with the strength of selection for the initiating mutation. However, it was inversely correlated with the selection for the secondary mutation in the wild-type somatic genotype ($r = -0.40$, $P < 0.001$), exemplifying a pattern in which weaker initial drivers are more strongly potentiated by the triggering of oncogenic transformation by prior mutations.

### *Antagonistic effects can reduce positive selection or induce negative selection on drivers*

In ES-LUAD, we identified eight gene pairs exhibiting statistically significant antagonistic sign epistasis, in which selection for a driver mutation shifted from positive to negative in the presence of another mutated gene. Mutations in *CTNNB1* and *GNAS* became selectively disadvantageous in the presence of *MET* and *APC* mutations. Similarly, selection for *SMAD4*

mutations was antagonized by prior *BRCA2* and *PIK3CA* mutations, and *MET* mutations were antagonized by mutations of *RB1*, *EGFR*, *CTNNB1*, and *GNAS*. In each of these eight cases, the reciprocal mutation order also induced negative selection (e.g., mutated *CTNNB1* induced negative selection for *MET* mutations). For over a hundred other ordered pairs, antagonistic sign epistasis was estimated, without reaching statistical significance. Antagonistic shifts in the magnitude of selection, in which selection for a mutation remained positive but was substantially reduced, were indicated in sixteen cases, of which two were significant: the presence of *KRAS* mutations caused a 73% reduction in selection for *TP53* mutations, and an 85% reduction in selection for *EGFR* mutations (**Fig. 4A**). Across all analyses of ES- and NS-LUAD, new mutations in *TP53* and *KRAS* were never inferred to experience stronger selection in the context of another driver mutation.

### *Antagonistic epistasis is more frequent in NS-LUAD*

Compared to ES-LUAD, NS-LUAD exhibited a higher frequency of antagonistic epistasis and fewer instances of synergistic epistasis (**Fig. 4B**). Of the 46 gene pairs exhibiting significant antagonistic epistasis in ES-LUAD, 16 also reached significance in NS-LUAD, and an additional 24 exhibited antagonistic trends without achieving significance (**Fig. 4C–D**). Thus, instances of selective antagonism inferred in ES-LUAD were nearly always present in NS-LUAD.

This consistency contrasted with a substantial inconsistency in selective synergies between ES- and NS-LUAD. Of the 72 gene pairs whose mutations exhibited significant synergistic epistasis in ES-LUAD, only eight were significantly synergistic in NS-LUAD. An additional 38 trended synergistic without reaching significance. Strikingly, 22 of these 72 instances were inferred to be significantly antagonistic in NS-LUAD, including many involving *BRCA2*, *GNAS*, and *SMARCA4*. For example, *KEAP1* mutation enhanced selection for *SMARCA4* mutation in ES-LUAD but antagonized it in NS-LUAD (**Fig. 4C–D**). These reversals

reveal distinct adaptive constraints imposed by the smoking-altered tumor microenvironment. Overall, antagonistic interactions were relatively stable across smoking status, whereas synergistic epistasis was highly contingent on smoking status.

Mutations in these 21 driver genes were subject to a broad range of epistatic and environmental interactions, as well as combinations thereof (**Fig. S2**). For example, *TP53* mutations synergistically increased selection for *EGFR* mutations in both ES- and NS-LUAD. However, the synergism was three times stronger in ES-LUAD (sixfold vs twofold increase; **Fig. 4E**). These context-dependent epistatic effects substantially contribute to the divergent trajectories of ES- and NS-LUAD.

Fewer examples of synergistic epistasis were observed in NS-LUAD compared to ES-LUAD. Nevertheless, synergistic epistasis contributed to a median ten-fold increase in selection on secondary mutations. Notable examples include enhanced selection for *EGFR* mutations following *TP53* mutation, for *CTNNB1* and *RB1* mutations after *EGFR* mutation, for *STK11* mutation after *KEAP1* mutation, and for *SETD2* mutation after *ARID1A* mutation (**Fig. 4B**). Even in relatively less mutated NS-LUAD, cooperative genetic interactions dramatically reshape the fitness landscape of evolving tumors.

### *Pairwise epistasis can be asymmetric in both sign and magnitude*

In both ES- and NS-LUAD, several gene pairs exhibited asymmetric epistasis, in which the direction or strength of interaction depended on the order of mutation acquisition. For instance, in ES-LUAD, mutation of either *KEAP1* or *STK11* enhanced selection on mutations of the other, but with markedly different magnitudes: primary *STK11* mutation was estimated to increase selection for a secondary *KEAP1* mutation by 60%, whereas primary *KEAP1* mutation significantly increased selection for a secondary mutation of *STK11* by over 300%. In other cases, the sign of epistasis reversed depending on mutational order. Most notably in ES-LUAD,

primary *KRAS* mutation synergistically increased selection for a secondary *KEAP1* mutation, but primary *KEAP1* mutation strongly antagonized selection for a secondary *KRAS* mutation (**Fig. 4A**). This asymmetry means that a mutation in *KRAS* followed by a mutation in *KEAP1* was advantageous, but a mutation in *KRAS* was disadvantageous or lethal when preceded by a mutation in *KEAP1*. This dramatic directional asymmetry and nine other such instances exemplify how somatic evolutionary trajectories can be constrained or potentiated by the specific sequence in which mutations occur.

## Higher-order epistasis shapes evolutionary trajectories of ES-and NS-LUAD

### KRAS*,* KEAP1*, and* STK11 *constitute an epistatically interacting triad in ES-LUAD*

Having identified *KRAS*, *KEAP1*, and *STK11* as experiencing strong, ever-smoker-specific selection, we next investigated whether their interactions extended beyond pairwise epistasis to form a higher-order epistatic network. Mutations in all three genes exhibited substantial fixation rates from nearly any initial somatic genotype, particularly en route to the triple-mutant genotype. Fixation rates for *KRAS* mutations were estimated to be very low when arising after *KEAP1* or *STK11* mutations alone compared to when arising as the initial mutation or in the *KEAP1-STK11* double-mutant (**Fig. 5A**). Nevertheless, these three genes form an epistatic triad in which mutations in any one tend to increase selection for the others.

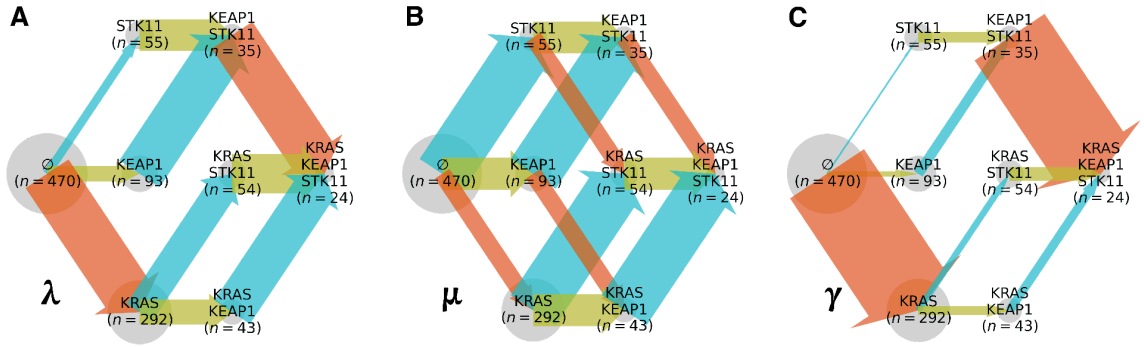



**Figure 5. Evolutionary trajectory for *KRAS*, *KEAP1*, and *STK11* mutations in ever-smoker lung adenocarcinoma.** (**A**) Fixation rates, (**B**) oncogenic mutation rates, and (**C**) scaled selection coefficients for mutations of *KRAS* (orange arrows; arrows width proportional to rate or coefficient), *KEAP1* (light blue arrows), and *STK11* (yellow arrows) conditioned on somatic genotype (gray circles; radius proportional to genotype prevalence).

Among the three genes, *STK11* manifested the highest oncogenic mutation rate, followed by *KEAP1*, and then *KRAS* (**Fig. 5B**). Given these differential underlying rates of mutation, we computed the consequent scaled selection coefficients for mutations in each genotype in the three-gene model, and compared selection between genotypes to quantify pairwise and higher-order selective epistasis. Many tumors possessed mutations in at least one of these genes (**Fig. 5C**, second column of nodes), providing high power to detect pairwise epistasis. Indeed, our three-gene model recapitulated the pairwise epistatic relationships inferred from the two-gene models (**Fig. 4C**). However, concurrent mutations of two or more of these genes were relatively infrequent (**Fig. 5C**, third and fourth columns of nodes), limiting power to detect higher-order epistasis with statistical confidence. For example, substantial uncertainty in the selection estimate for *KRAS* mutations in tumors harboring *STK11* and *KEAP1* mutations precluded a definitive inference of higher-order interactions. In contrast, because of the sufficiently high prevalence of *KRAS-KEAP1*- and *KRAS-KEAP1-STK11*-mutant genotypes, we were able to infer higher-order epistatic effects on *STK11* with greater confidence.

*STK11* mutations experienced significantly stronger selection in the presence of either *KRAS* or *KEAP1* mutations. If the synergistic effects of *KRAS* and *KEAP1* mutations on *STK11* compounded additively, then *STK11* mutations would experience substantially stronger selection in a *KRAS-KEAP1*-mutant genotype. However, selection for *STK11* mutations in a *KRAS-KEAP1* double-mutant genotype ($3.9 \times 10^5$) was comparable to that in a *KEAP1* or *KRAS* single-mutant genotype ($4.4 \times 10^5$ and $2.8 \times 10^5$; **Fig. 5C**). Accordingly, the compound synergistic effect of concurrent *KEAP1* and *KRAS* mutations was substantially less than the product of their disjoint synergistic effects, representing an example of sub-additive higher-order epistasis.

*Most pairwise epistatic effects compound without strong higher-order interaction*

The substantially sub-additive epistasis affecting mutations of *STK11* demonstrates that assessments of higher-order epistasis will be necessary to thoroughly understand cancer evolution. However, most effects of driver mutations compounded in a manner consistent with minimal higher-order epistasis. Examining all 1330 triads out of the 21 genes, we compared selection for a new mutation in single- versus double-mutant backgrounds. In the vast majority of cases, a combination of pairwise epistatic effects was sufficient to explain significant differences in selection for a new mutation between single- and double-mutant genotypes. For all cases in which selection for a secondary mutation in one gene was epistatically altered by presence of a primary mutation in a second gene and was unaffected by a primary mutation in a third gene, selection on the secondary mutation remained commensurately epistatically altered when both other variants of the triad were concurrently present. For example, in ES-LUAD, selection for *APC* mutations was synergistically enhanced by *SMARCA4* mutations but unaffected by *KRAS* mutations, and correspondingly, selection for mutations in *APC* was synergistically increased in a *SMARCA4-KRAS*-mutant genotype (**Fig. 6A,B**). Similarly, in

NS-LUAD, selection for *RB1* mutations was synergistically increased in both a *TP53*-mutant and *TP53-EGFR*-mutant genotype (**Fig. 6C**). Overall, pairwise interactions typically dominate the adaptive landscape of LUAD, but higher-order epistasis can have substantial influence on specific evolutionary trajectories.

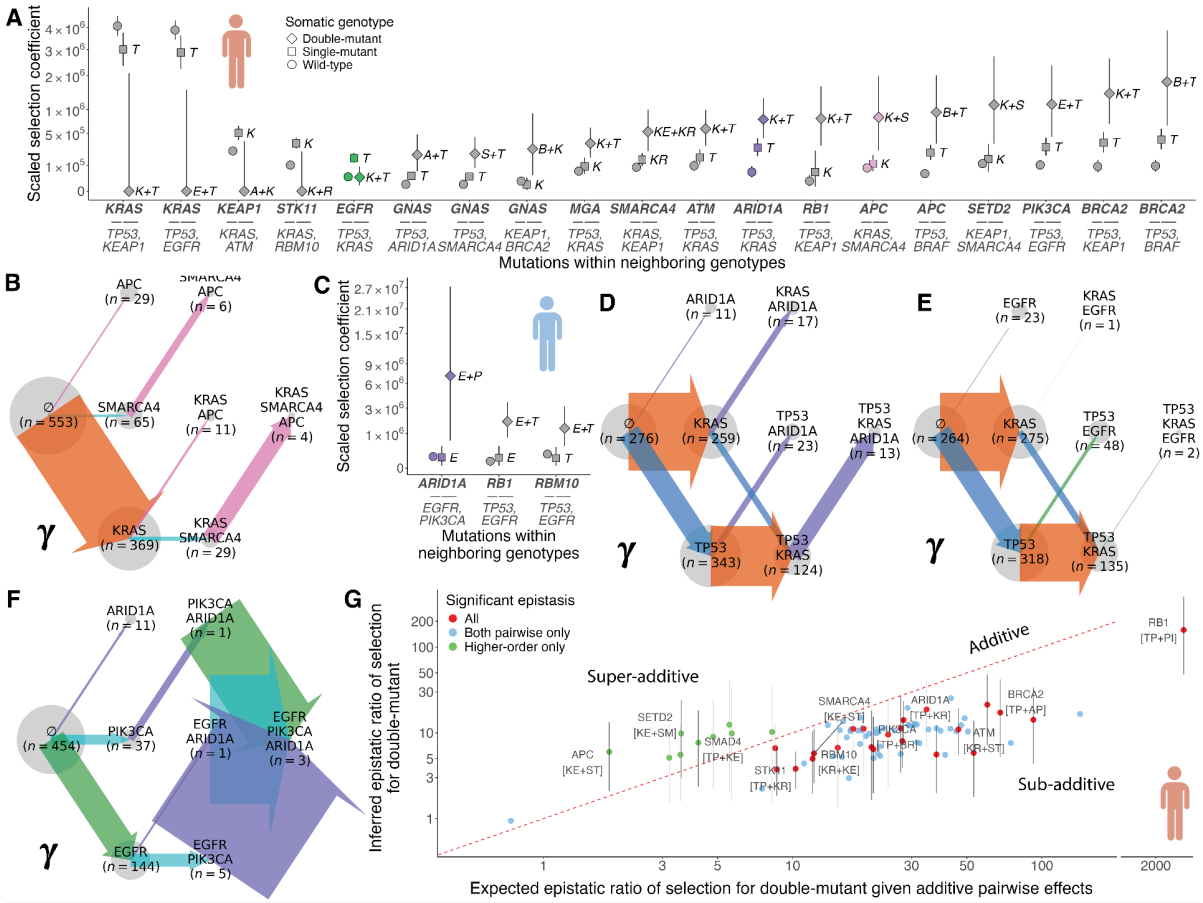


**Figure 6. Higher-order selective epistasis in lung adenocarcinoma (LUAD).** (**A**) Selection within gene triads (*y* axis is square-root scaled), for a mutation (bold font) in a context of no co-occurring mutations (circle), one co-occurring mutation (square; labeled with first letter(s) of the gene(s) indicated in the tick label that are mutated), and two co-occurring mutations (diamond) in ever-smoker (ES) LUAD [pink: visualized in panel B**;** purple: visualized in panel D; green: visualized in panel F]. (**B**) In ES-LUAD, the synergistic effect of *SMARCA4* mutations (light blue arrows) on selection for *APC* mutations (pink arrows; arrow width proportional to scaled selection coefficient; circle radius proportional to somatic genotype prevalence) does not diminish in the presence of a neutrally-interacting *KRAS* mutation (orange arrows). (**C**) Selection within gene triads in never-smoker (NS) LUAD [purple: visualized in (**F**); *y* axis is square-root scaled, symbols are as in panel A]. (**D**) Pairwise and higher-order epistasis in ES-LUAD: co-mutation of *TP53* (blue) and *KRAS* (orange), each of which independently increase selection for *ARID1A* mutations (purple), produces an even stronger synergy for *ARID1A* mutations; and (**E**) antagonistic effect of *KRAS* mutations (orange) on selection for *EGFR* mutations (green) outweighs the synergistic effect of *TP53* mutations (blue). (**F**) In NS-LUAD, co-mutation of *EGFR* (green) and *PIK3CA* (light blue) synergistically increases selection for *ARID1A* mutations (purple) despite neutral pairwise effects. (**G**) Comparison of the epistatic effect of double-mutant-genotype (label: bottom row, two-letter abbreviation) on selection for new mutation (label: top row) to the "expected" product of pairwise epistatic effects (blue points: both pairwise epistatic effects are statistically significant, but not the higher order effect; green: the higher-order effect is significant, but neither pairwise effect is; red: all effects are significant)

When two co-occurring mutations each independently enhanced selection for a third mutation, the synergistic effects often compounded. For instance, *ARID1A* mutations experienced increased selection in both KRAS- and *TP53*-mutant genotype. These synergistic effects compounded, resulting in mutations of *ARID1A* experiencing nearly three times stronger selection in a *KRAS-TP53*-mutant genotype than in a *TP53*-mutant genotype and fourteen times stronger selection than in a wild-type genotype (**Fig. 6D**). Similar patterns of compounded synergy were evident for *BRCA2* mutations in *KEAP1-TP53*- and *BRAF-TP53*-mutant genotypes, as well as for APC mutations in a *BRAF-TP53*-mutant background. Across cases, compounded synergy frequently drove selection to levels up to an order of magnitude higher than in genotypes lacking either partner mutation. This substantial compounding constitutes a

crucial element of how driver mutations accumulate in cancer genomes, positively reinforcing progression of cancer toward more aggressive forms.

### *Higher-order epistasis often manifests as sub-additive or emergent synergistic interactions*

In the case of compounding synergistic effects, it is also possible to observe super-additive higher-order epistasis, in which selection for a mutation in the presence of two independently synergistic mutations is even larger than what would be expected from the product of their independent synergistic effects. Indeed, some cases are consistent with either additive synergy or super-additive epistasis, such as mutations of *BRCA2* in a *KEAP1*-*TP53*-mutant genotype in ES-LUAD (**Fig. 6A**). More frequently, cases are consistent with either additive synergy or sub-additive epistasis, in which selection for a mutation in the presence of two independently synergistic mutations is smaller than expected from the sum of synergistic effects. Indeed, in all 73 cases wherein a new mutation arose in a genotype of two significantly synergistic mutations, estimates of selection for the new mutation were lower than would be expected from purely additive synergistic epistasis. Such widespread sub-additivity pervades the adaptive landscape, tempering rather than amplifying selective pressures, even in the context of strong pairwise synergy.

We also identified cases of higher-order epistasis involving the co-occurrence of synergistic and antagonistic partner mutations, as well as the emergence of synergy in the co-occurrence of neutrally-interacting partner mutations. In ES-LUAD, for example, selection for *EGFR* mutations was synergistically enhanced by prior *TP53* mutations but antagonized by *KRAS* mutations. In a *TP53*-*KRAS*-mutant genotype, *EGFR* mutations experienced significantly less selection than in a *TP53*-mutant *KRAS*-wildtype genotype (**Fig. 6A,E**). In fact, these *EGFR* mutations experienced levels of selection similar to those they would experience in a genotype

lacking either mutation, indicating a neutralizing result of the combination of synergistic and antagonistic epistasis. Similarly, in ES-LUAD, selection for *STK11* mutations was synergistically increased by *KRAS* mutations and antagonized by *RBM10* mutations. *STK11* mutations experienced significantly weaker selection in a *KRAS*-*RBM10*-mutant genotype than in a *KRAS*-mutant *RBM10*-wildtype genotype (**Fig. 6A**). Thus, in a combination of synergistic and antagonistic epistasis, the antagonistic effect tended to dampen, or potentially even overwhelm, the synergistic effect.

Another form of higher-order epistasis is the emergence of synergy or antagonism from a combination of mutations, each of which does not individually exhibit pairwise epistasis with the affected gene. We observed no instances of emergent antagonism. We did, however, observe at least one instance in which selection for a mutation was significantly increased by the co-occurrence of two mutations that did not individually have synergistic effects: in NS-LUAD, selection for mutations in *ARID1A* was not significantly altered when only one of the two genes was mutated, but was increased by an order of magnitude in an *EGFR*-*PIK3CA*-mutant genotype (**Fig. 6A**,**F**). Among multiple other gene triads, we detected similar instances of emergent synergy, such as for *APC* mutations in a *KEAP1*-*STK11*-mutant genotype (**Fig. 6G**): in this instance, significance was reached for synergy when compared to wild type but not when compared to the *KEAP1*- or *STK11*-mutant genotypes, which were themselves epistatically neutral.

Overall, we observed two major classes of higher-order epistasis: sub-additive synergy and emergent—and consequently super-additive—synergy. In particular, when both component epistatic effects were strongly synergistic, their combination was always sub-additively synergistic (**Fig. 6G**). On the other hand, when both component epistatic effects were neutral or insignificant, a super-additive synergistic effect occasionally emerged from their combination (**Fig. 6G**). Interestingly, as the strength of the component epistatic effects increases, the inferred

epistatic effect of the double-mutant combination rapidly plateaus at an approximately ten to twenty-fold increase in selection.

In this survey of the adaptive landscape of LUAD, we discovered several cases of higher-order epistasis that substantially influence the selective advantage of oncogenic mutations. Over a third of LUAD samples in ever-smokers possessed mutations in three or more of 21 tested genes—a proportion that would likely rise in an analysis of a greater number of driver genes. In these patients, higher-order epistasis may play a substantial role in the evolutionary trajectories of their tumors and their responses to targeted therapies. Indeed, we found that a wide variety of adaptive trajectories ensued from the combined epistatic effects of multiple mutations. Some of these trajectories cannot be predicted from pairwise analyses. However, our large-scale analysis of the adaptive landscape of 21 mutations has significantly supported only a subset of the many potential instances of higher-order epistasis between oncogenic mutations in LUAD. Indeed, multiple other instances of higher-order epistasis were estimated to be present. Their effect would be clarified with greater sample sizes.

## Discussion

Here we have characterized the distinct adaptive landscapes of lung adenocarcinoma in ever-smokers and never-smokers by first estimating gene-specific background mutation rates in ES- and NS-LUAD, then comparing the intensity of somatic selection and selective epistasis acting on those mutations between the two groups. Consistent with the established mutagenic effects of tobacco carcinogens [4], we detected elevated oncogenic mutation rates in all tested genes in ES- compared to NS-LUAD. Accounting for these background rates revealed that mutations in genes such as *EGFR*, *PIK3CA*, and *MET* provide a stronger selective advantage in NS-LUAD, while mutations in *KRAS*, *KEAP1*, and *STK11* provide a stronger selective

advantage in ES-LUAD. Focusing on *EGFR* and *KEAP1*, we found that these adaptive differences were at least partially explained by the differential effects of these mutations on epithelial-mesenchymal transition and DNA replication transcriptional program activity in ES- and NS-LUAD.

By quantifying the selective advantage of these mutations within specific somatic genotypes, our analysis recovered known epistatic interactions—such as the enhancement of selection on *RB1* and *EGFR* mutations by prior *TP53* mutations—and revealed context-specific interactions, such as the synergistic effect of *EGFR* mutations on *CTNNB1* mutations uniquely in NS-LUAD. We also detected substantial asymmetries: *CTNNB1* mutations did not impact *EGFR* mutations in ES- or NS-LUAD. Notably, we found more frequent synergy in ES-LUAD and substantially more pervasive antagonism in NS-LUAD, constituting fundamental differences in the evolutionary navigability of their adaptive landscapes. Finally, we showed that higher-order epistasis—particularly sub-additive synergistic and emergent synergistic epistasis—is frequent in and substantially guides the evolutionary trajectories of cancer. These findings of differential selection and epistasis reveal that exposure to tobacco smoke reshapes the adaptive landscape of LUAD not only by elevating mutation burden, but also by altering the selective value of mutations and their epistatic interactions—highlighting opportunities to leverage these adaptive differences for biomarker development and therapeutic stratification.

The prevalences of driver mutations have been reported to substantially differ between smokers and never-smokers in LUAD [21,22] as well as in non-small cell lung cancer [NSCLC; 23,24]. However, previous studies have either refrained from explaining these differences or attributed them primarily to tobacco-induced mutagenesis [22,23]. Our results show that they often cannot be explained by differential mutagenesis alone. For example, *KRAS* variants are over five times more frequent in ES- than NS-LUAD tumors, but the background, neutral rate at which *KRAS* mutations occur is only about three times higher in ES-LUAD. To resolve this

discrepancy, *KRAS* mutations must be nearly twice as oncogenic in ES- compared to NS-LUAD. In general, the differential prevalence of these driver mutations is often not solely attributable to tobacco-induced mutagenesis, but also due to their altered selective advantage in the tobacco-altered lung environment. This physiological-adaptive effect of smoking is crucial to account for: changes in the selective advantage of a mutation suggest altered functional impact within a given tissue environment, with direct implications for the efficacy of targeted therapies and the interpretation of biomarkers in smokers versus non-smokers.

Multiple perspectives have proposed that the adaptive landscape of oncogenesis is altered by physiological insults from endogenous and exogenous factors [14,17,75,76]. However, data-driven evidence for this hypothesis has remained limited [cf. 77]. Here, we provide compelling evidence that canonical driver mutations confer altered selective advantages in the smoking-affected lung microenvironment of ever-smokers with LUAD. These shifts in selection are linked to context-dependent phenotypic effects of mutations. For example, *EGFR* mutations are associated with greater activity of the epithelial-mesenchymal transition (EMT) transcriptional program in NS-LUAD, but not in ES-LUAD—likely explaining their stronger selective advantage in never-smokers. Such divergent phenotypic consequences can be attributed to smoking-induced alterations to transcriptional program activity [78–80], tumor microenvironment composition [81], and cell fitness. These findings provide a mechanistic basis for how exogenous exposures can reshape the evolutionary trajectory of cancer by altering the adaptive value of oncogenic mutations.

In the case of *TP53*, Rodin and Rodin [13] reported comparable background mutation rates between smokers and nonsmokers with lung cancer, and attributed the higher frequency of *TP53* mutations in smokers to increased selection in the smoking-altered microenvironment. By contrast, our analysis—leveraging a substantially larger dataset, incorporating tissue- and context-specific covariates, and applying advanced statistical methodologies to estimate

background mutation rates [6]—identified a markedly elevated *TP53* mutation rate in ES-LUAD. With this substantial differential mutagenesis accounted for, our analysis revealed that *TP53* mutations are actually under weaker positive selection in ever-smokers than in never-smokers. This finding challenges earlier conclusions and underscores the importance of accurately resolving both mutational processes and selection when interpreting variant prevalence across environmental exposures.

Despite mounting evidence of the substantial biological impact of tobacco smoke, clinical trials for lung cancer therapies often do not assess the influence of smoking status—and frequently do not even collect usage history [82], perhaps due to a conception that tobacco smoke has a negligible effect on outcomes [83]. However, our findings reveal that tobacco smoke profoundly reshapes the evolutionary trajectory of lung adenocarcinoma, necessitating more systematic incorporation of smoking history into trial design and treatment decision-making, particularly for targeted therapies. For example, the substantially stronger selection for *EGFR* mutations in NS-LUAD implies that EGFR tyrosine kinase inhibitors (TKIs) should yield greater clinical benefit in never-smokers. Indeed, a retrospective analysis of patients with *EGFR*-mutant LUAD found that the objective response rate (ORR) of patients receiving EGFR TKIs declined from 73% among never-smokers to 46% among individuals with over 10 pack-years of smoking history [84]. Similar patterns have been reported in additional retrospective studies of LUAD [85] and across NSCLC more broadly [86,87]. This concordance between selection-based predictions and observed outcomes highlights the translational value of selection coefficients as predictors of targeted therapy efficacy. By identifying likely responders based on environmental exposures that shape the tumor microenvironment, exposure-informed selection coefficients can improve clinical trial stratification, enhance statistical power, and refine treatment personalization in precision oncology.

Prior studies have questioned whether computational analyses can reliably detect epistasis, citing concerns about confounding by subtype-specific mutation profiles, variation in mutation burden, and limited statistical power [31,88]. To address these concerns, we analyzed ES- and NS-LUAD independently and comparatively, modeled subtype- and gene-specific background mutation rates, estimated effect sizes of mutations in specific somatic genotypes instead of suboptimal measures of co-occurrence or mutual exclusivity, and integrated data from multiple cohorts to ensure sufficient statistical power. Consequently, we were able to make robust inferences of both pairwise and higher-order selective epistasis, many of which align with experimental and clinical evidence. For example, consistent with our finding that *KRAS* mutations synergistically enhance selection for *KEAP1* and *STK11* mutations, prior studies show that *KEAP1* loss accelerates LUAD progression in *KRAS*-mutant mouse tumors [89,90] and compensates for the deleterious effects of *KRAS*-*STK11*-co-mutation [91], and mutations of *KEAP1* and/or *STK11* worsen survival and treatment response in *KRAS*-mutant LUAD [92,93] and NSCLC [94–99]. Furthermore, in accordance with the inferred mutually synergistic effects of mutations of *KEAP1* and *STK11* in ES-LUAD, co-mutation of *KEAP1* and *STK11* promote cell proliferation [100] and worsen survival in NSCLC independently of *KRAS* [101,102].

Our analyses also identified synergies between *EGFR* and other drivers: *BRAF* in ES-LUAD, *CTNNB1* in NS-LUAD, and *RB1* in both. These interactions are corroborated by experimental and clinical studies showing that *BRAF* mutations confer resistance to EGFR TKIs [103], that *CTNNB1* mutations are enriched in advanced *EGFR*-mutant NSCLC and associated with metastasis and TKI resistance [104], and that *RB1* mutations in *EGFR*-mutant LUAD mediate histologic transformation to small-cell lung cancer [105,106]. We also revealed that mutations of *TP53* and *KRAS* were widely antagonized by other driver mutations, consistent with their early, clonal emergence during LUAD evolution [50,107] and supporting their frequently primary role in tumor initiation. This panoply of alignments between our results and

experimental and clinical evidence encourages deeper investigation into the many epistatic relationships we have revealed that remain untested in the laboratory or clinic. They represent fertile ground for mechanistic investigation and promise new therapeutic targets and biomarker candidates.

Previous studies have shown that epistatic interaction networks vary across cancer types in humans [108] and across environmental contexts in other organisms [109–114]. However, extensive differences in selective epistasis within a single cancer type exposed to distinct tissue environments have not previously been demonstrated. Here, we have shown that tobacco smoke exposure markedly reshapes the selective epistatic interactions underlying LUAD evolution. Broadly, synergistic epistasis was more frequent in ES-LUAD, whereas antagonistic epistasis was more frequent in NS-LUAD. Indeed, in some cases, epistasis between driver genes was synergistic in ES-LUAD, but antagonistic in NS-LUAD. In other cases, the magnitude of synergy differed across exposures: synergy for mutation of *TP53* in *EGFR*-mutant tumors was stronger in ES- than in NS-LUAD. The higher frequency of synergy and lower frequency of antagonism in ES-LUAD suggests a more navigable adaptive landscape—characterized by a higher frequency of beneficial mutational combinations and fewer evolutionary barriers—than in NS-LUAD. This increased accessibility enables ES-LUADs to develop along a greater diversity of evolutionary trajectories, contributing to increased interpatient heterogeneity. Moreover, it may also help to explain why tobacco smoke elevates the risk of lung cancer to such an exceptional degree. In NS-LUAD, widespread antagonistic epistasis may reflect a greater fitness cost of mutation accumulation, possibly due to a more immunologically reactive tumor microenvironment [12,71,81,115] or due to a lower tolerance for mutation accumulation in the non-dysregulated lung tissue cells of never-smokers. Taken together, these findings reveal that environmental exposures modulate selective epistasis within a cancer type, suggesting the

importance of exposure history to predicting tumor evolutionary trajectories and to guiding epistasis-informed treatment strategies.

It has often been assumed that the adaptive landscape of cancer can be adequately described without epistasis [6,116–120] or with only pairwise epistasis [31,32,121]. However, our results demonstrate that not only pairwise but also higher-order epistasis—particularly sub-additive and emergent epistasis—play a crucial role in modulating the adaptive value of cancer drivers. For instance, in NS-LUAD, selection for *ARID1A* mutations increased by an order of magnitude in an *EGFR-PIK3CA*-mutant genotype, despite showing no significant pairwise synergy with either mutation alone. This emergent epistasis, arising only in the triadic context, would be missed by analyses restricted to pairwise comparisons, necessitating consideration of multi-mutant genotypes in both computational and experimental studies. Moreover, all significant pairwise synergistic effects combined sub-additively in triads, rapidly reaching a plateau of synergistic benefit. This pattern aligns with the diminishing-returns epistasis observed in asexual populations [44–47]. Overall, higher-order synergistic epistasis biases tumors toward some evolutionary trajectories, while higher-order antagonism prunes most trajectories. Both make tumor evolution more predictable. Thus, incorporation of higher-order epistasis into models of tumorigenesis is essential to understanding and guiding the evolutionary paths that tumors take under therapy.

Current NSCLC treatment guidelines for EGFR TKIs do not account for co-occurring mutations. However, our findings suggest that concurrent mutations could serve as biomarkers of treatment response. Antagonistic epistasis between *EGFR* mutations and mutations in *KRAS* or *KEAP1* implies that the efficacy of EGFR TKIs may be reduced in tumors harboring these concurrent alterations. This prediction aligns with clinical observations: *KRAS* mutations decrease survival in patients with *EGFR*-mutant LUAD patients treated with EGFR TKIs [122], and *KEAP1* mutations reduced EGFR TKI efficacy in *EGFR*-mutant lung cancer [123,124]. More

broadly, antagonistic epistasis is a signal of synthetic lethality [32]. Our results also demonstrate mutual negative selection between *BRCA2* and *SMAD4* loss-of-function mutations—suggesting that pharmacological inhibition of *SMAD4* could be therapeutically beneficial in *BRCA2*-mutant tumors, in analogy to usage of PARP inhibitors in *BRCA1/2* deficient tumors. In general, selective epistasis can clarify the frequent heterogeneity of response to targeted therapies [37,125], and can improve stratification of responders and nonresponders to targeted therapy in clinical trials [126,127], enhancing trial power and therapeutic precision.

Because *TP53* mutations synergize with mutations of *EGFR*, we would expect that in tumors with concurrent *EGFR* and *TP53* mutations, EGFR TKIs would be especially effective because they would both eliminate the baseline effect of the *EGFR* mutation and diminish the effect of the concurrent *TP53* mutation that is mediated through the *EGFR* mutation. However, *TP53* mutations are consistently associated with resistance to EGFR TKIs in *EGFR*-mutant LUAD [128–130] and NSCLC [131–135]. This contradiction may be explained by the pleiotropic effects of *TP53* loss, including increased somatic nucleotide mutation rates [130], widespread copy-number alterations [128,136], and dysregulation of gene expression programs linked to resistance [137] and progression [136]. These pleiotropic effects provide greater genetic variation that facilitates the evolution of resistance independently of the selective synergy between *TP53* loss and *EGFR* mutation.

Previous computational studies of pairwise interactions have demonstrated asymmetries in the magnitude of pairwise epistasis [cf. 30]. Our analysis has revealed not only epistasis of asymmetric magnitudes, but also asymmetric signs of selection, in which one ordering of the mutations is advantageous while the reverse is disadvantageous. These directional asymmetries likely arise from cellular or stable transcriptional transitions initiated by the primary mutation that unidirectionally alter the selective context for subsequent alterations—an effect not captured by conventional analyses of mutual exclusivity or co-occurrence. Indeed, the presence

of such asymmetries of epistasis are evident in established genetic models of tumor evolution, such as the *APC* → *KRAS* → *TP53* mutation sequence in colorectal cancer [138,139], whose strongly preferential ordering requires directional epistasis. Asymmetric epistasis has also been supported by genetic and computational genomic studies of *TP53* and *CCNE1* mutations [32], *TP53* and *RB1* mutations [30], and *JAK2* and *TET2* mutations [140]. The presence of directional epistasis implies preferential orders of mutation acquisition that offer new opportunities to predict and potentially redirect cancer evolution.

This study was subject to several limitations. First, its retrospective nature introduces the possibility that covariates of smoking behavior—such as comorbid health states, environmental exposures, inherited genetics, and sex biases—may contribute to the observed differences in the adaptive landscapes of ES-LUAD and NS-LUAD. However, none of these factors are known to exert a physiological impact on the lung that is comparable to smoking tobacco. Accordingly, the selective differences we report are most parsimoniously attributed to the direct and downstream effects of smoking. Second, smoking status in our study was primarily inferred from mutational signatures in tumor sequences, which can be susceptible to misclassification. However, this method has been validated for classification of smoking status in NSCLC and is believed to outperform self-reported smoking histories [21,25,141]; self-reported status has been found to be unreliable [142,143] and often does not account for the risk conferred by secondhand smoke [144]. Furthermore, because our approach explicitly quantifies the smoking-associated mutational burden in the lung periphery—where LUAD develops—it also implicitly identifies patients who have experienced significant injury from smoking in the regions where LUAD originates. Third, our estimates of selective epistasis assumed that oncogenic mutation rates were independent of the somatic genotype. This assumption may be violated in cases where mutations that are known to increase mutation rate, such as those in *TP53* and *BRCA2* [145,146] occur early; such increases in mutation rate potentially inflate estimates of

selection later in tumorigenesis. This effect warrants further investigation, but is unlikely to systematically bias our comparisons between ES- and NS-LUAD. Finally, our analysis was restricted to somatic SNVs and did not consider indels, copy number alterations (CNAs), or other structural variants (SVs). CNAs are common in cancer [147] and pairwise selective epistasis has been indicated between SNVs and CNAs and other SVs [148,149] or between SVs [37,108,150]. Unfortunately, accurate background mutation rates are currently unavailable for SVs. As such rates become calculable, integrated analysis of SNVs and SVs will reveal new epistatic relationships and clarify the somatic adaptive landscape of cancer.

## Conclusion

We have demonstrated that the adaptive landscapes of ES- and NS-LUAD diverge substantially, implicating tobacco smoking not only as a mutagen but also as a modulator of somatic selection. Our analysis also shows that symmetric and asymmetric selective epistasis among driver mutations is both substantial and pervasive in lung adenocarcinoma and generates the distinct evolutionary trajectories of ES-LUAD and NS-LUAD. Furthermore, we provide a systematic characterization of higher-order selective epistasis in LUAD, revealing its underrecognized role in shaping tumor evolution. Together, these results show that the contribution of mutations to tumor growth is profoundly context-dependent—modulated by both tumor somatic genetics and the environmental exposures it has experienced. This insight illuminates new opportunities for precision oncology through the integration of somatic genotype and exposure history into treatment strategies. Continued investigation into epistatic, environmental, and even gene-by-gene-by-environment interactions in lung adenocarcinoma and other cancers will be essential for the advancement of our understanding of cancer evolution and the improvement of patient-specific therapeutic strategies.

## Abbreviations

LUAD: Lung adenocarcinoma

ES: Ever-smoker

NS: Never-smoker

SNV: Single nucleotide variant

NSCLC: Non-small cell lung cancer

TKI: Tyrosine kinase inhibitor

ORR: Objective response rate

## References


1. HHS. How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-attributable Disease. U.S. Department of Health and Human Services; 2010.

2. Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, et al. Mutational signatures associated with tobacco smoking in human cancer. Science. 2016;354:618–22.

3. Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, Hainaut P. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. Oncogene. 2002;21:7435–51.

4. Yoshida K, Gowers KHC, Lee-Six H, Chandrasekharan DP, Coorens T, Maughan EF, et al. Tobacco exposure and somatic mutations in normal human bronchial epithelium. Nature. 2020;578:266.

5. Dasari K, Somarelli JA, Kumar S, Townsend JP. The somatic molecular evolution of cancer: Mutation, selection, and epistasis. Prog Biophys Mol Biol. 2021;165:56–65.

6. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal patterns of selection in cancer and somatic tissues. Cell. 2017;171:1029–41.e21.

7. Cannataro VL, Kudalkar S, Dasari K, Gaffney SG, Lazowski HM, Jackson LK, et al. APOBEC mutagenesis and selection for NFE2L2 contribute to the origin of lung squamous-cell carcinoma. Lung Cancer. 2022;171:34–41.

8. Tan C, Mandell JD, Dasari K, Cannataro VL, Alfaro-Murillo JA, Townsend JP. Heavy mutagenesis by tobacco leads to lung adenocarcinoma tumors with KRAS G12 mutations other than G12D, leading KRAS G12D tumors-on average-to exhibit a lower mutation burden. Lung Cancer. 2022;166:265–9.

9. Mochizuki A, Shiraishi K, Honda T, Higashiyama RI, Sunami K, Matsuda M, et al. Passive smoking-induced Mutagenesis as a promoter of lung carcinogenesis. J Thorac Oncol. 2024;19:984–94.

10. Lee J, Taneja V, Vassallo R. Cigarette Smoking and Inflammation: Cellular and Molecular Mechanisms. J Dent Res. 2012;91:142.

11. van der Vaart H, Timens W, Ten Hacken NHT. Acute effects of cigarette smoke on inflammation and oxidative stress: a review. Thorax. 2004;59:713–21.

12. Li X, Li J, Wu P, Zhou L, Lu B, Ying K, et al. Smoker and non-smoker lung adenocarcinoma is characterized by distinct tumor immune microenvironments. Oncoimmunology. 2018;7:e1494677.

13. Rodin SN, Rodin AS. Human lung cancer and p53: the interplay between mutagenesis and selection. Proc Natl Acad Sci U S A. 2000;97:12244–9.

14. Laconi E, Marongiu F, DeGregori J. Cancer as a disease of old age: changing mutational

and microenvironmental landscapes. Br J Cancer. 2020;122:943–52.

15. Marusyk A, DeGregori J. Declining cellular fitness with age promotes cancer initiation by selecting for adaptive oncogenic mutations. Biochim Biophys Acta. 2008;1785:1–11.

16. Lopez-Bigas N, Gonzalez-Perez A. Are carcinogens direct mutagens? Nat Genet. 2020;52:1137–8.

17. Stratton MR, Humphreys L, Alexandrov LB, Balmain A, Brennan P, Campbell PJ, et al. Implementing mutational epidemiology on a global scale: Lessons from mutographs. Cancer Discov. 2025;15:22–7.

18. Okamoto T, Suzuki Y, Fujishita T, Kitahara H, Shimamatsu S, Kohno M, et al. The prognostic impact of the amount of tobacco smoking in non-small cell lung cancer--differences between adenocarcinoma and squamous cell carcinoma. Lung Cancer. 2014;85:125–30.

19. Kenfield SA, Wei EK, Stampfer MJ, Rosner BA, Colditz GA. Comparison of Aspects of Smoking Among Four Histologic Types of Lung Cancer. Tob Control. 2008;17:198.

20. National Center for Health Statistics. Table 17, Current cigarette smoking among adults aged 18 and over, by sex, race, and age: United States, selected years 1965–2018. National Center for Health Statistics (US); 2021.

21. Devarakonda S, Li Y, Rodrigues FM, Sankararaman S, Kadara H, Goparaju C, et al. Genomic Profiling of Lung Adenocarcinoma in Never-Smokers. J Clin Oncol [Internet]. 2021 [cited 2024 Apr 25]; Available from: https://ascopubs.org/doi/10.1200/JCO.21.01691

22. Moorthi S, Paguirigan A, Itagi P, Ko M, Pettinger M, Hoge AC, et al. The genomic landscape of lung cancer in never-smokers from the Women's Health Initiative. JCI Insight [Internet]. 2024;9. Available from: https://insight.jci.org/articles/view/174643#:~:text=At%20the%20genetic%20level%2C%20lung, development%20(5%2C%2018).

23. Wang X, Ricciuti B, Nguyen T, Li X, Rabin MS, Awad MM, et al. Association between smoking history and tumor mutation burden in advanced non-small cell lung cancer. Cancer Res. 2021;81:2566–73.

24. Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell. 2012;150:1121–34.

25. Díaz-Gay M, Zhang T, Hoang PH, Leduc C, Baine MK, Travis WD, et al. The mutagenic forces shaping the genomes of lung cancer in never smokers. Nature. 2025;1–12.

26. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med. 2018;10:1–11.

27. Mandell JD, Cannataro VL, Townsend JP. Estimation of Neutral Mutation Rates and Quantification of Somatic Variant Selection Using cancereffectsizeR. Cancer Res. 2023;83:500–5.

28. Alfaro-Murillo JA, Townsend JP. Pairwise and higher-order epistatic effects among somatic

cancer mutations across oncogenesis. Math Biosci. 2023;366:109091.

29. Wang X, Fu AQ, McNerney ME, White KP. Widespread genetic epistasis among cancer genes. Nat Commun. 2014;5:1–10.

30. Iranzo J, Gruenhagen G, Calle-Espinosa J, Koonin EV. Pervasive conditional selection of driver mutations and modular epistasis networks in cancer. Cell Rep. 2022;40:111272.

31. Blair LM, Juan JM, Sebastian L, Tran VB, Nie W, Wall GD, et al. Oncogenic context shapes the fitness landscape of tumor suppression. Nat Commun. 2023;14:1–19.

32. Mina M, Iyer A, Ciriello G. Epistasis and evolutionary dependencies in human cancers. Curr Opin Genet Dev. 2022;77:101989.

33. Zhang Z, Yang Y, Zhou Y, Fang H, Yuan M, Sasser K, et al. A forward selection algorithm to identify mutually exclusive alterations in cancer studies. J Hum Genet. 2021;66:509–18.

34. Fedrizzi T, Ciani Y, Lorenzin F, Cantore T, Gasperini P, Demichelis F. Fast mutual exclusivity algorithm nominates potential synthetic lethal gene pairs through brute force matrix product computations. Comput Struct Biotechnol J. 2021;19:4394–403.

35. Liu S, Liu J, Xie Y, Zhai T, Hinderer EW, Stromberg AJ, et al. MEScan: a powerful statistical framework for genome-scale mutual exclusivity analysis of cancer mutations. Bioinformatics. 2021;37:1189–97.

36. Klein MI, Cannataro VL, Townsend JP, Newman S, Stern DF, Zhao H. Identifying modules of cooperating cancer drivers. Mol Syst Biol [Internet]. 2021 [cited 2024 Jul 2]; Available from: https://www.embopress.org/doi/10.15252/msb.20209810

37. Mina M, Raynaud F, Tavernari D, Battistello E, Sungalee S, Saghafinia S, et al. Conditional Selection of Genomic Alterations Dictates Cancer Evolution and Oncogenic Dependencies. Cancer Cell. 2017;32:155–68.e6.

38. Kryazhimskiy S, Rice DP, Jerison ER, Desai MM. Global epistasis makes adaptation predictable despite sequence-level stochasticity. Science [Internet]. 2014 [cited 2024 Apr 24]; Available from: https://www.science.org/doi/10.1126/science.1250939

39. Weinreich DM, Lan Y, Wylie CS, Heckendorn RB. Should evolutionary geneticists worry about higher-order epistasis? Curr Opin Genet Dev. 2013;23:700–7.

40. Sailer ZR, Harms MJ. High-order epistasis shapes evolutionary trajectories. PLoS Comput Biol. 2017;13:e1005541.

41. Weinreich DM, Lan Y, Jaffe J, Heckendorn RB. The influence of higher-order epistasis on biological fitness landscape topography. J Stat Phys. 2018;172:208–25.

42. Domingo J, Diss G, Lehner B. Pairwise and higher-order genetic interactions during the evolution of a tRNA. Nature. 2018;558:117–21.

43. Papkou A, Garcia-Pastor L, Escudero JA, Wagner A. A rugged yet easily navigable fitness landscape. Science. 2023;382:eadh3860.

44. Wei X, Zhang J. Patterns and Mechanisms of Diminishing Returns from Beneficial

Mutations. Mol Biol Evol. 2019;36:1008–21.

45. Aggeli D, Li Y, Sherlock G. Changes in the distribution of fitness effects and adaptive mutational spectra following a single first step towards adaptation. Nat Commun. 2021;12:1–14.

46. Chou H-H, Chiu H-C, Delaney NF, Segrè D, Marx CJ. Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science. 2011;332:1190–2.

47. Wünsche A, Dinh DM, Satterwhite RS, Arenas CD, Stoebel DM, Cooper TF. Diminishing-returns epistasis decreases adaptability along an evolutionary trajectory. Nature Ecology & Evolution. 2017;1:1–6.

48. Chen J, Yang H, Teo ASM, Amer LB, Sherbaf FG, Tan CQ, et al. Genomic landscape of lung adenocarcinoma in East Asians. Nat Genet. 2020;52:177–86.

49. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell. 2012;150:1107–20.

50. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the Evolution of Non-Small-Cell Lung Cancer. N Engl J Med. 2017;376:2109–21.

51. Jordan EJ, Kim HR, Arcila ME, Barron D, Chakravarty D, Gao J, et al. Prospective Comprehensive Molecular Characterization of Lung Adenocarcinomas for Efficient Patient Matching to Approved and Emerging Therapies. Cancer Discov. 2017;7:596–609.

52. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science. 2015;348:124–8.

53. Rizvi H, Sanchez-Vega F, La K, Chatila W, Jonsson P, Halpenny D, et al. Molecular Determinants of Response to Anti-Programmed Cell Death (PD)-1 and Anti-Programmed Death-Ligand 1 (PD-L1) Blockade in Patients With Non-Small-Cell Lung Cancer Profiled With Targeted Next-Generation Sequencing. J Clin Oncol. 2018;36:633–41.

54. Zhang T, Joubert P, Ansari-Pour N, Zhao W, Hoang PH, Lokanga R, et al. Genomic and evolutionary classification of lung cancer in never smokers. Nat Genet. 2021;53:1348.

55. Gillette MA, Satpathy S, Cao S, Dhanasekaran SM, Vasaikar SV, Krug K, et al. Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. Cell. 2020;182:200–25.e35.

56. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6:l1.

57. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014;511:543–50.

58. Kadara H, Choi M, Zhang J, Parra ER, Rodriguez-Canales J, Gaffney SG, et al. Whole-exome sequencing and immune profiling of early-stage lung adenocarcinoma with fully annotated clinical follow-up. Ann Oncol. 2017;28:75–82.

59. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. Bioinformatics. 2009;25:1841–2.

60. Manders F, Brandsma AM, de Kanter J, Verheul M, Oka R, van Roosmalen MJ, et al. MutationalPatterns: the one stop shop for the analysis of mutational processes. BMC Genomics. 2022;23:134.

61. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578:94–101.

62. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17:1–11.

63. Devarakonda S, Morgensztern D, Govindan R. Genomic alterations in lung adenocarcinoma. Lancet Oncol [Internet]. 2015 [cited 2024 Apr 26];16. Available from: https://pubmed.ncbi.nlm.nih.gov/26149886/

64. Greulich H. The Genomics of Lung Adenocarcinoma: Opportunities for Targeted Therapies. Genes Cancer. 2010;1:1200.

65. Seo JS, Ju YS, Lee WC, Shin JY, Lee JK, Bleazard T, et al. The transcriptional landscape and mutational profile of lung adenocarcinoma. Genome Res [Internet]. 2012 [cited 2024 Apr 26];22. Available from: https://pubmed.ncbi.nlm.nih.gov/22975805/

66. Li S, Choi YL, Gong Z, Liu X, Lira M, Kan Z, et al. Comprehensive Characterization of Oncogenic Drivers in Asian Lung Adenocarcinoma. J Thorac Oncol [Internet]. 2016 [cited 2024 Apr 26];11. Available from: https://pubmed.ncbi.nlm.nih.gov/27615396/

67. Abril-Pla O, Andreani V, Carroll C, Dong L, Fonnesbeck CJ, Kochurov M, et al. PyMC: a modern, and comprehensive probabilistic programming framework in Python. PeerJ Comput Sci. 2023;9:e1516.

68. Wilks SS. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. aoms. 1938;9:60–2.

69. Cannataro VL, Gaffney SG, Townsend JP. Effect sizes of somatic mutations in cancer. J Natl Cancer Inst. 2018;110:1171–7.

70. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

71. Sui Q, Liang J, Hu Z, Chen Z, Bi G, Huang Y, et al. Genetic and microenvironmental differences in non-smoking lung adenocarcinoma patients compared with smoking patients. Transl Lung Cancer Res. 2020;9:1407–21.

72. Frank R, Scheffler M, Merkelbach-Bruse S, Ihle MA, Kron A, Rauer M, et al. Clinical and Pathological Characteristics of KEAP1- and NFE2L2-Mutated Non-Small Cell Lung Carcinoma (NSCLC). Clin Cancer Res [Internet]. 2018 [cited 2024 Apr 25];24. Available from: https://pubmed.ncbi.nlm.nih.gov/29615460/

73. Nadal E, Palmero R, Muñoz-Pinedo C. Mutations in the Antioxidant KEAP1/NRF2 Pathway

Define an Aggressive Subset of NSCLC Resistant to Conventional Treatments. J Thorac Oncol [Internet]. 2019 [cited 2024 Apr 25];14. Available from: https://pubmed.ncbi.nlm.nih.gov/31668314/

74. Subramanian J, Govindan R. Lung Cancer in Never Smokers: A Review. J Clin Oncol [Internet]. 2016 [cited 2024 Apr 25]; Available from: https://ascopubs.org/doi/10.1200/JCO.2006.06.8015

75. Scott J, Marusyk A. Somatic clonal evolution: A selection-centric perspective. Biochim Biophys Acta. 2017;1867:139–50.

76. Fleenor CJ, Marusyk A, DeGregori J. Ionizing radiation and hematopoietic malignancies: altering the adaptive landscape. Cell Cycle. 2010;9:3005–11.

77. Hill W, Lim EL, Weeden CE, Lee C, Augustine M, Chen K, et al. Lung adenocarcinoma promotion by air pollutants. Nature. 2023;616:159–67.

78. Bossé Y, Postma DS, Sin DD, Lamontagne M, Couture C, Gaudreault N, et al. Molecular signature of smoking in human lung tissues. Cancer Res. 2012;72:3753–63.

79. Pintarelli G, Noci S, Maspero D, Pettinicchio A, Dugo M, De Cecco L, et al. Cigarette smoke alters the transcriptome of non-involved lung tissue in lung adenocarcinoma patients. Sci Rep. 2019;9:13039.

80. Ramirez JM, Ribeiro R, Soldatkina O, Moraes A, García-Pérez R, Oliveros W, et al. The molecular impact of cigarette smoking resembles aging across tissues. Genome Medicine. 2025;17:1–23.

81. Luo W, Zeng Z, Jin Y, Yang L, Fan T, Wang Z, et al. Distinct immune microenvironment of lung adenocarcinoma in never-smokers from smokers. Cell Rep Med. 2023;4:101078.

82. Price SN, Land SR, Pebley K, Fahey MC, Palmer AM, McCall MH, et al. Tobacco assessment in actively accruing National Cancer Institute clinical trials network trials. Nicotine Tob Res. 2025;ntaf071.

83. Peters EN, Warren GW, Sloan JA, Marshall JR. Tobacco assessment in completed lung cancer treatment trials: Pooling Project: Smoking and Lung Cancer. Cancer. 2016;122:3260–2.

84. Kim IA, Lee JS, Kim HJ, Kim WS, Lee KY. Cumulative smoking dose affects the clinical outcomes of EGFR-mutated lung adenocarcinoma patients treated with EGFR-TKIs: a retrospective study. BMC Cancer. 2018;18:768.

85. Kim MH, Kim HR, Cho BC, Bae MK, Kim EY, Lee CY, et al. Impact of cigarette smoking on response to epidermal growth factor receptor (EGFR)-tyrosine kinase inhibitors in lung adenocarcinoma with activating EGFR mutations. Lung Cancer. 2014;84:196–202.

86. Zhang P, Nie X, Bie Z, Li L. Impact of heavy smoking on the benefits from first-line EGFR-TKI therapy in patients with advanced lung adenocarcinoma. Medicine . 2018;97:e0006.

87. Igawa S, Sasaki J, Otani S, Ishihara M, Takakura A, Katagiri M, et al. Impact of smoking history on the efficacy of gefitinib in patients with non-small cell lung cancer harboring activating epidermal growth factor receptor mutations. Oncology. 2015;89:275–80.

88. van de Haar J, Canisius S, Yu MK, Voest EE, Wessels LFA, Ideker T. Identifying epistasis in cancer genomes: A delicate affair. Cell. 2019;177:1375–83.

89. Romero R, Sayin VI, Davidson SM, Bauer MR, Singh SX, LeBoeuf SE, et al. Keap1 loss promotes Kras-driven lung cancer and results in dependence on glutaminolysis. Nat Med. 2017;23:1362–8.

90. Best SA, Ding S, Kersbergen A, Dong X, Song J-Y, Xie Y, et al. Distinct initiating events underpin the immune and metabolic heterogeneity of KRAS-mutant lung adenocarcinoma. Nat Commun. 2019;10:4190.

91. Galan-Cobo A, Sitthideatphaiboon P, Qu X, Poteete A, Pisegna MA, Tong P, et al. LKB1 and KEAP1/NRF2 pathways cooperatively promote metabolic reprogramming with enhanced glutamine dependence in KRAS-mutant lung adenocarcinoma. Cancer Res. 2019;79:3251–67.

92. Ricciuti B, Arbour KC, Lin JJ, Vajdi A, Vokes N, Hong L, et al. Diminished Efficacy of Programmed Death-(Ligand)1 Inhibition in STK11- and KEAP1-Mutant Lung Adenocarcinoma Is Affected by KRAS Mutation Status. J Thorac Oncol. 2022;17:399–410.

93. Kwack WG, Shin SY, Lee SH. Primary resistance to immune checkpoint blockade in an STK11/TP53/KRAS-mutant lung adenocarcinoma with high PD-L1 expression. Onco Targets Ther. 2020;13:8901–5.

94. Arbour KC, Jordan E, Kim HR, Dienstag J, Yu HA, Sanchez-Vega F, et al. Effects of Co-occurring Genomic Alterations on Outcomes in Patients with KRAS-Mutant Non–Small Cell Lung Cancer. Clin Cancer Res. 2018;24:334–40.

95. Proulx-Rocray F, Routy B, Nassabein R, Belkaid W, Tran-Thanh D, Malo J, et al. The prognostic impact of KRAS, TP53, STK11 and KEAP1 mutations and their influence on the NLR in NSCLC patients treated with immunotherapy. Cancer Treat Res Commun. 2023;37:100767.

96. Baptiste Oudart J, Garinet S, Leger C, Barlesi F, Mazières J, Jeannin G, et al. STK11/LKB1 alterations worsen the poor prognosis of KRAS mutated early-stage non-squamous non-small cell lung carcinoma, results based on the phase 2 IFCT TASTE trial. Lung Cancer. 2024;190:107508.

97. Bange E, Marmarelis ME, Hwang W-T, Yang Y-X, Thompson JC, Rosenbaum J, et al. Impact of KRAS and TP53 co-mutations on outcomes after first-line systemic therapy among patients with STK11-mutated advanced non-small-cell lung cancer. JCO Precis Oncol. 2019;3:1–11.

98. West HJ, McCleland M, Cappuzzo F, Reck M, Mok TS, Jotte RM, et al. Clinical efficacy of atezolizumab plus bevacizumab and chemotherapy in KRAS-mutated non-small cell lung cancer with STK11, KEAP1, or TP53 comutations: subgroup results from the phase III IMpower150 trial. J Immunother Cancer. 2022;10:e003027.

99. Aredo JV, Padda SK, Kunder CA, Han SS, Neal JW, Shrager JB, et al. Impact of KRAS mutation subtype and concurrent pathogenic mutations on non-small cell lung cancer outcomes. Lung Cancer. 2019;133:144–50.

100. Wohlhieter CA, Richards AL, Uddin F, Hulton CH, Quintanal-Villalonga À, Martin A, et al. Concurrent mutations in STK11 and KEAP1 promote ferroptosis protection and SCD1

dependence in lung cancer. Cell Rep. 2020;33:108444.

101. Julian C, Pal N, Gershon A, Evangelista M, Purkey H, Lambert P, et al. Overall survival in patients with advanced non-small cell lung cancer with KRAS G12C mutation with or without STK11 and/or KEAP1 mutations in a real-world setting. BMC Cancer [Internet]. 2023;23. Available from: https://bmccancer.biomedcentral.com/articles/10.1186/s12885-023-10778-6#Sec6

102. Boeschen M, Kuhn CK, Wirtz H, Seyfarth H-J, Frille A, Lordick F, et al. Comparative bioinformatic analysis of KRAS, STK11 and KEAP1 (co-)mutations in non-small cell lung cancer with a special focus on KRAS G12C. Lung Cancer. 2023;184:107361.

103. Schaufler D, Ast DF, Tumbrink HL, Abedpour N, Maas L, Schwäbe AE, et al. Clonal dynamics of BRAF-driven drug resistance in EGFR-mutant lung cancer. npj Precision Oncology. 2021;5:1–12.

104. Gini B, Thomas N, Blakely CM. Impact of concurrent genomic alterations in epidermal growth factor receptor (EGFR)-mutated lung cancer. J Thorac Dis. 2020;12:2883–95.

105. Offin M, Chan JM, Tenet M, Rizvi HA, Shen R, Riely GJ, et al. Concurrent RB1 and TP53 alterations define a subset of EGFR-mutant lung cancers at risk for histologic transformation and inferior clinical outcomes. J Thorac Oncol. 2019;14:1784–93.

106. Pros E, Saigi M, Alameda D, Gomez-Mariano G, Martinez-Delgado B, Alburquerque-Bejar JJ, et al. Genome-wide profiling of non-smoking-related lung cancer cells reveals common RB1 rearrangements associated with histopathologic transformation in EGFR-mutant tumors. Ann Oncol. 2020;31:274–82.

107. Frankell AM, Dietzen M, Al Bakir M, Lim EL, Karasaki T, Ward S, et al. The evolution of lung cancer and impact of subclonal selection in TRACERx. Nature. 2023;616:525–33.

108. Park S, Lehner B. Cancer type-dependent genetic interactions between cancer driver alterations indicate plasticity of epistasis across cell types. Mol Syst Biol. 2015;11:824.

109. Guénolé A, Srivas R, Vreeken K, Wang ZZ, Wang S, Krogan NJ, et al. Dissection of DNA damage responses using multiconditional genetic interaction maps. Mol Cell. 2013;49:346–58.

110. Harrison R, Papp B, Pál C, Oliver SG, Delneri D. Plasticity of genetic interactions in metabolic networks of yeast. Proc Natl Acad Sci U S A. 2007;104:2307–12.

111. Zhu C-T, Ingelmo P, Rand DM. G×G×E for lifespan in Drosophila: mitochondrial, nuclear, and dietary interactions that modify longevity. PLoS Genet. 2014;10:e1004354.

112. Shi X, Reinstadler B, Shah H, To T-L, Byrne K, Summer L, et al. Combinatorial GxGxE CRISPR screen identifies SLC25A39 in mitochondrial glutathione transport linking iron homeostasis to OXPHOS. Nat Commun. 2022;13:2483.

113. Flynn KM, Cooper TF, Moore FB-G, Cooper VS. The environment affects epistatic interactions to alter the topology of an empirical fitness landscape. PLoS Genet. 2013;9:e1003426.

114. Costanzo M, Hou J, Messier V, Nelson J, Rahman M, VanderSluis B, et al. Environmental

robustness of the global yeast genetic interaction network. Science [Internet]. 2021;372. Available from: http://dx.doi.org/10.1126/science.abf8424

115. Sun Y, Yang Q, Shen J, Wei T, Shen W, Zhang N, et al. The Effect of Smoking on the Immune Microenvironment and Immunogenicity and Its Relationship With the Prognosis of Immune Checkpoint Inhibitors in Non-small Cell Lung Cancer. Front Cell Dev Biol. 2021;9:745859.

116. Boström M, Larsson E. Somatic mutation distribution across tumour cohorts provides a signal for positive selection in cancer. Nat Commun. 2022;13:7023.

117. Bolton KL, Ptashkin RN, Gao T, Braunstein L, Devlin SM, Kelly D, et al. Cancer therapy shapes the fitness landscape of clonal hematopoiesis. Nat Genet. 2020;52:1219–26.

118. Salehi S, Kabeer F, Ceglia N, Andronescu M, Williams MJ, Campbell KR, et al. Clonal fitness inferred from time-series modelling of single-cell cancer genomes. Nature. 2021;595:585–90.

119. Liggett LA, DeGregori J. Changing mutational and adaptive landscapes and the genesis of cancer. Biochim Biophys Acta Rev Cancer. 2017;1867:84–94.

120. Hsu T-K, Asmussen J, Koire A, Choi B-K, Gadhikar MA, Huh E, et al. A general calculus of fitness landscapes finds genes under selection in cancers. Genome Res. 2022;32:916–29.

121. Misra N, Szczurek E, Vingron M. Inferring the paths of somatic evolution in cancer. Bioinformatics. 2014;30:2456–63.

122. Marchetti A, Milella M, Felicioni L, Cappuzzo F, Irtelli L, Del Grammastro M, et al. Clinical implications of KRAS mutations in lung cancer patients treated with tyrosine kinase inhibitors: an important role for mutations in minor clones. Neoplasia. 2009;11:1084–92.

123. Hellyer JA, Stehr H, Das M, Padda SK, Ramchandran K, Neal JW, et al. Impact of KEAP1/NFE2L2/CUL3 mutations on duration of response to EGFR tyrosine kinase inhibitors in EGFR mutated non-small cell lung cancer. Lung Cancer. 2019;134:42–5.

124. Krall EB, Wang B, Munoz DM, Ilic N, Raghavan S, Niederst MJ, et al. KEAP1 loss modulates sensitivity to kinase targeted therapy in lung cancer. Elife [Internet]. 2017;6. Available from: http://dx.doi.org/10.7554/eLife.18970

125. Iyer G, Hanrahan AJ, Milowsky MI, Al-Ahmadie H, Scott SN, Janakiraman M, et al. Genome sequencing identifies a basis for everolimus sensitivity. Science. 2012;338:221.

126. Weigelt B, Reis-Filho JS. Epistatic interactions and drug response. J Pathol. 2014;232:255–63.

127. Wilkins JF, Cannataro VL, Shuch B, Townsend JP. Analysis of mutation, selection, and epistasis: an informed approach to cancer clinical trials. Oncotarget. 2018;9:22243–53.

128. Hobor S, Al Bakir M, Hiley CT, Skrzypski M, Frankell AM, Bakker B, et al. Mixed responses to targeted therapy driven by chromosomal instability through p53 dysfunction and genome doubling. Nat Commun. 2024;15:4871.

129. Bria E, Pilotto S, Amato E, Fassan M, Novello S, Peretti U, et al. Molecular heterogeneity assessment by next-generation sequencing and response to gefitinib of EGFR mutant advanced lung adenocarcinoma. Oncotarget. 2015;6:12783–95.

130. Vokes NI, Chambers E, Nguyen T, Coolidge A, Lydon CA, Le X, et al. Concurrent TP53 Mutations Facilitate Resistance Evolution in EGFR-Mutant Lung Adenocarcinoma. J Thorac Oncol. 2022;17:779–92.

131. Canale M, Petracci E, Delmonte A, Chiadini E, Dazzi C, Papi M, et al. Impact of TP53 Mutations on Outcome in EGFR-Mutated Patients Treated with First-Line Tyrosine Kinase Inhibitors. Clin Cancer Res. 2017;23:2195–202.

132. Ibusuki R, Iwama E, Shimauchi A, Tsutsumi H, Yoneshima Y, Tanaka K, et al. TP53 gain-of-function mutations promote osimertinib resistance via TNF-α-NF-κB signaling in EGFR-mutated lung cancer. NPJ Precis Oncol. 2024;8:60.

133. Canale M, Petracci E, Delmonte A, Bronte G, Chiadini E, Ludovini V, et al. Concomitant TP53 Mutation Confers Worse Prognosis in EGFR-Mutated Non-Small Cell Lung Cancer Patients Treated with TKIs. J Clin Med Res [Internet]. 2020;9. Available from: http://dx.doi.org/10.3390/jcm9041047

134. Choudhury NJ, Lavery JA, Brown S, de Bruijn I, Jee J, Tran TN, et al. The GENIE BPC NSCLC cohort: A real-world repository integrating standardized clinical and genomic data for 1,846 patients with non-small cell lung cancer. Clin Cancer Res. 2023;29:3418–28.

135. Fu J, Tong Y, Xu Z, Li Y, Zhao Y, Wang T, et al. Impact of TP53 mutations on EGFR-tyrosine kinase inhibitor efficacy and potential treatment strategy. Clin Lung Cancer. 2023;24:29–39.

136. Donehower LA, Soussi T, Korkut A, Liu Y, Schultz A, Cardenas M, et al. Integrated analysis of TP53 gene and pathway alterations in The Cancer Genome Atlas. Cell Rep. 2019;28:1370–84.e5.

137. Rho JK, Choi YJ, Ryoo B-Y, Na III, Yang SH, Kim CH, et al. p53 enhances gefitinib-induced growth inhibition and apoptosis by regulation of Fas in non-small cell lung cancer. Cancer Res. 2007;67:1163–9.

138. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990;61:759–67.

139. Paterson C, Clevers H, Bozic I. Mathematical model of colorectal cancer initiation. Proc Natl Acad Sci U S A. 2020;117:20681–8.

140. Ortmann CA, Kent DG, Nangalia J, Silber Y, Wedge DC, Grinfeld J, et al. Effect of mutation order on myeloproliferative neoplasms. N Engl J Med. 2015;372:601–12.

141. Ernst SM, Mankor JM, van Riet J, von der Thüsen JH, Dubbink HJ, Jgjv A, et al. Tobacco Smoking-Related Mutational Signatures in Classifying Smoking-Associated and Nonsmoking-Associated NSCLC. J Thorac Oncol [Internet]. 2023 [cited 2024 Apr 24];18. Available from: https://pubmed.ncbi.nlm.nih.gov/36528243/

142. Hobbs SD, Wilmink AB, Adam DJ, Bradbury AW. Assessment of smoking status in patients

with peripheral arterial disease. J Vasc Surg [Internet]. 2005 [cited 2024 Apr 24];41. Available from: https://pubmed.ncbi.nlm.nih.gov/15838479/

143. Hald J, Overgaard J, Grau C. Evaluation of objective measures of smoking status--a prospective clinical study in a group of head and neck cancer patients treated with radiotherapy. Acta Oncol [Internet]. 2003 [cited 2024 Apr 24];42. Available from: https://pubmed.ncbi.nlm.nih.gov/12801134/

144. Brennan P, Buffler PA, Reynolds P, Wu AH, Wichmann HE, Agudo A, et al. Secondhand smoke exposure in adulthood and risk of lung cancer among never smokers: a pooled analysis of two large studies. International journal of cancer [Internet]. 2004 [cited 2024 Apr 24];109. Available from: https://pubmed.ncbi.nlm.nih.gov/14735478/

145. Morris SM. A role for p53 in the frequency and mechanism of mutation. Mutat Res [Internet]. 2002 [cited 2024 Apr 24];511. Available from: https://pubmed.ncbi.nlm.nih.gov/11906841/

146. Zámborszky J, Szikriszt B, Gervai JZ, Pipek O, Póti Á, Krzystanek M, et al. Loss of BRCA1 or BRCA2 markedly increases the rate of base substitution mutagenesis and has distinct effects on genomic deletions. Oncogene. 2016;36:746–55.

147. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, et al. The landscape of somatic copy-number alteration across human cancers. Nature. 2010;463:899–905.

148. Besedina E, Supek F. Copy number losses of oncogenes and gains of tumor suppressor genes generate common driver mutations. Nat Commun. 2024;15:6139.

149. Tomanek I, Guet CC. Adaptation dynamics between copy-number and point mutations. Elife [Internet]. 2022;11. Available from: http://dx.doi.org/10.7554/eLife.82240

150. Ciriello G, Cerami E, Sander C, Schultz N. Mutual exclusivity analysis identifies oncogenic network modules. Genome Res. 2012;22:398–406.

151. Dasari K. Townsend-Lab-Yale/lung-smoking: Software for “Tobacco smoke alters the trajectory of lung adenocarcinoma evolution via effects on somatic selection and epistasis.” [cited 2025 Jul 23]; Available from: https://zenodo.org/records/16379015

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Availability of data and materials

The datasets used, generated, and analyzed during the current study are available in the Zenodo repository, https://zenodo.org/records/16379015 [151].

### Competing interests

The authors declare that they have no competing interests.

### Funding

Support for this research was supplied by a Yale College First-Year Summer Research Fellowship in the Sciences and Engineering to KD as well as by developmental program funds from NIH P50CA196530 and the Elihu Endowment at Yale to JPT.

### Authors' contributions

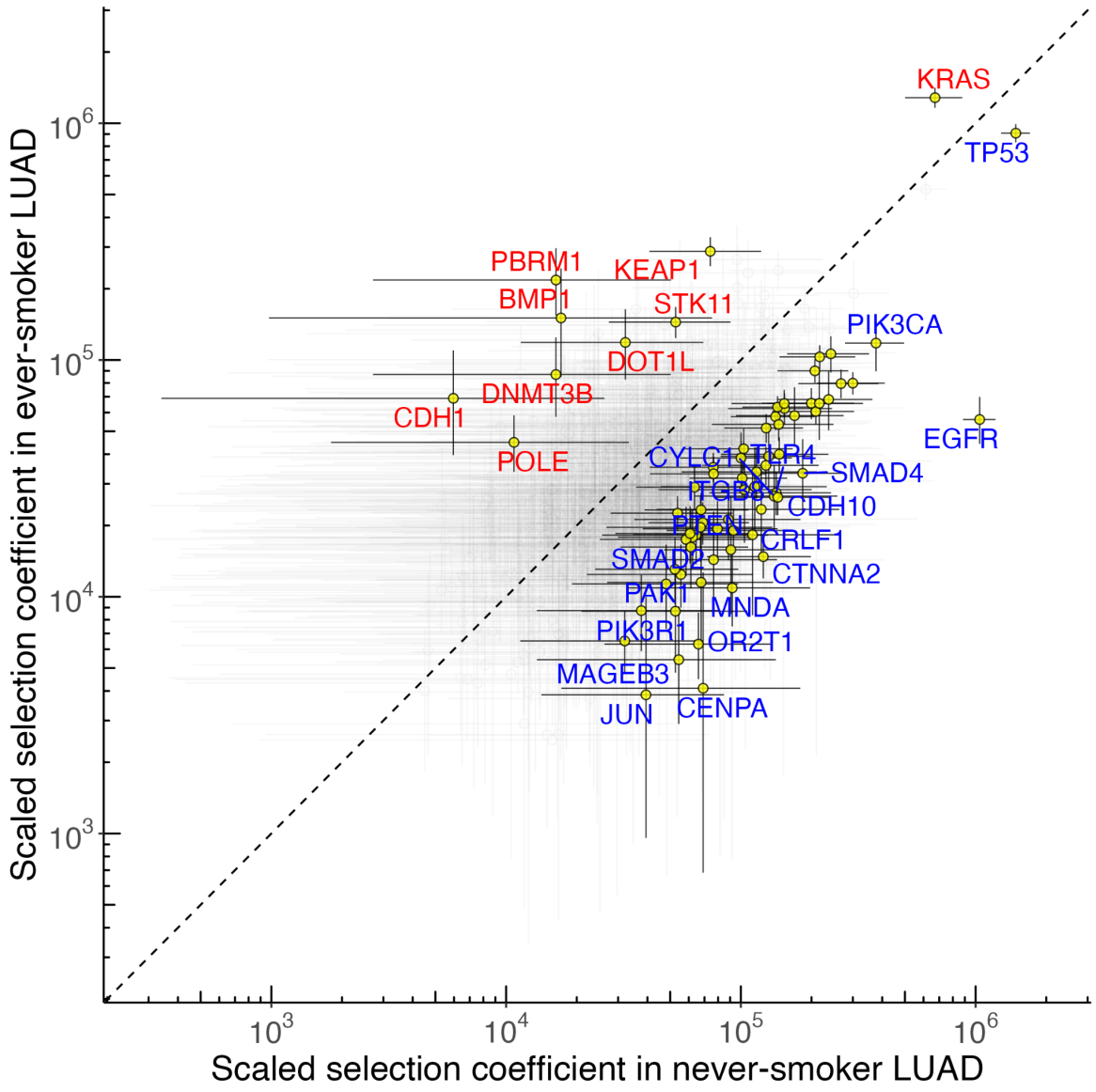
JPT, JAA-M, and KD conceptualized the project. KD curated the data and performed the formal analysis. KD and JAA-M wrote the software. KD, JAA-M, and JPT contributed to the methodology and investigation. KD drafted the original manuscript and JPT and JAA-M reviewed and edited the manuscript. All authors read and approved the final manuscripts.

### Acknowledgements

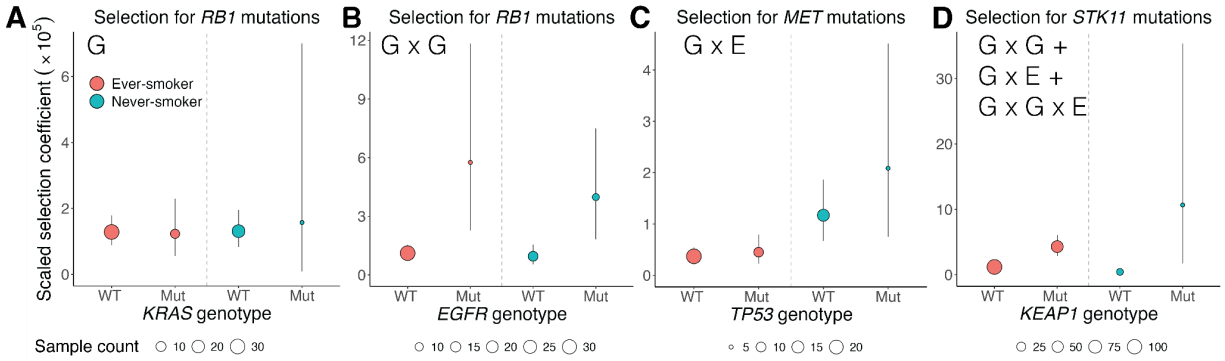
We thank Jeff Mandell for his advice on mutational signature attribution and mutation rate estimation, and Elizabeth Perry for assistance with gene expression analysis.

## Supplementary Materials

### Supplementary Figures and Tables




**Supplementary Figure 1.** Strengths of somatic selection on 1,200 oncogenic gene mutations between lung adenocarcinoma (LUAD) in ever-smokers (*n* = 1,066) and in never-smokers (*n* = 656; significant: yellow circle, insignificant: no circle, red gene name: stronger selection in ES-LUAD, blue gene name: stronger selection in NS-LUAD).




**Supplementary Figure 2. Four patterns of the influences of genetic and environmental interactions in ever-smoker and never-smoker lung adenocarcinoma.** (**A**) *RB1* mutations with wild-type (WT) or mutant *KRAS*, (**B**) *RB1* mutations with wild-type or mutant *EGFR*, (**C**) *PIK3CA* mutations with wild-type or mutant *KRAS*, and (**D**) *EGFR* mutations with wild-type or mutant *TP53*.

**Table S1. Characteristics of datasets included in study.**

| Study | Institution/project | Sequencing methods | Treatment prior to sample retrieval | Number of samples included in final analysis |
|---|---|---|---|---|
| The Cancer Genome Atlas Research Network., 2014 | The Cancer Genome Atlas | Whole-exome sequencing | 3 / 451[*] | 451 |
| Jordan et al., 2017 | Memorial Sloan Kettering Cancer Center (MSKCC) | Targeted sequencing (MSK-IMPACT 341, MSK-IMPACT 410) | 101 / 305 | 305 |
| Chen et al., 2020 | OncoSG | Whole-exome sequencing | ≥ 70 / 247[‡, §] | 247 |
| Zhang et al., 2021 | National Cancer Institute | Whole-genome sequencing | None | 187 |
| Rizvi et al., 2018 | MSKCC | Targeted sequencing (MSK-IMPACT 341, MSK-IMPACT 410, MSK-IMPACT 468) | 164 / 164[‖] | 164 |

| | | | | |
|---|---|---|---|---|
| Imielinski et al., 2012. | Broad Institute | Whole-exome sequencing, whole-genome sequencing, or both | None | 153 |
| Gillette et al., 2020 | Clinical Proteomic Tumor Analysis Consortium | Whole-exome sequencing | None | 74 |
| Kadara et al., 2017 | Yale University | Whole-exome sequencing | Not available | 73 |
| Jamal-Hanjani et al., 2017 | TracerX | Whole-exome sequencing | None | 61 |
| Rizvi et al., 2015 | MSKCC | Whole-exome sequencing | 7 / 7 | 7 |

* Treatment not indicated
† Combination of targeted therapy, immunotherapy, and/or chemotherapy
‡ Treatment information is incomplete for 75 samples
§ Tyrosine-kinase inhibitor and/or chemotherapy
‖ Anti–PD-(L)1 therapy, sometimes combined with anti–CTLA-4 therapy
Pembrolizumab