Tobacco smoke alters the landscape of mutation, selection, and epistasis in lung adenocarcinoma

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

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

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

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

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

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

Abstract

Tobacco smoke elevates mutagenesis. Its physiological effects may additionally influence the adaptive benefits of specific mutations to oncogenesis. However, the relative contributions of these mutagenic and physiological effects to oncogenesis have not been quantified. Here we deconvolved the prevalence of 21 driver mutations into components of mutation and selection in ever-smoker (ES) and never-smoker (NS) lung adenocarcinoma (LUAD). Mutagenesis was broadly elevated in ES-LUAD. Moreover, mutations of some genes experienced stronger selection in ES-LUAD (*KRAS*, *KEAP1*, and *STK11*) or NS-LUAD (*TP53*, *EGFR*, *PIK3CA*, *SMAD4*, and more). Analyzing the influence of somatic genotype, we found widespread antagonistic epistasis and strong synergistic epistasis, with more frequent synergy and substantially less frequent antagonism in ES-LUAD. When multiple mutations were present, higher-order epistasis exhibited strong effects on the oncogenic adaptive landscape. This disambiguation of the mutagenic and selective effects of smoking facilitates quantitative prediction of treatment response based on smoking history and tumor somatic genotype.

INTRODUCTION

Tobacco smoke contains carcinogens known to cause cancer, in part by inducing mutagenesis (*1*). Exposure to these carcinogens can trigger distinct forms of DNA damage, generating recognizable patterns of mutations (*2*), a fraction of which then contribute to oncogenesis (*3–9*). In addition to this mutagenesis-mediated oncogenic effect, tobacco smoke can contribute to oncogenesis by altering the lung microenvironment: inducing chronic inflammation, oxidative stress, immune dysfunction, and cellular damage (*1*, *10–12*). Smoke-induced alterations of lung physiology could increase the likelihood that specific mutations contribute to oncogenesis—that is, alter their specific somatic selective advantage (*13–16*). However, the extent to which such physiological effects alter the adaptive landscape of cancer has not been previously quantified. These physiological effects may be a substantial contributor to differences in the genetic profiles (*17–20*) and biological characteristics (*21–25*) of the tumors of ever-smokers and never-smokers, as well as the outcomes of patients (*26*, *27*). Of lung cancer subtypes, LUAD afflicts the largest proportion of never-smokers, providing better-balanced tumor sample sizes for elucidating underlying differences (*28*). In the US as of 2002, nearly 23% of LUADs occur in never-smokers (*29*). This proportion is likely growing with the decline of cigarette usage (*30*). 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.

Existing computational models can infer baseline somatic mutation rates from observed rates of synonymous mutations within tumor samples, implicitly (*6*) and explicitly (*31*) accounting for elevated mutagenesis due to carcinogens such as tobacco smoke. These baseline mutation rates can be compared to observed mutation prevalence to infer the degree of selection acting on the mutations (*8*, *32*, *33*), providing a framework to disambiguate the oncogenic contributions of the mutagenic and physiological effects of smoking. This strength of selection far more accurately predicts the contribution of mutations to tumor growth than does mutation prevalence (*34*, *35*). However, selection on every mutation is in each

case likely influenced not only by the microenvironmental effects of tobacco smoke, but also by pairwise (*36*, *37*) and higher-order selective epistasis between mutations. In selective epistasis, one mutation affects the selection operating on other mutations. These interactions between the selective effects of mutations shape the adaptive landscape of cancer evolution (*5*, *38*, *39*).

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 (*40*, *41*) in the context of heterogeneous mutation rates (*37*, *42–44*). However, with the exception of Coselens (*37*), 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 therapeutic strategies.

Moreover, all of these approaches are restricted to pairwise inferences (*37*, *41*) or to inference of discrete modules (*40*, *42–44*), 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 (*45–50*), including the degree to which selective epistasis diminishes along the adaptive trajectory (*51–54*). 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 (*33*). 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.

Here, we quantify the mutagenic and selective effects of tobacco smoke on single-nucleotide variants (SNVs) in LUAD. To do so, we compiled 8,868 lung adenocarcinoma tumor sample genomes.

Of these, we classified 1,723 by smoking status into ever-smoker or never-smoker cohorts. We estimated background mutation rates in both cohorts. We compared these estimated mutation rates to observed mutation prevalence in tumors of each cohort, calculating for each mutation the average strength of selection necessary to resolve the discrepancy. In addition to this calculation that averages across somatic genotypes, we calculated selection for mutations within specific somatic genotypes, enabling quantification of pairwise and higher-order epistasis. We found numerous substantial differences in the mutational and adaptive landscapes of LUAD between ever-smokers and never-smokers. These differences guide the development and application of therapies to account for genetic and environmental interactions in cancer evolution.

RESULTS

Rates of mutation and selection determine variant prevalence in lung adenocarcinoma (LUAD)

Deconvolving the prevalence of somatic tumor sequence variants in our literature-based set of 21 reputed driver genes into background oncogenic mutation rates and scaled selection coefficients revealed mutations in several genes that were under intense positive selection. Across all LUAD samples in our dataset (n = 9,457), *TP53* and *KRAS* were the most frequently mutated genes, followed distantly by *EGFR*, *KEAP1*, and *STK11* (**Fig. 1A**). The background oncogenic mutation rates for the majority of genes spanned slightly less than an order of magnitude—between 2.2×10^{-7} and 8.8×10^{-7} (**Fig. 1B**). Exceptions to that range included *CTNNB1*, which was subject to an unusually low mutation rate of 1.2×10^{-7}, and *STK11*, *GNAS*, and *ALK*, which were subject to mutation rates greater than 1×10^{-6}.

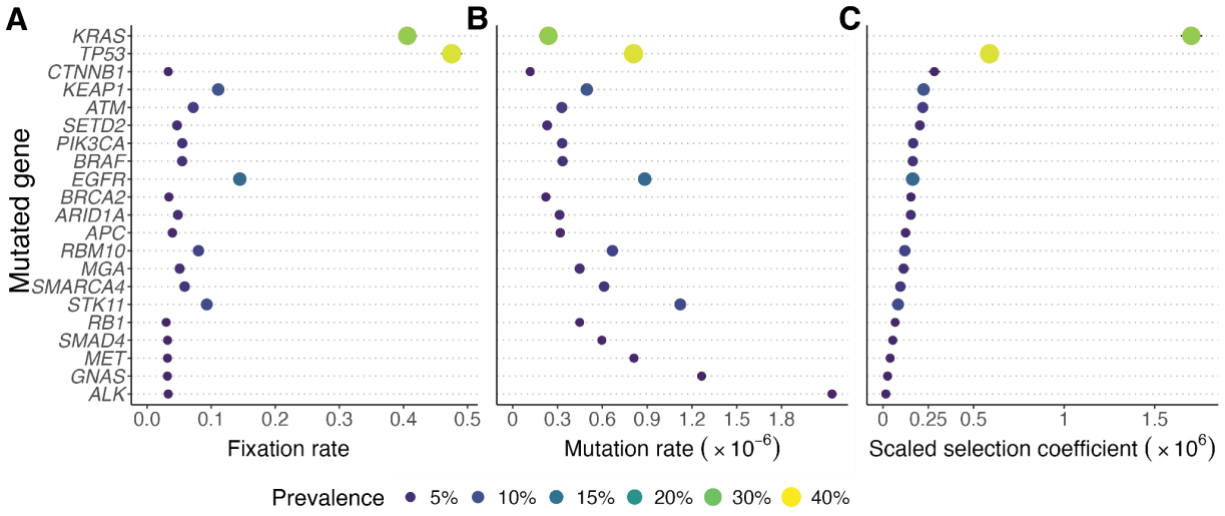


Figure 1. Fixation rates and scaled selection coefficients for 21 known and potential driver mutations among all patients (*n* = 9,457) with lung adenocarcinoma. (**A**) Poisson-corrected fixation rates (rates at which each mutation occurs and becomes fixed within the population). (**B**) Background rate at which oncogenic gene mutations occur. (**C**) Strength of selection experienced by oncogenic gene mutations.

Scaled selection coefficients for these mutations followed a long-tailed distribution (**Fig. 1C**). Mutations in *KRAS* experienced the strongest selection, followed distantly by mutations of *TP53*, and then by mutations in *CTNNB1*, *KEAP1*, *ATM*, *SETD2*, and other genes that experienced substantial but weaker selection (**Fig. 1C**). *CTNNB1* mutations are notable for appearing in only 3.2% of patients. Nevertheless, 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*. In general, there were substantial discrepancies between the relative prevalence of a mutation (**Fig. 1A**) and the relative intensity of selection it experienced (**Fig. 1C**).

KRAS mutations are observed over seven times more frequently than *SETD2* mutations. The two genes have comparable oncogenic mutation rates—2.4×10^{-7} for *KRAS* and 2.3×10^{-7} for *SETD2*. Consequently, the greater relative prevalence of *KRAS* mutations across tumors can be attributed to stronger positive selection. In contrast, *CTNNB1* and *MET* mutations are observed at nearly the same prevalence. However, the oncogenic mutation rate of *MET* is roughly seven times that of *CTNNB1*.

Consequently, to achieve the same prevalence in tumors as mutations of *MET*, *CTNNB1* mutations must experience roughly seven times as much selection. The relative prevalence of other pairs of mutations can be similarly explained by distinct contributions of mutation and selection.

These estimates of oncogenic mutation rates and selection coefficients are quantified based on all tumors, and therefore represent the oncogenic mutation rates and selection coefficients for an average LUAD tumor. However, the mutagenic and physiological effects of smoking likely influence both the underlying mutation rates and the strength of selection for oncogenic mutations of specific genes. Therefore, we performed isolated analyses of selection using tumor sequences from ever-smokers and never-smokers.

Mutation and selection differ substantially between ever-smoker LUAD (ES-LUAD) and never-smoker LUAD (NS-LUAD)

After division of all the samples by smoking status using signature analysis and clinical classification, we obtained two cohorts: ever-smokers (n = 1,066) and never-smokers (n = 656). To distinguish the mutagenic and physiological-adaptive effects of smoking, we computed oncogenic mutation rates and selection coefficients for our 21 reputed driver genes in both cohorts independently. We found significant differences in both mutation and selection in ES-LUAD and NS-LUAD. The oncogenic mutation rates for all genes were higher in ES-LUAD, ranging from moderately increased (*RB1*, *CTNNB1*, and *ARID1A*; 13%–41%) to over five times greater (*EGFR*, *SMAD4*, *GNAS*, and *ALK*; 403–1288%; **Fig. 2A**). These elevated mutation rates demonstrate the intense mutagenic effect of tobacco smoking on the peripheral tissue of the lung.

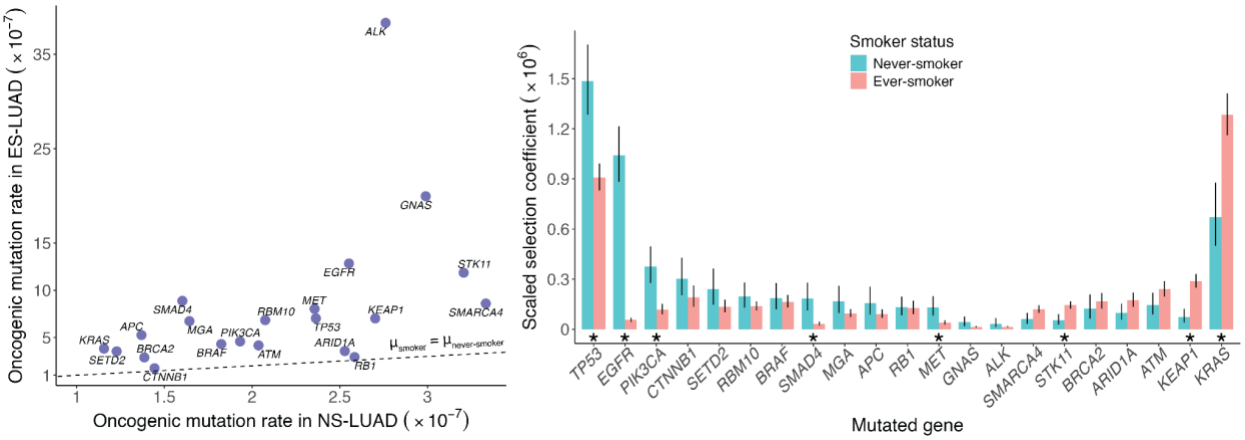


Figure 2. Mutational and adaptive landscapes of ever-smoker lung adenocarcinoma (ES-LUAD; *n* = 1,066) and never-smoker lung adenocarcinoma (NS-LUAD; *n* = 656). (**A**) Background oncogenic mutation rates for 21 known and potential driver genes in ES-LUAD and NS-LUAD (dashed line: equality). (**B**) Strengths of selection (*: non-overlapping 95% confidence intervals) for mutations of these genes in NS-LUAD (teal bars) and ES-LUAD (salmon bars), averaged across somatic genotypes.

In accord with previous analyses that have demonstrated a substantially different prevalence of major driver genes by smoking status (*25*, *55*, *56*), we found significant differences between the selection coefficients in ES-LUAD and NS-LUAD. Among our set of 21 reputed driver genes, mutations of eight genes are significantly differentially selected (**Fig. 2B**). *TP53*, *EGFR*, *PIK3CA*, *SMAD4*, and *MET* 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.

Of the mutations that were significantly differentially selected in ES-LUAD and NS-LUAD, some still experienced strong selection in both contexts. *TP53* and *KRAS* mutations were among the mutations under the strongest selection in both cancer types, despite experiencing differential selection (**Fig. 2B**). However, most mutations that experienced significantly different selection were strongly selected in one context and weakly selected in the other. For example, *KEAP1* mutations experienced the third strongest selection in ES-LUAD (after *TP53*, and *KRAS*) and experienced nearly four times stronger selection in ES-LUAD than in NS-LUAD. *KEAP1* mutations have been observed to be more common in ever-smoker

populations (*57*, *58*). Similarly, *STK11* mutations experienced over twice the strength of selection in ES-LUAD compared to NS-LUAD. *EGFR* mutations experienced the greatest difference in selection, with nearly twenty-fold increase in selection in NS-LUAD relative to ES-LUAD. They were also the second most highly selected mutations in NS-LUAD, behind *TP53*. *EGFR* mutations are known to be much more common in never-smoker patient populations than in ever-smoker populations (*55*, *59*). *PIK3CA*, *SMAD4*, and *MET* mutations also experienced over a threefold increase in selection in NS-LUAD relative to ES-LUAD.

Across a set of 1200 LUAD-associated genes, nine were significantly more selected in ES-LUAD (*KRAS*, *KEAP1*, *STK11*, *PBRM1*, *DOT1L*, *BMP1*, *DNMT3B*, *CDH1*, and *POLE*), and seventy-one were significantly more selected in NS-LUAD (**Fig. S1**). Mutations of *EGFR*, *CENPA*, *JUN*, *CTNNA2*, *PIK3R1*, *PAK1*, *SMAD2*, *SMAD4*, *PTEN*, and *TLR4* were among the mutations that experienced the greatest ratio of strength of selection in NS-LUAD compared to ES-LUAD. Typically, mutations experienced lower selection in ES-LUAD than in NS-LUAD (**Fig. S1**).

Overall, mutations often experienced significantly different intensities of selection in ES-LUAD and NS-LUAD, differentiating the adaptive landscapes of these two groups. Effects of each somatic mutation can be further analyzed to reveal epistasis with other mutations that mutually shape the adaptive landscapes of tumors in ever-smokers and never-smokers. Therefore we assessed the influence of pairwise epistasis on selection for mutations.

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

To determine the prevalence and strength of pairwise epistasis, we computed—for each pair of genes—selection coefficients for each mutated gene in “normal” somatic genotypes (wild type for both genes) and in somatic genotypes where the other gene was mutated. In both ES-LUAD and NS-LUAD, we calculated all 420 pairwise models for each combination of two genes from the 21 studied.

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

We grouped samples according to their smoking history and somatic genotype for *TP53* and *KRAS* (the two most frequently mutated genes in LUAD). This grouping enabled estimation of fixation rates of each mutation from each somatic genotype (**Fig. 3A**). In ES-LUAD, *TP53* mutations exhibited a high fixation rate when *KRAS* was wild-type. With *KRAS* mutated, *TP53* mutations exhibited a lower, yet still substantial fixation rate (**λ, Fig. 3B**). In contrast, the fixation rates of *KRAS* mutations were high regardless of whether *TP53* was mutated or not. In ES-LUAD, the most likely evolutionary trajectory involved fixation of a *TP53* mutation followed by fixation of a *KRAS* mutation. The mutation rate underlying gene variants observed in our dataset was larger for *TP53* variants than *KRAS* variants in ES-LUAD (μ = 7.0 × 10^{-7} for *TP53*, μ = 3.8 × 10^{-7} for *KRAS*; **Fig. 3B**). Therefore, a mutation of *KRAS* occurring in a wild-type genotype experienced stronger selection than a mutation of *TP53* occurring in a wild-type genotype (γ, **Fig. 3B**), despite the lower fixation rate of the *KRAS* mutation. Selection for mutation of *TP53* in ES-LUAD was antagonized by the presence of a mutation in *KRAS*, but selection for mutation of *KRAS* was unaffected by *TP53* status (γ, **Fig. 3B**). A mutation of *TP53* when *KRAS* was wildtype and a mutation of *KRAS* in either *TP53* genotype experienced similarly strong intensities of selection. In this pairwise analysis, the least selected mutation along the trajectory was that of *TP53* when *KRAS* was mutant.

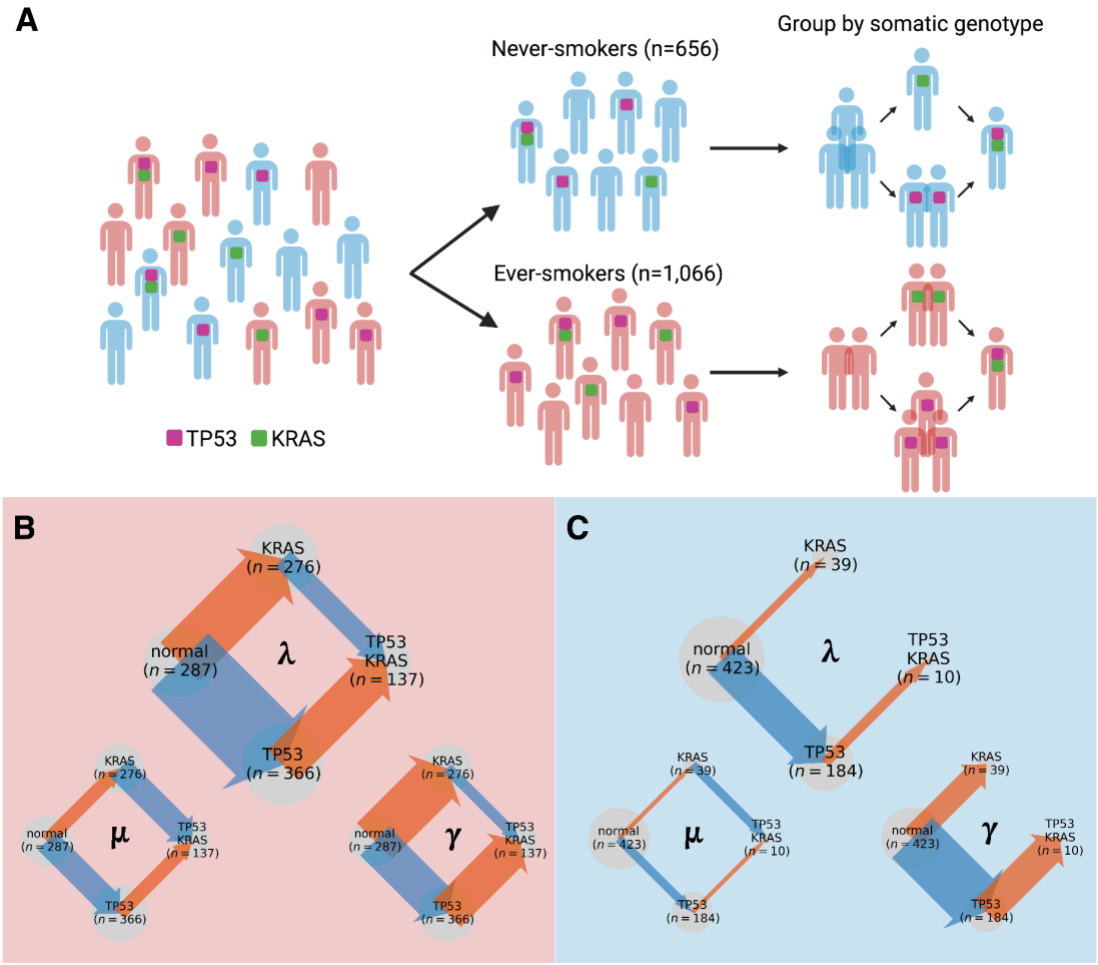


Figure 3. Pairwise selective epistasis inferred within the evolutionary trajectories of lung adenocarcinoma (LUAD), depicted for *TP53* and *KRAS*. (**A**) Genomic sequence data from lung adenocarcinomas can be stratified by clinically- and genetically-determined patient smoking status (never-smokers: blue; ever-smokers: red) 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 selection (γ; arrow width) on mutations of *TP53* (blue arrows) and *KRAS* (orange arrows).

The fixation rates for all mutations in NS-LUAD were lower than the corresponding rates for ES-LUAD (*λ*, **Fig. 3C**), and the differences between fixation rates of *TP53* and *KRAS* mutations were exacerbated. However, the most likely trajectory was the same as in ES-LUAD: fixation of a *TP53*

mutation followed by a *KRAS* mutation. The oncogenic mutation rate of *TP53* was 104% larger than the oncogenic mutation rate of *KRAS* in NS-LUAD, and both mutations occurred at about a third of the rate of the corresponding mutations in ES-LUAD (μ, **Fig. 3C**). In comparison to ES-LUAD, *TP53* mutations experienced stronger antagonistic epistasis as a consequence of mutated *KRAS* in NS-LUAD. Indeed, after *KRAS* mutations in ES-LUAD, TP53 mutations were potentially under negative selection: the maximum likelihood estimate for γ was below one (**Fig. 3B**).

Synergistic epistasis is frequent and strong in ES-LUAD

Examining all pairwise combinations of the 21 genes in ES-LUAD and NS-LUAD, we found that both synergistic and antagonistic epistasis were prevalent in LUAD. These epistatic effects constituted substantial impacts on the selection experienced by mutations (**Fig. 4A**). Mutations that experienced strong selection in ES-LUAD, except for *TP53* and *KRAS*, experienced substantial synergistic epistasis: effects generally ranged from a two to a sixfold increase in selection. For example, selection on mutations in *KEAP1*, *ATM*, *CTNNB1*, and *ARID1A* were enhanced by prior mutation of *TP53* or *KRAS* (**Fig. 4B**). Strong synergistic epistasis in ES-LUAD was also inferred for mutants of *ARID1A* in the context of *ALK*, *ATM*, and *RBM10* mutations, for mutants of *STK11* in the context of mutated *KEAP1*, for mutants of *BRAF* in the context of mutated *EGFR*, and for mutants of *ATM* in the context of mutated *STK11*. Moderately strong positive selection on mutations in the tumor-suppressor genes *BRCA2*, *SETD2*, *RBM10*, *RB1*, *MGA*, and *APC*—as well as in the SWI/SNF chromatin-remodeling genes *ARID1A* and *SMARCA4*—was often enhanced by prior mutations in a heterogeneous set of other genes, and was infrequently substantially antagonized by other mutations. Mutations in *TP53* and *EGFR* frequently exhibited synergistic epistasis with the mutated tumor suppressor genes *BRCA2*, *SETD2*, *RBM10*, *RB1*, *MGA*, and *APC*; whereas *BRAF*, *ARID1A*, *PIK3CA*, *RBM10*, *SMARCA4*, *and ALK* occasionally did. Mutations in the tumor suppressors *TP53* and *RBM10* synergistically enhanced selection on mutations in the mutated chromatin-remodeling genes *ARID1A* and *SMARCA4* (**Fig. 4B,D**).

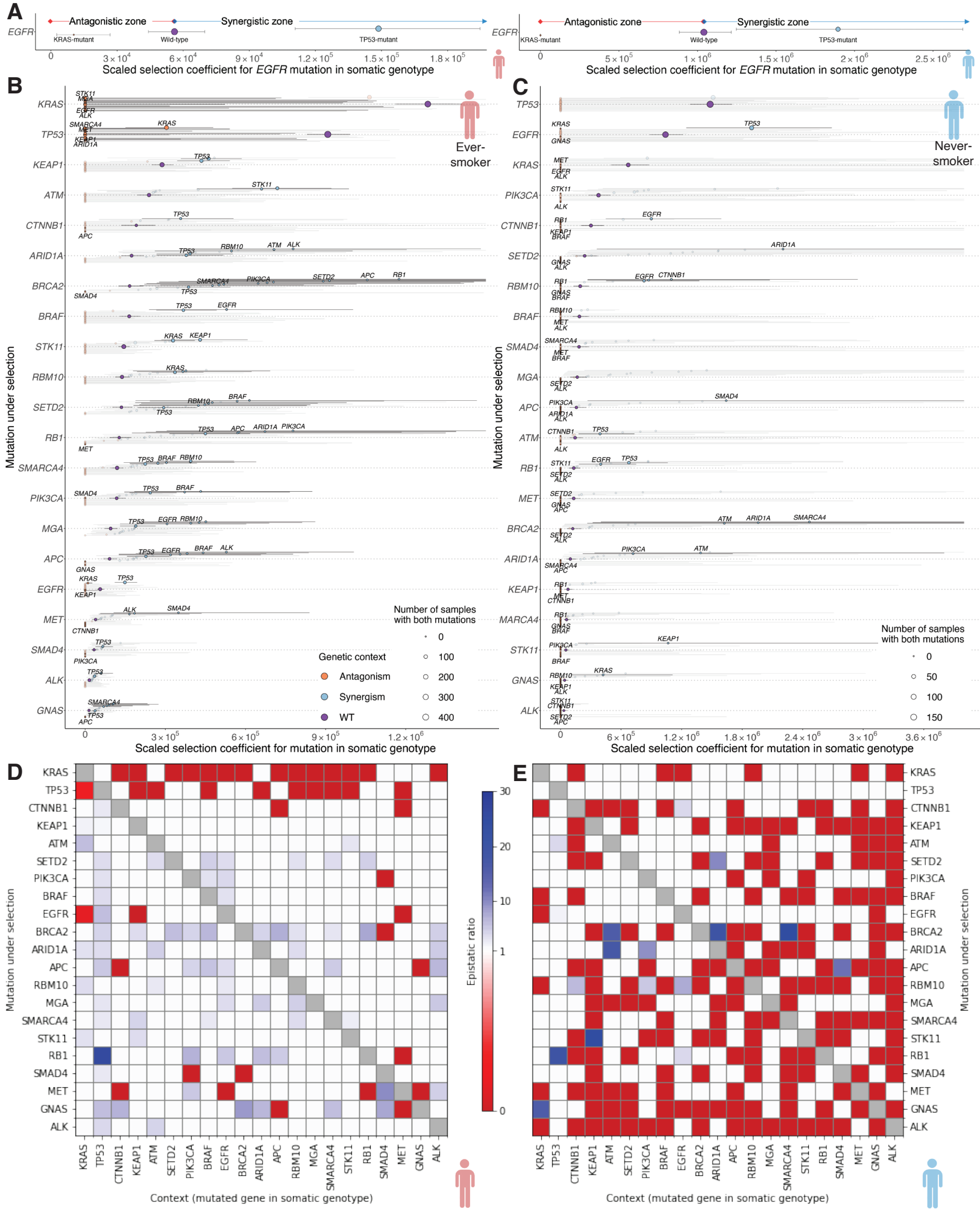

A
EGFR
Antagonistic zone
Synergistic zone
KRAS-mutant
Wild-type
TP53-mutant
Scaled selection coefficient for EGFR mutation in somatic genotype
B
Ever-smoker
Mutation under selection
Scaled selection coefficient for mutation in somatic genotype
Number of samples with both mutations
Genetic context
Antagonism
Synergism
WT
C
Never-smoker
Mutation under selection
Scaled selection coefficient for mutation in somatic genotype
Number of samples with both mutations
D
Mutation under selection
Context (mutated gene in somatic genotype)
Epistatic ratio
E
Mutation under selection
Context (mutated gene in somatic genotype)

Figure 4. Pairwise epistasis in ever-smoker (ES) and never-smoker (NS) lung adenocarcinoma (LUAD). (**A**) Strengths of selection experienced by *EGFR* mutations in normal versus *TP53* and *KRAS* mutant genotypes (synergistic: blue; antagonistic: red). In (**B**) ES-LUAD and (**C**) NS-LUAD, selection for mutations from wild-type (purple) and in the presence of a previous mutation (significant: labeled; insignificant: gray-hued). Significant pairwise epistatic interactions depicted in heatmap form for (**D**) ES-LUAD and (**E**) NS-LUAD. 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 normal somatic genotype.

The epistatic patterns of oncogenes were highly variable. Selection for mutation of *KRAS* was not enhanced by previous mutations of any of the tested genes. Selection for mutations in other genes, such as *GNAS*, was enhanced by previous mutations in up to six other genes (**Fig. 4B,D**). Selection on many mutations of oncogenes was enhanced by previous mutation of *TP53*, and was antagonized by mutations in other oncogenes—namely *CTNNB1*, *APC*, and *MET*. Across mutations of all tested genes, the median effect of statistically significant synergistic epistasis was a 4.3-fold increase in selection. The strength of synergistic epistasis did not correlate with the selection experienced by the prior mutation. However, it did correlate with the baseline selection experienced by the secondary mutation. For mutations already experiencing strong selection in a wild-type genotype, synergistic effects tended to be small, approaching a three-fold increase in selection. For mutations experiencing weaker baseline selection, synergistic epistasis typically exhibited larger effects, approaching a six-fold increase in selection.

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

In ES-LUAD, antagonistic sign epistasis–where selection for a mutated gene changes from positive to negative in the presence of another mutated gene—was inferred for 215 of 420 pairwise orderings of the 21 mutated genes. For mutations of three of these 21 genes, selection in one or more specific mutant genotypes was both significantly lower than selection from wild-type and also significantly negative: mutations of *CTNNB1* and *GNAS* were accordingly antagonized by mutations of *MET* and *APC*, mutations of *SMAD4* were accordingly antagonized by mutations of *BRCA2* and *PIK3CA*, and mutations

of *MET* were accordingly antagonized by mutations of *RB1*, *EGFR*, *CTNNB1*, and *GNAS*. Additionally, each of these instances of antagonistic sign epistasis was also statistically significant in the opposite direction. For instance, the presence of *CTNNB1* mutations induced negative selection for *APC* mutations. For over a hundred other ordered pairs, antagonistic sign epistasis was estimated, without reaching statistical significance in our analysis.

Antagonistic shifts in the magnitude of selection were identified in sixteen cases, of which two were significant: prior *KRAS* mutations caused a 73% reduction in selection for mutations of *TP53*, and an 85% reduction in selection for mutations of *EGFR* (**Fig. 4B**). *TP53* mutations did not substantially alter selection for mutations of *KRAS*, while *EGFR* mutations exhibited antagonistic sign epistasis toward mutations of *KRAS*.

Mutations of *TP53* and *KRAS* were antagonized by mutations in the majority of other genes. Other genes did not show such widespread antagonism. Prior mutations in *BRAF*, *KEAP1*, *STK11*, *SMARCA4*, *RBM10*, and *MGA* antagonized selection for new mutations in *TP53* and *KRAS*. *TP53* mutations were also antagonized by prior mutations in *KRAS*, *ARID1A*, *MET*, and *ATM*. *KRAS* mutations were antagonized by prior mutations in *EGFR*, *PIK3CA*, *CTNNB1*, *BRCA2*, *SETD2*, *RB1*, *ALK*, and *APC*. Indeed, in no case were new mutations in *TP53* and *KRAS* inferred to experience stronger selection in a mutant tumor genotype relative to a wild-type genotype in ES-LUAD or NS-LUAD.

Antagonistic epistasis is more frequent in NS-LUAD

In comparison to ES-LUAD, we observed more instances of antagonistic epistasis and fewer instances of synergistic epistasis in NS-LUAD (**Fig. 4C,E**). Of the 46 instances of significant antagonistic epistasis identified in ES-LUAD, 16 were also statistically significantly antagonistic in NS-LUAD. Another 24 of the 46 total were estimated as antagonistic but not with statistical significance in NS-LUAD (**Fig. 4D–E**). Thus, there was substantial consistency of selective antagonism in ES-LUAD with selective antagonism in NS-LUAD.

This consistency contrasted with a substantial inconsistency between selective synergy in ES-LUAD with selective synergy in NS-LUAD. Of the 72 instances of significant synergistic epistasis identified in ES-LUAD, only eight were identified as significantly synergistic in NS-LUAD. Another 38 of the 73 were estimated as synergistic but not with statistical significance. Twenty-two of these instances were inferred to be significantly antagonistic, with a majority of the antagonism affecting mutations of *BRCA2*, *GNAS*, and *SMARCA4*. For example, mutation of *KEAP1* was inferred to antagonize selection for mutation of *BRCA2* in NS-LUAD, despite synergistically increasing selection for mutation of *BRCA2* in ES-LUAD (**Fig. 4D–E**). Overall, the strength and sign of pairwise epistasis differed substantially between ES-LUAD and NS-LUAD, with greater consistency among antagonistic epistasis and less consistency among synergistic epistasis.

Fewer examples of synergistic epistasis were observed in NS-LUAD compared to ES-LUAD. However, the degree of synergy was remarkable: the 17 instances of statistically significant synergistic epistasis were associated with a median ten-fold increase in selection. The number of NS-LUAD tumors examined was lower than the number of ES-LUAD tumors; accordingly, the uncertainty on all estimates was greater than in ES-LUAD. Nevertheless, synergistic epistasis contributed to significantly stronger selection on several mutations, including mutation of *EGFR* after mutation of *TP53*, mutation of *CTNNB1* after *EGFR*, mutation of *BRCA2* after mutation of *SMARCA4* or *ARID1A*, mutation of *STK11* after *KEAP1*, mutation of *SETD2* after *ARID1A*, and mutation of *APC* after *SMAD4* (**Fig. 4C**).

Pairwise epistasis is sometimes asymmetric in sign and magnitude

In both ES-LUAD and NS-LUAD, some epistatic relationships were markedly asymmetric, indicated by either different magnitudes of effect or different signs of effect depending on which mutation occurred first. For example, in ES-LUAD, primary mutations in *KEAP1* or *STK11* were both inferred to increase selection for secondary mutation of the other gene, but apparently to different extents: primary *STK11* mutation was estimated to increase selection for a secondary *KEAP1* mutation by 60%, while

primary *KEAP1* mutation was estimated to statistically significantly increase selection for a secondary *STK11* mutation by over 300%. Mutations of *KRAS* and mutations of *KEAP1* exhibited not just different magnitudes of the same direction of epistatic effect, but rather exhibited opposite signs of epistatic effect on each other in ES-LUAD. Primary mutation of *KRAS* synergistically increased selection for a secondary *KEAP1* mutation. However, primary mutation of *KEAP1* strongly antagonized selection for a secondary *KRAS* mutation (**Fig. 4B**). This asymmetry indicates 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*. Many other pairs of mutated genes exhibited such asymmetries, including nine pairs for which—like *KRAS* and *KEAP1*—the direction of epistatic effect depended on the ordering of mutations.

Epistasis and environmental interactions can influence selection in combination (G × G, G × E, and G × G × E)

Mutations in the 21 reputed driver genes experienced a broad range of epistatic interactions and microenvironmental interactions, as well as combinations thereof. Some mutations are unaffected by mutations in some genes and by microenvironment, such as mutation of *RB1*, whose selection is not affected by prior mutation of *KRAS*, nor directly affected by a history of tobacco smoking (**Fig. 5A**) Nevertheless, selection for mutation of *RB1* is affected by prior mutation of *TP53* (**Fig. 4B**). Indeed, *RB1* is not unusual in being subject to selective epistasis as a consequence of mutation of some genes but not others: every tested mutation was statistically significantly affected by prior mutation in at least one other gene, demonstrating the pervasiveness of epistasis in tumor evolution.

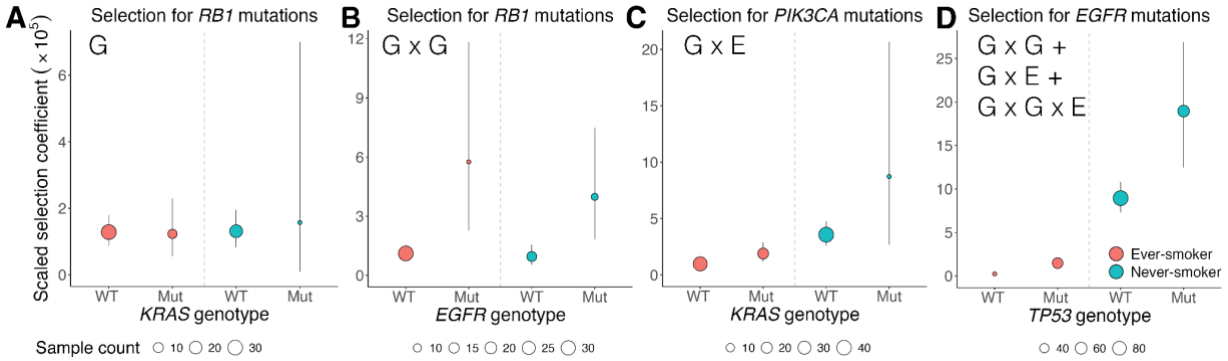


Figure 5. Four patterns of the influences of genetic and environmental interactions in ever-smoker and never-smoker lung adenocarcinoma. (**A**) *RB1* mutations with wild-type (WT) or mutant *KRAS*, (**B**) *RB1* mutations with wild-type or mutant *EGFR*, (**C**) *PIK3CA* mutations with wild-type or mutant *KRAS*, and (**D**) *EGFR* mutations with wild-type or mutant *TP53*.

A number of mutations were substantially influenced by epistasis and were trivially influenced by the microenvironment. For example, *EGFR* mutations antagonized selection for *RB1* mutations regardless of whether the mutations occurred in an ES- or NS-LUAD (**Fig. 5B**). Likewise, the synergistic effect of *EGFR* mutations on selection for *RBM10* mutations and the antagonistic effect of *SMAD4* mutations on selection for *BRCA2* mutations were similar in magnitude between both microenvironments. In contrast, certain pairs of mutations featured only an environmental influence: selection for *PIK3CA* mutations was significantly stronger in NS-LUAD and was not significantly altered by prior *KRAS* mutation (**Fig. 5C**). Similarly, selection for mutant *MET* was stronger in NS-LUAD and was not altered by prior mutation of *TP53*.

Frequently, selection on mutations was influenced both by the microenvironment and by epistasis. Indeed, all mutations that were significantly differentially selected between ES- and NS-LUAD interacted with other mutations that further influenced their selection (**Fig. 2A, Fig. 4D–E**). Furthermore, the strength of epistasis was sometimes itself influenced by the microenvironment, creating a gene-by-gene-by-environment interaction. For example, *TP53* mutations synergistically increased selection for *EGFR* mutations in both microenvironments. However, the synergistic effect was stronger in ES-LUAD,

wherein it led to a six-fold increase in selection compared to a two-fold increase in selection in NS-LUAD (**Fig. 5E**).

Gene-by-gene-by-environment interactions like this one were also observed between *KEAP1* and *STK11* mutations and *SMAD4* and *MET* mutations among other potential instances, contributing to further differentiation of the evolutionary trajectories of ES- and NS-LUAD.

Higher-order epistasis guides the evolutionary trajectories of ES-LUAD and NS-LUAD

KRAS, KEAP1, *and* STK11 *constitute an epistatic network in ES-LUAD*

Having previously identified *KRAS*, *KEAP1*, and *STK11* as experiencing strong, ever-smoker-specific selection, we then tested whether these mutations cooperated with or antagonized each other through higher-order epistasis (**Fig. 6**). All three mutations exhibited a high fixation rate from almost any initial somatic genotype, and fixation rates were particularly high for mutations when they led to the state with all three genes mutated (**Fig. 6A**). However, the fixation rates for *KRAS* mutations after either *KEAP1* mutation or *STK11* mutation were significantly lower than the fixation rates for *KRAS* mutations from wild-type.

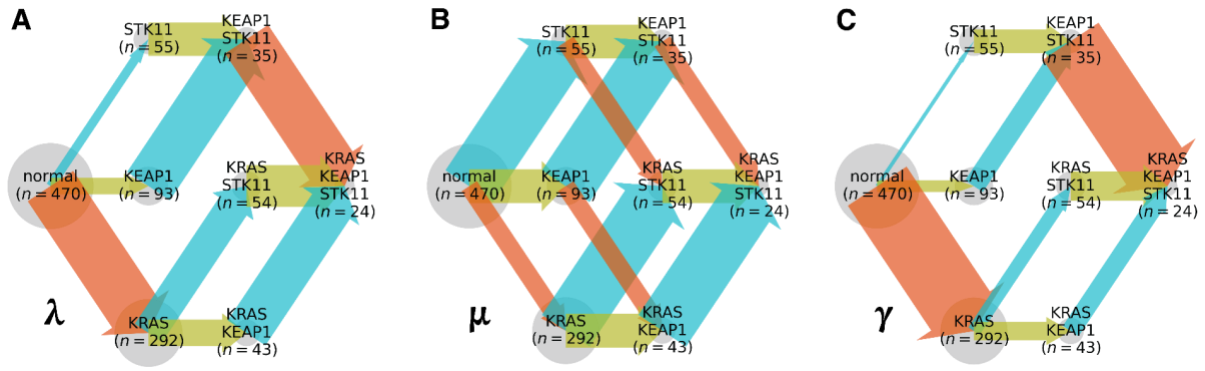


Figure 6. Evolutionary trajectory for *KRAS*, *KEAP1*, and *STK11* mutations in ever-smoker lung adenocarcinoma. (**A**) Fixation rates, (**B**) oncogenic mutation rates, and (**C**) scaled selection coefficients for mutations of *KRAS* (orange arrow; arrow width proportional to rate or coefficient), *KEAP1* (light blue arrow), and *STK11* (yellow arrow) conditioned on somatic genotype (gray circle; radius proportional to genotype prevalence).

Based on analysis of synonymous-site variation and mutation rate covariates, *STK11* manifested the highest oncogenic mutation rate of the three genes, followed by *KEAP1*, and then *KRAS* (**Fig. 6B**). Given these differential underlying rates of mutation, we computed the consequent pairwise and higher-order selective epistasis. Many samples possessed mutations in at least one of these genes (**Fig. 6C**, second layer). Consequently, our analysis of this three-gene model was well powered to detect pairwise epistasis. Indeed, this analysis recapitulated the pairwise epistatic effects inferred in the respective pairwise mutation models (**Fig. 4D**). There was a relative paucity of samples possessing concurrent mutations in two or more of the genes (**Fig. 6C**, third and fourth layers). Consequently, our analysis was less well powered to detect statistical significance of higher-order epistasis. For example, extensive uncertainty in the estimated selection experienced by *KRAS* mutations in an *STK11*-*KEAP1*-mutant genotype hindered ascertainment of some aspects of higher-order epistasis. However, because we had higher power to estimate selection for *STK11* in a *KRAS*-*KEAP1*-mutant genotype, we were able to make confident inferences about the nature of higher-order epistasis affecting mutations of *STK11*.

For example, *STK11* mutations experienced significantly stronger selection in the context of

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, the selection experienced by *STK11* mutations in a *KRAS-KEAP1*-mutant genotype, 3.9×10^5, was comparatively similar to the selection experienced by *STK11* mutations in a *KEAP1*-mutant genotype (4.4×10^5) or in a *KRAS*-mutant genotype (2.8×10^5; **Fig. 6C**). 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.

Pairwise epistatic effects typically compound in a manner consistent with minimal higher-order epistasis

The substantially sub-additive epistasis affecting mutations of *KRAS*, *KEAP1*, and *STK11* demonstrates that assessments of higher-order epistasis will be necessary to thoroughly understand cancer evolution. However, pairwise epistatic effects typically compound consistent with minimal higher-order epistasis. Examining all 1330 triads out of the 21 genes, we assessed epistasis when one to two of the triad were mutated. Pairwise epistatic effects were sufficient to explain most cases in which selection for a mutation was significantly different between genotypes containing one versus two mutations in other genes. In cases where selection for a mutation in a first gene was synergistically increased by an extant mutation in a second gene, and was unaffected by an extant mutation in a third gene, selection for the mutation of the first gene was commensurately synergistically increased when the second and third gene were concurrently mutated. For example, in ES-LUAD, selection for mutations in *APC* was synergistically increased by prior mutations in *SMARCA4*, and was unaffected by prior *KRAS* mutations, and correspondingly, selection for mutations in *APC* was synergistically increased in a *SMARCA4-KRAS*-mutant genotype, with no significant difference in selection for mutations in *APC* compared to a *SMARCA4*-mutant *KRAS*-wildtype genotype (**Fig. 7A,C**). Other instances of the same phenomenon in

ES-LUADs include selection for mutations of *SMARCA4* in a *KRAS-KEAP1*-mutant genotype and for mutations of either *MGA* or *ATM* in a *TP53-KRAS*-mutant genotype (**Fig. 7A**). In NS-LUADs, a similar set of epistatic relationships was observed for mutations of *RB1* in a *TP53-EGFR*-mutant genotype (**Fig. 7B**).

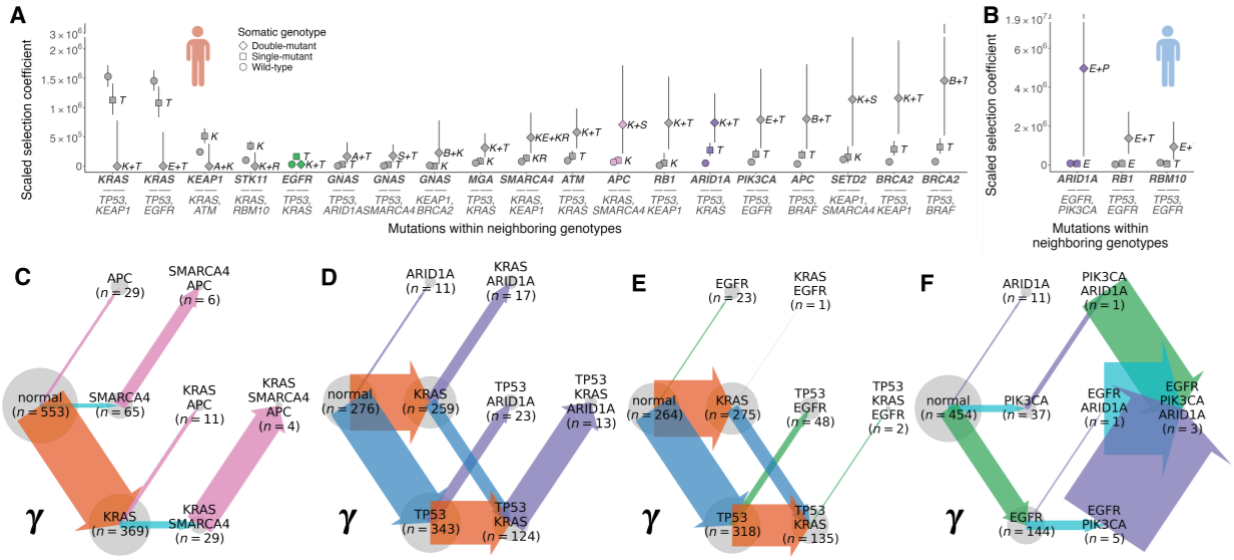


Figure 7. Higher-order selective epistasis in lung adenocarcinoma (LUAD). Selection for a mutation in a context of no co-occurring mutations (circle), one co-occurring mutation (square), and two co-occurring mutations (diamond) in (**A**) ever-smoker (ES) and (**B**) never-smoker (NS) LUAD. Pairwise and higher-order epistasis in ES-LUAD: (**C**) 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); (**D**) 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 (in this panel, all arrow widths are scaled down compared to other panels by a factor of two).

Another generality is that when selection for a mutation in one gene was antagonized by a mutation in a second gene and was unaffected by a mutation in a third gene, then selection for the mutation was still similarly antagonized when the second and third gene were concurrently mutated. For example,

selection for *KRAS* mutations was antagonized by prior *EGFR* or *KEAP1* mutations and was unaffected by prior *TP53* mutations, and correspondingly, selection for *KRAS* mutations was antagonized by a *TP53*-*EGFR*-mutant genotype and by a *TP53*-*KEAP1*-mutant genotype (**Fig. 7A**). There were no instances in which statistically significant pairwise epistasis between two mutated genes failed to manifest due to the concurrent presence of a third mutation that was neutral in terms of its epistatic effects on selection for their first mutated gene.

When two co-occurring mutations each exhibited individual selective synergy with a new mutation, the synergistic effects often stacked together, strongly increasing selection for it. For example, selection for mutations in *ARID1A* was synergistically increased in both a *KRAS*-mutant genotype and a *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. 7D**). Compounded synergistic effects were also inferred for mutations of *BRCA2* in a *KEAP1*-*TP53*-mutant genotype and a *BRAF*-*TP53*-mutant genotype and for mutations of *APC* in a *BRAF*-*TP53*-mutant genotype. Mutations experiencing compounded synergistic epistasis experienced up to an order of magnitude stronger selection in the double-mutant genotype relative to the normal somatic genotype. This substantial compounding constitutes a crucial element of how driver mutations accumulate in cancer genomes, contributing to a positively reinforced progression of cancer toward more aggressive forms.

Higher-order epistasis results in differential selective benefit in somatic genotypes featuring two or more mutations

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. 7A**). Similarly, some 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 selection for a mutation could be assessed in a genotype with two statistically significantly synergistic mutations, the maximum-likelihood estimates of selection were lower than would be expected from additive synergistic epistasis.

We observed higher-order epistasis involving the co-occurrence of synergistic and antagonistic mutations, as well as the emergence of synergy in the co-occurrence of neutrally-interacting mutations. As an example of the former, in ES-LUAD, selection for *EGFR* mutations was synergistically increased by *TP53* mutations and 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. 7A,E**). In fact, these *EGFR* mutations experienced levels of selection similar to those they would experience in a normal genotype, 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 7A**). Thus, in a combination of synergistic and antagonistic epistasis, the antagonistic effect tended to dampen, or potentially even overwhelm, the synergistic effect. This dampening resulted in weaker selection for the affected mutation relative to its selection in the presence of a synergistic mutation alone.

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. Among cases where two mutations individually do not have antagonistic effects toward a third mutation, we observed no instances in which selection for that third mutation was statistically significantly antagonized by their co-occurrence. We did, however, observe at least one instance in which

selection for a mutation was statistically significantly enhanced by the co-occurrence of two mutations that did not individually have synergistic effects: in NS-LUAD, selection for mutations in *ARID1A* increased by an order of magnitude in *EGFR-PIK3CA*-mutant genotypes, despite not being significantly altered when only one of the two genes was mutated (**Fig 7A**,**F**). This case demonstrates the potentially rare but important possibility that epistasis may only emerge from the combinations of mutations. Such emergent epistasis would be obscured in analyses restricted to pairwise epistasis, emphasizing the importance of considering multi-mutant genotypes.

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 analysis of the adaptive landscape of 21 mutations has strongly supported only a subset of the many potential instances of higher-order epistasis between oncogenic mutations in LUAD. Multiple other instances of higher-order epistasis were estimated to be present, but did not reach statistical significance with accessible tumor sequence sample sizes.

DISCUSSION

Here we have characterized the adaptive landscapes of lung adenocarcinoma in ever-smokers and never-smokers by estimating underlying gene-specific oncogenic mutation rates, the intensity of somatic selection acting on those mutations, and the strength of selective epistasis. We have shown that the mutational and adaptive landscape of ES-LUAD and NS-LUAD differ substantially. These differences manifest from the mutagenic effects of smoking, the environmentally driven differences in the selective benefit of mutations in isolation, and the differences in the sign and magnitude of epistasis.

Consistent with reported mutagenic effects of tobacco carcinogens (*4*), we detected elevated oncogenic mutation rates in all tested genes in ES-LUAD compared to NS-LUAD. In addition, we find strong evidence for differential selection on canonical driver mutations in ES- and NS-LUAD. These differences are likely due to the direct damage or downstream physiological effects of smoking on the tumor microenvironment. Rodin and Rodin (*13*) reported roughly equivalent background mutation rates for *TP53* in lung cancer between smokers and nonsmokers, and attributed the higher frequency of *TP53* mutations in smokers to increased selection in the smoking-altered microenvironment. Our mutation rate inference used several times more sample mutation data, incorporated tissue-specific covariates, and relied on substantially improved methods to estimate background mutation rates (*6*), and it revealed a higher background mutation rate for *TP53* in ES-LUAD and consequently that mutations of *TP53* experience weaker selection in a smoking-altered microenvironment. We also found that scaled selection coefficients for *KRAS*, *KEAP1*, and *STK11* mutations were significantly greater in ES-LUAD than in NS-LUAD, indicating that these mutations are more adaptive in the smoking-affected tumor microenvironment. In contrast, mutations in *EGFR*, *PIK3CA*, and *SMAD4* experience stronger selection in NS-LUAD, suggesting that these mutations are more strongly adaptive in a comparatively healthy lung microenvironment. Such differences in selection support current hypotheses that the adaptive landscape of oncogenesis is altered by the physiological insults contributed by endogenous and exogenous factors, such as age, tobacco smoke, alcohol, and radiation (*14*, *60–62*). Based on these results, we should expect

that, for example, the clinical benefit of receiving *EGFR* inhibitors should be greater in never-smokers than ever-smokers. Indeed, a retrospective analysis of patients with *EGFR*-mutant LUAD found that the objective response rate (ORR) of patients receiving EGFR tyrosine kinase inhibitors (TKIs) decreased significantly as tobacco smoking history increased: patients with no smoking history had an ORR of 73% compared to 46% in patients with over 10 pack years of smoking history (*63*). Higher response rates to EGFR TKIs among never-smokers compared to heavy-smokers were also demonstrated in multiple other retrospective analyses of LUAD (*64*) and non-small cell lung cancer (NSCLC; *65*, *66*). These results demonstrate the utility of selection coefficients as predictors of the efficacy of targeted therapies in patient populations that have been subjected to distinct environmental toxicities influencing the tumor microenvironment. Indeed, exposure-influenced selection coefficients on a therapeutically targeted gene can identify likely responders to targeted therapies, increasing the power for clinical trials to observe a treatment benefit.

Our analysis revealed the widespread presence of pairwise epistasis between driver mutations, with effects ranging from antagonistic sign epistasis to multi-fold increases in selection. Nearly all tested mutations are significantly antagonistic toward *TP53* and *KRAS* mutations. *TP53* and *KRAS* mutations are frequent in LUADs. Consequently, this widespread antagonism indicates a greater contribution of *TP53* and *KRAS* mutations to early cancer development, compared to late cancer development when other mutations have already accumulated. Indeed, every driver gene examined exhibited evidence of epistasis occurring in multiple pairings. For each gene, at least one of these epistatic pairings was significant in ES-LUAD, NS-LUAD, or both.

Many of the significant pairings can be linked to experimental and clinical support. In accordance with the inferred synergistic effect of *KRAS* mutations on mutations of *KEAP1* and *STK11*, *KEAP1* loss accelerates LUAD progression in *KRAS*-mutant mouse tumors (*67*, *68*) and compensates for the deleterious effects of *KRAS*-*STK11*-co-mutation (*69*), and mutations of *KEAP1* and/or *STK11* worsen survival and treatment response in *KRAS*-mutant LUAD (*70*, *71*) and NSCLC (*72–77*). 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 (*78*) and worsen survival in NSCLC independently of *KRAS* mutation status (*79*, *80*). Furthermore, we found that *EGFR* mutations synergized with mutations of *BRAF* in ES-LUAD, mutations of *CTNNB1* in NS-LUAD, and mutations of *RB1* in both ES- and NS-LUAD. Accordingly, experimental and clinical studies have demonstrated that *BRAF* mutations drive EGFR TKI resistance in *EGFR*-mutant cancers (*81*), *CTNNB1* mutations are overrepresented in advanced *EGFR*-mutant NSCLC, where they are associated with metastasis and EGFR TKI resistance (*82*), and *RB1* mutations preferentially occur in *EGFR*-mutant LUAD, where they mediate histological transformation from LUAD to small cell lung cancer (*83*, *84*). Most of the selective epistatic relationships we have determined remain experimentally untested and therefore represent promising research directions for the discovery of direct and indirect interactions between mutant proteins that may be relevant to carcinogenesis and treatment.

The frequent presence of strong epistasis in LUAD identifies numerous opportunities for precision targeting of therapy. For instance, strong epistasis implies utility of co-occurring mutations as a biomarker. *EGFR* mutations experience stronger selection in a *TP53*-mutant genotype than a wild-type genotype. In this case, inhibition of EGFR would in principle eliminate both the baseline effect of the *EGFR* mutation and also diminish the effect of the concurrent *TP53* mutation that is mediated through the *EGFR* mutation. Accordingly, *EGFR* mutations would be expected to be more effectively targeted by EGFR TKIs in *TP53*-mutant LUAD. By the same logic, *EGFR* mutations occurring within antagonistic *KRAS*- or *KEAP1*-mutant genotypes would be less tumorigenic and would be expected to be poorly targeted by EGFR-directed therapies. Indeed, patients with *EGFR*-mutant LUAD treated with an EGFR TKI exhibited significantly decreased survival if they had co-occurring *KRAS* mutations (*85*). Moreover, the presence of *KEAP1* mutations decreases the efficacy of EGFR tyrosine kinase inhibitors in *EGFR*-mutant lung cancer (*86*, *87*). Because *TP53* mutations synergize with mutations of *EGFR*, we would expect based on the selective epistasis alone that TP53 mutation would be negatively associated with

resistance to EGFR TKIs. In contrast with this expectation, TP53 mutation is positively associated with resistance to EGFR TKIs in *EGFR*-mutant LUAD (*88–90*) and NSCLC (*91–95*). *TP53* loss is known not only to be strongly adaptive in lung cancer, but also to increase the rate of somatic nucleotide mutations (*90*) and copy-number alterations (*88*, *96*)—providing greater genetic variation that facilitates the evolution of resistance—and to alter gene expression programs relevant to resistance (*92*) and cancer progression (*96*). On the other hand, antagonistic epistasis can be indicative of synthetic lethality. For instance, *BRCA2* and *SMAD4* mutations are negatively selected in the presence of the other. Therefore, direct or indirect inhibition of *SMAD4* activity may be a potential treatment for *BRCA2*-mutant tumors or vice-versa. In general, selective epistasis can explain the heterogeneity of response to targeted therapies (*44*, *97*), and can be exploited to identify subpopulations of likely responders and nonresponders to targeted therapy (*98*, *99*), providing a means to substantially increase the power of clinical trials to detect therapeutic benefit from new drug candidates.

Networks of inferred epistatic interactions are known to be substantially different between cancer types (*100*). It is also known that selective epistatic interactions can change with environmental conditions (*101–106*). However, extensive differences in selective epistasis has not previously been demonstrated within a cancer type experiencing different selective environments. We have demonstrated that selective epistasis can markedly vary within a cancer type between patients with different environmental exposures to tobacco smoke. ES-LUAD and NS-LUAD featured distinct epistatic networks. Large-scale differences were evident: synergistic epistasis was relatively frequently detected in ES-LUAD and very infrequently detected in NS-LUAD. In a few cases, epistasis between driver genes was synergistic in ES-LUAD but antagonistic in NS-LUAD. In other cases, the strength of epistasis was different: for example, the synergistic effect of *TP53* mutations on *EGFR* mutations was stronger in ES-LUAD relative to NS-LUAD. Antagonistic epistasis was widespread between drivers in NS-LUAD and was infrequent in ES-LUAD. The higher frequency of synergism and lower frequency of antagonism in ES-LUAD relative to NS-LUAD suggests that the adaptive landscape of ES-LUAD is more accessible

than the landscape of NS-LUAD. Such increased accessibility enables ES-LUADs to develop along a greater diversity of evolutionary trajectories, manifesting in increased heterogeneity of the disease between patients. It may also help to explain—beyond carcinogenic mutation per se—the profound extent to which tobacco smoking elevates the risk of lung cancer beyond solely carcinogenic mutagenesis. Within NS-LUAD, the widespread prevalence of antagonistic epistasis may be explained by a higher fitness cost of increased mutation load. This fitness cost could be due to differences in tumor immune microenvironment between never-smokers and ever-smokers (*12*, *24*, *25*, *107*) or due to a lower tolerance for mutation accumulation in the fragile lung tissue cells of never-smokers. Taken together, these results highlight gene-by-gene-by-environment interactions that alter the patterns of selective epistasis in ES-LUAD and correspondingly inform the efficacy of epistasis-based treatment strategies.

Previous computational studies of pairwise interactions have revealed asymmetric magnitudes of pairwise epistasis (cf. *37*). Our analysis has demonstrated the presence of widespread epistasis with not only asymmetric magnitudes but also asymmetric signs of selection—epistasis in which one ordering of the mutations is advantageous while the reverse is disadvantageous. These asymmetries may be explained by cellular transitions induced by the primary mutation that unidirectionally alters selection for the secondary mutation, and such asymmetry would not be evident from measures of mutual exclusivity or co-occurrence. Indeed, the presence of such asymmetries of epistasis have been extensively supported by genetic and computational genomic studies of *TP53* and *CCNE1* mutations (*39*) and *JAK2* and *TET2* mutations (*108*). Asymmetric epistasis further implies the existence of preferential mutation orderings that would, in combination, yield predictable trajectories of cancer evolution.

It has often been assumed that the adaptive landscape of cancer can be well described without epistasis (*6*, *109–113*) or with only pairwise epistasis (*38*, *39*, *114*). However, our results suggest that not only pairwise but also higher-order epistasis strongly influences the adaptive value of cancer drivers. For example, in NS-LUAD, selection for *ARID1A* mutations increased by an order of magnitude in an *EGFR*-*PIK3CA*-mutant genotype despite a lack of statistically significant synergistic pairwise epistasis with

either mutated gene individually. Another form of higher-order epistasis manifested in ES-LUAD via antagonistic mutations in *KRAS*, which dampened the synergistic effects of *TP53* mutations on *EGFR* mutations, and via antagonistic mutations in *KEAP1*, which undermined the synergistic effects of *TP53* mutations on *EGFR* mutations. Furthermore, our results suggest multiple potential instances of super-additive synergistic epistasis and the potentially widespread prevalence of sub-additive synergistic epistasis, a finding consistent with the diminishing-returns epistasis often observed in asexually reproducing populations (*51–54*). Knowledge of higher-order epistasis is especially important in tumors with a large mutation load—such as those commonly afflicting patients with a long smoking history—for the prediction of the importance of somatic variants to cancer and accordingly for prediction of the effects of variant-targeted therapeutics. Higher-order synergistic epistasis, in particular, renders some evolutionary trajectories more likely than others, while higher-order antagonistic epistasis prunes the set of likely evolutionary trajectories. Both make tumor evolution more predictable—a property that should guide the development of therapy. Therefore, it is vital that future efforts to characterize the adaptive landscape of cancer incorporate analyses of the influence of higher-order epistasis.

This study was subject to several limitations. This analysis was retrospective. Therefore, covariates with smoking behavior—such as health states, environmental exposures, inherited genetics, and sex biases—likely contribute to the observed differences in the adaptive landscapes of ES-LUAD and NS-LUAD. However, none of these factors have a comparable impact on lung physiology. Consequently, our reported selective differences can primarily be attributed to the direct and downstream effects of smoking. Smoking status in our study was classified according to mutational signatures in tumor sequences, which can be subject to misclassification. However, such methods have been validated for use in classifying smoking status (*115*) self-reported status have been found to be unreliable (*116*, *117*) and often do not account for the risk conferred by secondhand smoke (*118*). Overall, because our approach explicitly quantifies the smoking-associated mutational burden in the lung periphery where LUAD develops, it also implicitly distinguishes patients who have and have not experienced significant

physiological effects of smoke in the lung periphery. In both ES-LUAD and NS-LUAD, our estimates of selective epistasis assumed that oncogenic mutation rates were independent of the somatic genotype. However, mutations in *TP53* and *BRCA2* are known to increase somatic mutation rates (*119*, *120*), and such increases in mutation rate later in the oncogenic trajectory could lead to inflation of inferred selection for mutations occurring later in the trajectory. This effect warrants further investigation, but would likely not influence our findings on the differences in somatic adaptive landscapes between smokers and non-smokers. 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 (*121*) and pairwise selective epistasis has been indicated between SNVs and CNAs and other SVs (*122*, *123*) or between SVs (*44*, *100*, *124*). Unfortunately, accurate background mutation rates are currently unavailable for SVs. If such rates are calculable, integrated analysis of SNVs and SVs will reveal new epistatic relationships and clarify the somatic adaptive landscape of lung adenocarcinoma.

These results contribute evidence for substantial differences in the adaptive landscapes of ES-LUAD and NS-LUAD, suggesting major effects of smoking on cancer risk beyond mutagenesis. We also demonstrate that epistasis is pervasive and powerful in lung adenocarcinoma, with a strong influence on the divergent evolutionary trajectories of ES-LUAD and NS-LUAD, mediated through both symmetric and asymmetric epistasis. Furthermore, we provide a systematic characterization of higher-order selective epistasis in LUAD, reporting its underappreciated role in cancer evolution. This study demonstrates that the contribution of mutations to tumor growth is heavily influenced by the genetic and environmental context in which it arises. This context illuminates new opportunities for precision oncology through the informed application of targeted therapies according to individual patient somatic genotypes and environmental exposures. Further study into genetic and environmental interactions in lung adenocarcinoma and other cancers will advance our understanding of cancer evolution and improve our ability to tailor cancer treatments to individual patients.

MATERIALS AND METHODS

Experimental Design

The objective of this study was to test for differences in mutagenic rates, scaled selection coefficients, and gene-by-gene epistatic effects between lung adenocarcinoma in ever-smokers (ES-LUAD) and never-smokers (NS-LUAD). Publicly available mutation data for 8,868 LUAD tumor samples was aggregated from multiple institutional sources (see *Data sources*), and samples were classified as either ES-LUAD or NS-LUAD based on the presence or absence of the smoking-associated mutational signature SBS4 in exome or genome sequences or, for targeted sequences, based on the patient's self-reported smoking history (see *Mutational signature attributions*). Established data-driven approaches were used to compute mutagenic rates (see *Estimation of oncogenic mutation rates)*, selection, and epistasis (see *Estimation of scaled selection coefficients and selective epistasis*) within the cohort of ES-LUAD samples (n = 1,066) and the cohort of NS-LUAD samples (n = 657). Non-overlap of 95% confidence intervals was used to conservatively assess statistical significance for selective differences between ES-LUAD and NS-LUAD.

Data sources

Clinicogenomic data was sourced from (i) eight published datasets hosted by cBioPortal (*125–133*), (ii) The Cancer Genome Atlas and Foundation Medicine Adult Clinical Cancer datasets hosted on Genomic Data Commons (*134, 135*), (iii) AACR Project Genie release 9.0 (*136*), and (iv) 108 samples sequenced at Yale (*137*). No studies were included that specifically selected a population based on their mutational profile. Included samples were sourced only from primary sites of LUAD. If a study submitted multiple samples from the same patient, one was chosen by random sampling, and if a patient was included in multiple studies, the sample with the highest quality sequence data and clinical annotations was used. A breakdown of datasets and associated studies, sample counts, sequencing methods, treatment information, and use in analysis is provided in **Supplementary Table 1**.

Genomic coordinates for all mutations were lifted over to the hg19 reference assembly using the liftOver function of rtracklayer v1.62.0(*138*). The R package cancereffectsizeR v2.10.1 was used to filter out variants that failed genome lifting, were misclassified as single-nucleotide variants (SNVs), or were otherwise invalid for analysis. For panel-sequenced samples, regions covered by each panel were either obtained from direct panel sequencing information (AACR GENIE) or inferred from the list of genes included in each panel using their hg19 coordinates (cBioPortal & Foundation Medicine).

Mutational signature attributions

Sample-specific signature attributions were inferred for whole-exome or whole-genome sequences, but not panel-sequence data, because signature attribution is less reliable when there are few silent variants identified. Signature attribution was performed in cancereffectsizeR using MutationalPatterns v3.12.0(*139*), specifying COSMIC v3.2 single-base substitution signatures (*31*, *140*). COSMIC signatures that were unlikely to have been observed in lung adenocarcinomas (*140*) were excluded from consideration. COSMIC signatures with treatment-based etiologies were excluded from signature deconvolutions of samples known to be treatment-naive. Tumor samples in which the SBS4 signature was detected above the default MutationalPatterns threshold of 5% of the total signature contributions were classified as ever-smoker LUAD (ES-LUAD), and otherwise were classified as never-smoker LUAD (NS-LUAD). Zhang et al. detected no signature of SBS4 in the samples used in their study; therefore these samples were automatically classified as NS-LUAD. As in Rosenthal et al. (*141*), samples whose whole-exome sequencing revealed fifty or fewer single-nucleotide variants (SNVs) were excluded from classification into ES-LUAD or NS-LUAD due to imprecision of signature deconvolution. Depending on the dataset, our mutation signature-based classifications ranged from 67–90% in their consistency with clinical smoking status annotations, with differences likely due to significant contributions of secondhand smoke, ever-smoker reporting associated with a light smoking history with minimal mutational and physiological consequences, or misreporting of smoking history. Panel-sequenced samples were classified

according to clinical annotations of smoking history whenever available. Panel-sequenced samples annotated as exposed to “former light” smoking history were not included in this classification and were excluded from further analysis.

Estimation of oncogenic mutation rates

Gene-specific oncogenic mutation rates were calculated as in Alfaro-Murillo et al. (*33*) for the ES-LUAD and NS-LUAD cohorts. To do so, we defined oncogenic variants of the gene as those that were observed at least once in any sample included in the analysis. The mutation rate of an oncogenic variant at a certain site within a gene was calculated as the product of three factors: the rate at which any variant, oncogenic or not, could arise in the gene, the probability that this arising variant would be of the same trinucleotide mutation type as the assessed oncogenic variant, and the probability that this arising variant would occur within a specific site among all corresponding trinucleotide contexts within the gene.

The rate at which any variant could occur in the gene was obtained from cancereffectsizeR (*32*) based on expected synonymous mutation counts using lung-tissue-specific covariates describing expression levels, chromatin marks, and replication timing as in dNdScv (*6*). These rates were calculated separately for the ES-LUAD and NS-LUAD cohorts.

The probability of the arising variant matching the trinucleotide mutation type of the oncogenic variant was measured as the proportional frequency of that trinucleotide mutation type in the genomes of all samples in the cohort.

The probability of the arising variant occuring at the same site was quantified as the inverse of the number of sites within the gene with the same trinucleotide context as the assessed oncogenic variant.

The oncogenic mutation rate of the gene is lower than the mutation rate of the gene, and was composed as the sum of the mutation rates of all oncogenic variants.

Estimation of scaled selection coefficients and selective epistasis

For a set of 1,200 genes compiled from previous literature (*134*, *142–145*), we quantified the effects of mutations on the growth and survival of cancer cell lineages by calculation of scaled selection coefficients. These coefficients are roughly the ratio of the fixation rate for the oncogenic variant—the rate at which the site is mutated and subsequently selected to a high enough frequency to be observed within the cancer cell population—to the background mutation rate for the variant. Precisely, the flux of fixation of the mutant allele was calculated by analytical maximization of the binomial likelihood for the number of samples possessing at least one non-silent SNV in the gene. Scaled selection coefficients for each mutated gene were computed in three cohorts: in all samples (including those whose smoking history could not be determined), in ever-smokers, and in never-smokers. Within each cohort, scaled selection coefficients were estimated from a set of samples with associated whole-exome, whole-genome, or targeted gene panel sequences (TGS). Results were consistent when TGS samples were excluded; inclusion of TGS samples provided greater power to detect statistical significance of differences in selection between cohorts.

From these 1,200 genes, we selected 21 for analysis of selective epistasis. These 21 genes were reputed to be substantial drivers of LUAD based on literature support (*134*, *143*) and by their average strength of selection across genotypes. To quantify selective epistasis, we calculated scaled selection coefficients conditional on the somatic genotype as in Alfaro-Murillo et al. (*33*): first, fixation rates were calculated for mutations in each gene of a set of several genes. These gene-specific fixation rates were calculated for mutations occuring within each somatic genotype, defined here as a combination of mutant and wild-type genes restricted to the tested gene set. This calculation was performed for the set of several genes by maximization of the multinomial likelihood for the number of samples with each somatic genotype using PyMC (*146*). The probability of a somatic genotype was derived from a continuous-time Markov chain model (*33*). Asymmetric 95% confidence intervals were individually estimated for each

fixation rate with respect to the set of the other fixation rates using a theorem from Wilk (*147*). To obtain scaled selection coefficients for each mutation occurring within each somatic genotype, we divided the respective Poisson-corrected fixation rate by the background oncogenic mutation rate of the gene, which was assumed to be consistent across somatic genotypes.

Including both the ever-smoker and the never-smoker cohorts, we evaluated 3,080 models encompassing all combinations of two and three genes from these 21 genes. Because our definition of the somatic genotype only consists of the several genes of interest, we obtained sample counts for each genotype by grouping tumor samples according to their somatic genotype for the genes included in the model and then tallying. In all cases, panel-sequenced samples were only included if the panel sequence results included all of the genes tested in the model.

To classify a mutant somatic genotype as interacting synergistically or antagonistically with a certain mutation, we compared the scaled selection coefficients obtained for a mutation when it occurred in a mutant somatic genotype versus the scaled selection coefficients obtained for the same mutation in a normal somatic genotype, defined herein as having no mutations in the genes that were included in the model. If the scaled selection coefficient for a mutation in gene A when it occurred in a wild-type genotype ($\emptyset \rightarrow A$) was less than the scaled selection coefficient for a mutation in gene A when it occurred in a genotype with a mutation in gene B ($B \rightarrow A + B$), then the epistatic effect of mutant gene B on mutant gene A was deemed to be synergistic. If the scaled selection coefficient for $\emptyset \rightarrow A$ was greater than the scaled selection coefficient for $B \rightarrow A + B$, then the epistatic effect of mutant gene B on mutant gene A was deemed to be antagonistic. We defined antagonistic sign epistasis as cases in which selection for a new mutation switches from positive selection (scaled selection $\gamma > 1$) to negative selection (scaled selection $\gamma < 1$) in the presence of the antagonistic genotype. We defined antagonistic magnitude epistasis as cases in which selection for a new mutation remains positive in the antagonistic genotype but is weaker in magnitude.

Statistical Analysis

Statistical significance for differences in selection was assessed conservatively, using non-overlap of the 95% confidence intervals (CIs) for each estimate. Analysis was performed in Python (3.9.5) and R (4.3.0). A full list of package versions is provided in the following GitHub repository, alongside the software to download and preprocess the data, run the analysis performed in this project, and generate figures: https://github.com/Townsend-Lab-Yale/lung-smoking.

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. L. B. Alexandrov, Y. S. Ju, K. Haase, P. Van Loo, I. Martincorena, S. Nik-Zainal, Y. Totoki, A. Fujimoto, H. Nakagawa, T. Shibata, P. J. Campbell, P. Vineis, D. H. Phillips, M. R. Stratton, Mutational signatures associated with tobacco smoking in human cancer. *Science* **354**, 618–622 (2016).

3. G. P. Pfeifer, M. F. Denissenko, M. Olivier, N. Tretyakova, S. S. Hecht, P. Hainaut, Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene* **21**, 7435–7451 (2002).

4. K. Yoshida, K. H. C. Gowers, H. Lee-Six, D. P. Chandrasekharan, T. Coorens, E. F. Maughan, K. Beal, A. Menzies, F. R. Millar, E. Anderson, S. E. Clarke, A. Pennycuick, R. M. Thakrar, C. R. Butler, N. Kakiuchi, T. Hirano, R. E. Hynds, M. R. Stratton, I. Martincorena, S. M. Janes, P. J. Campbell, Tobacco exposure and somatic mutations in normal human bronchial epithelium. *Nature* **578**, 266 (2020).

5. K. Dasari, J. A. Somarelli, S. Kumar, J. P. Townsend, The somatic molecular evolution of cancer: Mutation, selection, and epistasis. *Prog. Biophys. Mol. Biol.* **165**, 56–65 (2021).

6. I. Martincorena, K. M. Raine, M. Gerstung, K. J. Dawson, K. Haase, P. Van Loo, H. Davies, M. R. Stratton, P. J. Campbell, Universal patterns of selection in cancer and somatic tissues. *Cell* **171**, 1029–1041.e21 (2017).

7. V. L. Cannataro, S. Kudalkar, K. Dasari, S. G. Gaffney, H. M. Lazowski, L. K. Jackson, I. Yildiz, R. K. Das, B. E. Gould Rothberg, K. S. Anderson, J. P. Townsend, APOBEC mutagenesis and selection for NFE2L2 contribute to the origin of lung squamous-cell carcinoma. *Lung Cancer* **171**, 34–41 (2022).

8. C. Tan, J. D. Mandell, K. Dasari, V. L. Cannataro, J. A. Alfaro-Murillo, J. P. Townsend, 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* **166**, 265–269 (2022).

9. A. Mochizuki, K. Shiraishi, T. Honda, R. I. Higashiyama, K. Sunami, M. Matsuda, Y. Shimada, Y. Miyazaki, Y. Yoshida, S.-I. Watanabe, Y. Yatabe, R. Hamamoto, T. Kohno, Passive smoking-induced Mutagenesis as a promoter of lung carcinogenesis. *J. Thorac. Oncol.* **19**, 984–994 (2024).

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

11. H. van der Vaart, W. Timens, N. H. T. Ten Hacken, Acute effects of cigarette smoke on inflammation and oxidative stress: a review. *Thorax* **59**, 713–721 (2004).

12. X. Li, J. Li, P. Wu, L. Zhou, B. Lu, K. Ying, E. Chen, Y. Lu, P. Liu, Smoker and non-smoker lung adenocarcinoma is characterized by distinct tumor immune microenvironments. *Oncoimmunology*

7, e1494677 (2018).

13. S. N. Rodin, A. S. Rodin, Human lung cancer and p53: the interplay between mutagenesis and selection. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 12244–12249 (2000).

14. E. Laconi, F. Marongiu, J. DeGregori, Cancer as a disease of old age: changing mutational and microenvironmental landscapes. *Br. J. Cancer* **122**, 943–952 (2020).

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

16. N. Lopez-Bigas, A. Gonzalez-Perez, Are carcinogens direct mutagens? *Nat. Genet.* **52**, 1137–1138 (2020).

17. P. Wang, S. Sun, S. Lam, W. W. Lockwood, New insights into the biology and development of lung cancer in never smokers—implications for early detection and treatment. *J. Transl. Med.* **21**, 1–16 (2023).

18. A. Grosse, C. Grosse, M. Rechsteiner, A. Soltermann, Analysis of the frequency of oncogenic driver mutations and correlation with clinicopathological characteristics in patients with lung adenocarcinoma from Northeastern Switzerland. *Diagn. Pathol.* **14**, 18 (2019).

19. C. K. Liam, C. Y. Yew, Y. K. Pang, C. K. Wong, M. E. Poh, J. L. Tan, C. I. Soo, T. C. Loh, K. K. Chin, V. Munusamy, Y. S. Liam, N. H. Ibrahim, Common driver mutations and programmed death-ligand 1 expression in advanced non-small cell lung cancer in smokers and never smokers. *BMC Cancer* **23**, 659 (2023).

20. A. M. Chapman, K. Y. Sun, P. Ruestow, D. M. Cowan, A. K. Madl, Lung cancer mutation profile of EGFR, ALK, and KRAS: Meta-analysis and comparison of never and ever smokers. *Lung Cancer* **102**, 122–134 (2016).

21. G. Pintarelli, S. Noci, D. Maspero, A. Pettinicchio, M. Dugo, L. De Cecco, M. Incarbone, D. Tosi, L. Santambrogio, T. A. Dragani, F. Colombo, Cigarette smoke alters the transcriptome of non-involved lung tissue in lung adenocarcinoma patients. *Sci. Rep.* **9**, 13039 (2019).

22. A. Tang, U. Ahmad, A. J. Toth, N. Bourdakos, S. Raja, D. P. Raymond, E. H. Blackstone, S. C. Murthy, Non-small cell lung cancer in never- and ever-smokers: Is it the same disease? *J. Thorac. Cardiovasc. Surg.* **161**, 1903–1917.e9 (2021).

23. M. Dias, R. Linhas, S. Campainha, S. Conde, A. Barroso, Lung cancer in never-smokers - what are the differences? *Acta Oncol.* **56**, 931–935 (2017).

24. W. Luo, Z. Zeng, Y. Jin, L. Yang, T. Fan, Z. Wang, Y. Pan, Y. Yang, M. Yao, Y. Li, X. Xiao, G. Wang, C. Wang, S. Chang, G. Che, L. Zhang, Y. Li, Y. Peng, W. Li, Distinct immune microenvironment of lung adenocarcinoma in never-smokers from smokers. *Cell Rep Med* **4**, 101078 (2023).

25. Q. Sui, J. Liang, Z. Hu, Z. Chen, G. Bi, Y. Huang, M. Li, C. Zhan, Z. Lin, Q. Wang, Genetic and microenvironmental differences in non-smoking lung adenocarcinoma patients compared with smoking patients. *Transl Lung Cancer Res* **9**, 1407–1421 (2020).

26. L. Löfling, A. Karimi, F. Sandin, S. Bahmanyar, H. Kieler, M. Lambe, K. Lamberg, G. Wagenius,

Clinical characteristics and survival in non-small cell lung cancer patients by smoking history: a population-based cohort study. *Acta Oncol.* **58**, 1618–1627 (2019).

27. J. Cho, S. M. Choi, J. Lee, C.-H. Lee, S.-M. Lee, D.-W. Kim, J.-J. Yim, Y. T. Kim, C.-G. Yoo, Y. W. Kim, S. K. Han, Y. S. Park, Proportion and clinical features of never-smokers with non-small cell lung cancer. *Chin. J. Cancer* **36**, 20 (2017).

28. T. Okamoto, Y. Suzuki, T. Fujishita, H. Kitahara, S. Shimamatsu, M. Kohno, Y. Morodomi, D. Kawano, Y. Maehara, The prognostic impact of the amount of tobacco smoking in non-small cell lung cancer--differences between adenocarcinoma and squamous cell carcinoma. *Lung Cancer* **85**, 125–130 (2014).

29. S. A. Kenfield, E. K. Wei, M. J. Stampfer, B. A. Rosner, G. A. Colditz, Comparison of Aspects of Smoking Among Four Histologic Types of Lung Cancer. *Tob. Control* **17**, 198 (2008).

30. 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).

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

32. J. D. Mandell, V. L. Cannataro, J. P. Townsend, Estimation of Neutral Mutation Rates and Quantification of Somatic Variant Selection Using cancereffectsizeR. *Cancer Res.* **83**, 500–505 (2023).

33. J. A. Alfaro-Murillo, J. P. Townsend, Pairwise and higher-order epistatic effects among somatic cancer mutations across oncogenesis. *Math. Biosci.* **366**, 109091 (2023).

34. V. L. Cannataro, S. G. Gaffney, J. P. Townsend, Effect sizes of somatic mutations in cancer. *J. Natl. Cancer Inst.* **110**, 1171–1177 (2018).

35. J. D. Mandell, J. N. Fisk, E. Cyrenne, M. L. Xu, V. L. Cannataro, J. P. Townsend, Not only mutations but also tumorigenesis can be substantially attributed to DNA damage from reactive oxygen species in RUNX1::RUNX1T1-fusion-positive acute myeloid leukemia. *Leukemia* **36**, 2931–2933 (2022).

36. X. Wang, A. Q. Fu, M. E. McNerney, K. P. White, Widespread genetic epistasis among cancer genes. *Nat. Commun.* **5**, 1–10 (2014).

37. J. Iranzo, G. Gruenhagen, J. Calle-Espinosa, E. V. Koonin, Pervasive conditional selection of driver mutations and modular epistasis networks in cancer. *Cell Rep.* **40**, 111272 (2022).

38. L. M. Blair, J. M. Juan, L. Sebastian, V. B. Tran, W. Nie, G. D. Wall, M. Gerceker, I. K. Lai, E. A. Apilado, G. Grenot, D. Amar, G. Foggetti, M. Do Carmo, Z. Ugur, D. Deng, A. Chenchik, M. Paz Zafra, L. E. Dow, K. Politi, J. J. MacQuitty, D. A. Petrov, M. M. Winslow, M. J. Rosen, I. P. Winters, Oncogenic context shapes the fitness landscape of tumor suppression. *Nat. Commun.* **14**, 1–19 (2023).

39. M. Mina, A. Iyer, G. Ciriello, Epistasis and evolutionary dependencies in human cancers. *Curr. Opin. Genet. Dev.* **77**, 101989 (2022).

40. Z. Zhang, Y. Yang, Y. Zhou, H. Fang, M. Yuan, K. Sasser, H. Hamadeh, X. S. Xu, A forward selection algorithm to identify mutually exclusive alterations in cancer studies. *J. Hum. Genet.* **66**, 509–518 (2021).

41. T. Fedrizzi, Y. Ciani, F. Lorenzin, T. Cantore, P. Gasperini, F. Demichelis, Fast mutual exclusivity algorithm nominates potential synthetic lethal gene pairs through brute force matrix product computations. *Comput. Struct. Biotechnol. J.* **19**, 4394–4403 (2021).

42. S. Liu, J. Liu, Y. Xie, T. Zhai, E. W. Hinderer, A. J. Stromberg, N. L. Vanderford, J. M. Kolesar, H. N. B. Moseley, L. Chen, C. Liu, C. Wang, MEScan: a powerful statistical framework for genome-scale mutual exclusivity analysis of cancer mutations. *Bioinformatics* **37**, 1189–1197 (2021).

43. M. I. Klein, V. L. Cannataro, J. P. Townsend, S. Newman, D. F. Stern, H. Zhao, Identifying modules of cooperating cancer drivers. *Mol. Syst. Biol.*, doi: 10.15252/msb.20209810 (2021).

44. M. Mina, F. Raynaud, D. Tavernari, E. Battistello, S. Sungalee, S. Saghafinia, T. Laessle, F. Sanchez-Vega, N. Schultz, E. Oricchio, G. Ciriello, Conditional Selection of Genomic Alterations Dictates Cancer Evolution and Oncogenic Dependencies. *Cancer Cell* **32**, 155–168.e6 (2017).

45. S. Kryazhimskiy, D. P. Rice, E. R. Jerison, M. M. Desai, Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science*, doi: 10.1126/science.1250939 (2014).

46. D. M. Weinreich, Y. Lan, C. S. Wylie, R. B. Heckendorn, Should evolutionary geneticists worry about higher-order epistasis? *Curr. Opin. Genet. Dev.* **23**, 700–707 (2013).

47. Z. R. Sailer, M. J. Harms, High-order epistasis shapes evolutionary trajectories. *PLoS Comput. Biol.* **13**, e1005541 (2017).

48. D. M. Weinreich, Y. Lan, J. Jaffe, R. B. Heckendorn, The influence of higher-order epistasis on biological fitness landscape topography. *J. Stat. Phys.* **172**, 208–225 (2018).

49. J. Domingo, G. Diss, B. Lehner, Pairwise and higher-order genetic interactions during the evolution of a tRNA. *Nature* **558**, 117–121 (2018).

50. A. Papkou, L. Garcia-Pastor, J. A. Escudero, A. Wagner, A rugged yet easily navigable fitness landscape. *Science* **382**, eadh3860 (2023).

51. X. Wei, J. Zhang, Patterns and Mechanisms of Diminishing Returns from Beneficial Mutations. *Mol. Biol. Evol.* **36**, 1008–1021 (2019).

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

53. H.-H. Chou, H.-C. Chiu, N. F. Delaney, D. Segrè, C. J. Marx, Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **332**, 1190–1192 (2011).

54. A. Wünsche, D. M. Dinh, R. S. Satterwhite, C. D. Arenas, D. M. Stoebel, T. F. Cooper, Diminishing-returns epistasis decreases adaptability along an evolutionary trajectory. *Nature Ecology & Evolution* **1**, 1–6 (2017).

55. S. Devarakonda, Y. Li, F. M. Rodrigues, S. Sankararaman, H. Kadara, C. Goparaju, I. Lanc, K. Pepin, S. N. Waqar, D. Morgensztern, J. Ward, A. Masood, R. Fulton, L. Fulton, M. A. Gillette, S.

Satpathy, S. A. Carr, I. Wistuba, H. Pass, R. K. Wilson, L. Ding, R. Govindan, Genomic Profiling of Lung Adenocarcinoma in Never-Smokers. *J. Clin. Oncol.*, doi: 10.1200/JCO.21.01691 (2021).

56. S. Moorthi, A. Paguirigan, P. Itagi, M. Ko, M. Pettinger, A. C. Hoge, A. Nag, N. A. Patel, F. Wu, C. Sather, K. M. Levine, M. P. Fitzgibbon, A. R. Thorner, G. L. Anderson, G. Ha, A. H. Berger, The genomic landscape of lung cancer in never-smokers from the Women's Health Initiative. *JCI Insight* **9** (2024).

57. R. Frank, M. Scheffler, S. Merkelbach-Bruse, M. A. Ihle, A. Kron, M. Rauer, F. Ueckeroth, K. König, S. Michels, R. Fischer, A. Eisert, J. Fassunke, C. Heydt, M. Serke, Y. D. Ko, U. Gerigk, T. Geist, B. Kaminsky, L. C. Heukamp, M. Clement-Ziza, R. Büttner, J. Wolf, Clinical and Pathological Characteristics of KEAP1- and NFE2L2-Mutated Non-Small Cell Lung Carcinoma (NSCLC). *Clin. Cancer Res.* **24** (2018).

58. E. Nadal, R. Palmero, C. Muñoz-Pinedo, Mutations in the Antioxidant KEAP1/NRF2 Pathway Define an Aggressive Subset of NSCLC Resistant to Conventional Treatments. *J. Thorac. Oncol.* **14** (2019).

59. J. Subramanian, R. Govindan, Lung Cancer in Never Smokers: A Review. *J. Clin. Oncol.*, doi: 10.1200/JCO.2006.06.8015 (2016).

60. J. Scott, A. Marusyk, Somatic clonal evolution: A selection-centric perspective. *Biochim. Biophys. Acta* **1867**, 139–150 (2017).

61. C. J. Fleenor, A. Marusyk, J. DeGregori, Ionizing radiation and hematopoietic malignancies: altering the adaptive landscape. *Cell Cycle* **9**, 3005–3011 (2010).

62. W. Hill, E. L. Lim, C. E. Weeden, C. Lee, M. Augustine, K. Chen, F.-C. Kuan, F. Marongiu, E. J. Evans Jr, D. A. Moore, F. S. Rodrigues, O. Pich, B. Bakker, H. Cha, R. Myers, F. van Maldegem, J. Boumelha, S. Veeriah, A. Rowan, C. Naceur-Lombardelli, T. Karasaki, M. Sivakumar, S. De, D. R. Caswell, A. Nagano, J. R. M. Black, C. Martínez-Ruiz, M. H. Ryu, R. D. Huff, S. Li, M.-J. Favé, A. Magness, A. Suárez-Bonnet, S. L. Priestnall, M. Lüchtenborg, K. Lavelle, J. Pethick, S. Hardy, F. E. McRonald, M.-H. Lin, C. I. Troccoli, M. Ghosh, Y. E. Miller, D. T. Merrick, R. L. Keith, M. Al Bakir, C. Bailey, M. S. Hill, L. H. Saal, Y. Chen, A. M. George, C. Abbosh, N. Kanu, S.-H. Lee, N. McGranahan, C. D. Berg, P. Sasieni, R. Houlston, C. Turnbull, S. Lam, P. Awadalla, E. Grönroos, J. Downward, T. Jacks, C. Carlsten, I. Malanchi, A. Hackshaw, K. Litchfield, TRACERx Consortium, J. DeGregori, M. Jamal-Hanjani, C. Swanton, Lung adenocarcinoma promotion by air pollutants. *Nature* **616**, 159–167 (2023).

63. I. A. Kim, J. S. Lee, H. J. Kim, W. S. Kim, K. Y. Lee, Cumulative smoking dose affects the clinical outcomes of EGFR-mutated lung adenocarcinoma patients treated with EGFR-TKIs: a retrospective study. *BMC Cancer* **18**, 768 (2018).

64. M. H. Kim, H. R. Kim, B. C. Cho, M. K. Bae, E. Y. Kim, C. Y. Lee, J. S. Lee, D. R. Kang, J. H. Kim, Impact of cigarette smoking on response to epidermal growth factor receptor (EGFR)-tyrosine kinase inhibitors in lung adenocarcinoma with activating EGFR mutations. *Lung Cancer* **84**, 196–202 (2014).

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

66. S. Igawa, J. Sasaki, S. Otani, M. Ishihara, A. Takakura, M. Katagiri, N. Masuda, 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* **89**, 275–280 (2015).

67. R. Romero, V. I. Sayin, S. M. Davidson, M. R. Bauer, S. X. Singh, S. E. LeBoeuf, T. R. Karakousi, D. C. Ellis, A. Bhutkar, F. J. Sánchez-Rivera, L. Subbaraj, B. Martinez, R. T. Bronson, J. R. Prigge, E. E. Schmidt, C. J. Thomas, C. Goparaju, A. Davies, I. Dolgalev, A. Heguy, V. Allaj, J. T. Poirier, A. L. Moreira, C. M. Rudin, H. I. Pass, M. G. Vander Heiden, T. Jacks, T. Papagiannakopoulos, Keap1 loss promotes Kras-driven lung cancer and results in dependence on glutaminolysis. *Nat. Med.* **23**, 1362–1368 (2017).

68. S. A. Best, S. Ding, A. Kersbergen, X. Dong, J.-Y. Song, Y. Xie, B. Reljic, K. Li, J. E. Vince, V. Rathi, G. M. Wright, M. E. Ritchie, K. D. Sutherland, Distinct initiating events underpin the immune and metabolic heterogeneity of KRAS-mutant lung adenocarcinoma. *Nat. Commun.* **10**, 4190 (2019).

69. A. Galan-Cobo, P. Sitthideatphaiboon, X. Qu, A. Poteete, M. A. Pisegna, P. Tong, P.-H. Chen, L. K. Boroughs, M. L. M. Rodriguez, W. Zhang, F. Parlati, J. Wang, V. Gandhi, F. Skoulidis, R. J. DeBerardinis, J. D. Minna, J. V. Heymach, LKB1 and KEAP1/NRF2 pathways cooperatively promote metabolic reprogramming with enhanced glutamine dependence in KRAS-mutant lung adenocarcinoma. *Cancer Res.* **79**, 3251–3267 (2019).

70. B. Ricciuti, K. C. Arbour, J. J. Lin, A. Vajdi, N. Vokes, L. Hong, J. Zhang, M. Y. Tolstorukov, Y. Y. Li, L. F. Spurr, A. D. Cherniack, G. Recondo, G. Lamberti, X. Wang, D. Venkatraman, J. V. Alessi, V. R. Vaz, H. Rizvi, J. Egger, A. J. Plodkowski, S. Khosrowjerdi, S. Digumarthy, H. Park, N. Vaz, M. Nishino, L. M. Sholl, D. Barbie, M. Altan, J. V. Heymach, F. Skoulidis, J. F. Gainor, M. D. Hellmann, M. M. Awad, Diminished Efficacy of Programmed Death-(Ligand)1 Inhibition in STK11- and KEAP1-Mutant Lung Adenocarcinoma Is Affected by KRAS Mutation Status. *J. Thorac. Oncol.* **17**, 399–410 (2022).

71. W. G. Kwack, S. Y. Shin, S. H. Lee, Primary resistance to immune checkpoint blockade in an STK11/TP53/KRAS-mutant lung adenocarcinoma with high PD-L1 expression. *Onco. Targets. Ther.* **13**, 8901–8905 (2020).

72. K. C. Arbour, E. Jordan, H. R. Kim, J. Dienstag, H. A. Yu, F. Sanchez-Vega, P. Lito, M. Berger, D. B. Solit, M. Hellmann, M. G. Kris, C. M. Rudin, A. Ni, M. Arcila, M. Ladanyi, G. J. Riely, Effects of Co-occurring Genomic Alterations on Outcomes in Patients with KRAS-Mutant Non–Small Cell Lung Cancer. *Clin. Cancer Res.* **24**, 334–340 (2018).

73. F. Proulx-Rocray, B. Routy, R. Nassabein, W. Belkaid, D. Tran-Thanh, J. Malo, M. Tonneau, O. E. Ouarzadi, M. Florescu, M. Tehfe, N. Blais, 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.* **37**, 100767 (2023).

74. J. Baptiste Oudart, S. Garinet, C. Leger, F. Barlesi, J. Mazières, G. Jeannin, C. Audigier-Valette, D. Morot-Sibilot, A. Langlais, E. Amour, N. Mathiot, G. Birsen, H. Blons, M. Wislez, 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* **190**, 107508 (2024).

75. E. Bange, M. E. Marmarelis, W.-T. Hwang, Y.-X. Yang, J. C. Thompson, J. Rosenbaum, J. M. Bauml, C. Ciunci, E. W. Alley, R. B. Cohen, C. J. Langer, E. Carpenter, C. Aggarwal, 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.* **3**, 1–11 (2019).

76. H. J. West, M. McCleland, F. Cappuzzo, M. Reck, T. S. Mok, R. M. Jotte, M. Nishio, E. Kim, S. Morris, W. Zou, D. Shames, M. Das Thakur, G. Shankar, M. A. Socinski, 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* **10**, e003027 (2022).

77. J. V. Aredo, S. K. Padda, C. A. Kunder, S. S. Han, J. W. Neal, J. B. Shrager, H. A. Wakelee, Impact of KRAS mutation subtype and concurrent pathogenic mutations on non-small cell lung cancer outcomes. *Lung Cancer* **133**, 144–150 (2019).

78. C. A. Wohlhieter, A. L. Richards, F. Uddin, C. H. Hulton, À. Quintanal-Villalonga, A. Martin, E. de Stanchina, U. Bhanot, M. Asher, N. S. Shah, O. Hayatt, D. J. Buonocore, N. Rekhtman, R. Shen, K. C. Arbour, M. Donoghue, J. T. Poirier, T. Sen, C. M. Rudin, Concurrent mutations in STK11 and KEAP1 promote ferroptosis protection and SCD1 dependence in lung cancer. *Cell Rep.* **33**, 108444 (2020).

79. C. Julian, N. Pal, A. Gershon, M. Evangelista, H. Purkey, P. Lambert, Z. Shi, Q. Zhang, 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* **23** (2023).

80. M. Boeschen, C. K. Kuhn, H. Wirtz, H.-J. Seyfarth, A. Frille, F. Lordick, U. T. Hacker, U. Obeck, M. Stiller, H. Bläker, M. von Laffert, 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* **184**, 107361 (2023).

81. D. Schaufler, D. F. Ast, H. L. Tumbrink, N. Abedpour, L. Maas, A. E. Schwäbe, I. Spille, S. Lennartz, J. Fassunke, M. Aldea, B. Besse, D. Planchard, L. Nogova, S. Michels, C. Kobe, T. Persigehl, T. Westphal, S. Koleczko, R. Fischer, J.-P. Weber, J. Altmüller, R. K. Thomas, S. Merkelbach-Bruse, O. Gautschi, L. Mezquita, R. Büttner, J. Wolf, M. Peifer, J. Brägelmann, M. Scheffler, M. L. Sos, Clonal dynamics of BRAF-driven drug resistance in EGFR-mutant lung cancer. *npj Precision Oncology* **5**, 1–12 (2021).

82. B. Gini, N. Thomas, C. M. Blakely, Impact of concurrent genomic alterations in epidermal growth factor receptor (EGFR)-mutated lung cancer. *J. Thorac. Dis.* **12**, 2883–2895 (2020).

83. M. Offin, J. M. Chan, M. Tenet, H. A. Rizvi, R. Shen, G. J. Riely, N. Rekhtman, Y. Daneshbod, A. Quintanal-Villalonga, A. Penson, M. D. Hellmann, M. E. Arcila, M. Ladanyi, D. Pe'er, M. G. Kris, C. M. Rudin, H. A. Yu, 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.* **14**, 1784–1793 (2019).

84. E. Pros, M. Saigi, D. Alameda, G. Gomez-Mariano, B. Martinez-Delgado, J. J. Alburquerque-Bejar, J. Carretero, R. Tonda, A. Esteve-Codina, I. Catala, R. Palmero, M. Jove, C. Lazaro, A. Patiño-Garcia, I. Gil-Bazo, S. Verdura, A. Teulé, J. Torres-Lanzas, D. Sidransky, N. Reguart, R. Pio, O. Juan-Vidal, E. Nadal, E. Felip, L. M. Montuenga, M. Sanchez-Cespedes, Genome-wide profiling of non-smoking-related lung cancer cells reveals common RB1 rearrangements associated with histopathologic transformation in EGFR-mutant tumors. *Ann. Oncol.* **31**, 274–282 (2020).

85. A. Marchetti, M. Milella, L. Felicioni, F. Cappuzzo, L. Irtelli, M. Del Grammastro, M. Sciarrotta, S. Malatesta, C. Nuzzo, G. Finocchiaro, B. Perrucci, D. Carlone, A. J. Gelibter, A. Ceribelli, A. Mezzetti, S. Iacobelli, F. Cognetti, F. Buttitta, Clinical implications of KRAS mutations in lung cancer patients treated with tyrosine kinase inhibitors: an important role for mutations in minor

clones. *Neoplasia* **11**, 1084–1092 (2009).

86. J. A. Hellyer, H. Stehr, M. Das, S. K. Padda, K. Ramchandran, J. W. Neal, M. Diehn, H. A. Wakelee, 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* **134**, 42–45 (2019).

87. E. B. Krall, B. Wang, D. M. Munoz, N. Ilic, S. Raghavan, M. J. Niederst, K. Yu, D. A. Ruddy, A. J. Aguirre, J. W. Kim, A. J. Redig, J. F. Gainor, J. A. Williams, J. M. Asara, J. G. Doench, P. A. Janne, A. T. Shaw, R. E. McDonald Iii, J. A. Engelman, F. Stegmeier, M. R. Schlabach, W. C. Hahn, KEAP1 loss modulates sensitivity to kinase targeted therapy in lung cancer. *Elife* **6** (2017).

88. S. Hobor, M. Al Bakir, C. T. Hiley, M. Skrzypski, A. M. Frankell, B. Bakker, T. B. K. Watkins, A. Markovets, J. R. Dry, A. P. Brown, J. van der Aart, H. van den Bos, D. Spierings, D. Oukrif, M. Novelli, T. Chakrabarti, A. H. Rabinowitz, L. Ait Hassou, S. Litière, D. L. Kerr, L. Tan, G. Kelly, D. A. Moore, M. J. Renshaw, S. Venkatesan, W. Hill, A. Huebner, C. Martínez-Ruiz, J. R. M. Black, W. Wu, M. Angelova, N. McGranahan, J. Downward, J. Chmielecki, C. Barrett, K. Litchfield, S. K. Chew, C. M. Blakely, E. C. de Bruin, F. Foijer, K. H. Vousden, T. G. Bivona, TRACERx consortium, R. E. Hynds, N. Kanu, S. Zaccaria, E. Grönroos, C. Swanton, Mixed responses to targeted therapy driven by chromosomal instability through p53 dysfunction and genome doubling. *Nat. Commun.* **15**, 4871 (2024).

89. E. Bria, S. Pilotto, E. Amato, M. Fassan, S. Novello, U. Peretti, T. Vavalà, S. Kinspergher, L. Righi, A. Santo, M. Brunelli, V. Corbo, E. Giglioli, I. Sperduti, M. Milella, M. Chilosi, A. Scarpa, G. Tortora, Molecular heterogeneity assessment by next-generation sequencing and response to gefitinib of EGFR mutant advanced lung adenocarcinoma. *Oncotarget* **6**, 12783–12795 (2015).

90. N. I. Vokes, E. Chambers, T. Nguyen, A. Coolidge, C. A. Lydon, X. Le, L. Sholl, J. V. Heymach, M. Nishino, E. M. Van Allen, P. A. Jänne, Concurrent TP53 Mutations Facilitate Resistance Evolution in EGFR-Mutant Lung Adenocarcinoma. *J. Thorac. Oncol.* **17**, 779–792 (2022).

91. M. Canale, E. Petracci, A. Delmonte, E. Chiadini, C. Dazzi, M. Papi, L. Capelli, C. Casanova, N. De Luigi, M. Mariotti, A. Gamboni, R. Chiari, C. Bennati, D. Calistri, V. Ludovini, L. Crinò, D. Amadori, P. Ulivi, Impact of TP53 Mutations on Outcome in EGFR-Mutated Patients Treated with First-Line Tyrosine Kinase Inhibitors. *Clin. Cancer Res.* **23**, 2195–2202 (2017).

92. J. K. Rho, Y. J. Choi, B.-Y. Ryoo, I. I. I. Na, S. H. Yang, C. H. Kim, J. C. Lee, p53 enhances gefitinib-induced growth inhibition and apoptosis by regulation of Fas in non-small cell lung cancer. *Cancer Res.* **67**, 1163–1169 (2007).

93. R. Ibusuki, E. Iwama, A. Shimauchi, H. Tsutsumi, Y. Yoneshima, K. Tanaka, I. Okamoto, TP53 gain-of-function mutations promote osimertinib resistance via TNF-α-NF-κB signaling in EGFR-mutated lung cancer. *NPJ Precis Oncol* **8**, 60 (2024).

94. M. Canale, E. Petracci, A. Delmonte, G. Bronte, E. Chiadini, V. Ludovini, A. Dubini, M. Papi, S. Baglivo, N. De Luigi, A. Verlicchi, R. Chiari, L. Landi, G. Metro, M. A. Burgio, L. Crinò, P. Ulivi, Concomitant TP53 Mutation Confers Worse Prognosis in EGFR-Mutated Non-Small Cell Lung Cancer Patients Treated with TKIs. *J. Clin. Med. Res.* **9** (2020).

95. N. J. Choudhury, J. A. Lavery, S. Brown, I. de Bruijn, J. Jee, T. N. Tran, H. Rizvi, K. C. Arbour, K. Whiting, R. Shen, M. Hellmann, P. L. Bedard, C. Yu, N. Leighl, M. LeNoue-Newton, C. Micheel, J. L. Warner, M. S. Ginsberg, A. Plodkowski, J. Girshman, P. Sawan, S. Pillai, S. M. Sweeney, K. L. Kehl, K. S. Panageas, N. Schultz, D. Schrag, G. J. Riely, AACR GENIE BPC Core Team, 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.* **29**, 3418–3428 (2023).

96. L. A. Donehower, T. Soussi, A. Korkut, Y. Liu, A. Schultz, M. Cardenas, X. Li, O. Babur, T.-K. Hsu, O. Lichtarge, J. N. Weinstein, R. Akbani, D. A. Wheeler, Integrated analysis of TP53 gene and pathway alterations in The Cancer Genome Atlas. *Cell Rep.* **28**, 1370–1384.e5 (2019).

97. G. Iyer, A. J. Hanrahan, M. I. Milowsky, H. Al-Ahmadie, S. N. Scott, M. Janakiraman, M. Pirun, C. Sander, N. D. Socci, I. Ostrovnaya, A. Viale, A. Heguy, L. Peng, T. A. Chan, B. Bochner, D. F. Bajorin, M. F. Berger, B. S. Taylor, D. B. Solit, Genome sequencing identifies a basis for everolimus sensitivity. *Science* **338**, 221 (2012).

98. B. Weigelt, J. S. Reis-Filho, Epistatic interactions and drug response. *J. Pathol.* **232**, 255–263 (2014).

99. J. F. Wilkins, V. L. Cannataro, B. Shuch, J. P. Townsend, Analysis of mutation, selection, and epistasis: an informed approach to cancer clinical trials. *Oncotarget* **9**, 22243–22253 (2018).

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

101. A. Guénolé, R. Srivas, K. Vreeken, Z. Z. Wang, S. Wang, N. J. Krogan, T. Ideker, H. van Attikum, Dissection of DNA damage responses using multiconditional genetic interaction maps. *Mol. Cell* **49**, 346–358 (2013).

102. R. Harrison, B. Papp, C. Pál, S. G. Oliver, D. Delneri, Plasticity of genetic interactions in metabolic networks of yeast. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 2307–2312 (2007).

103. C.-T. Zhu, P. Ingelmo, D. M. Rand, G×G×E for lifespan in Drosophila: mitochondrial, nuclear, and dietary interactions that modify longevity. *PLoS Genet.* **10**, e1004354 (2014).

104. X. Shi, B. Reinstadler, H. Shah, T.-L. To, K. Byrne, L. Summer, S. E. Calvo, O. Goldberger, J. G. Doench, V. K. Mootha, H. Shen, Combinatorial GxGxE CRISPR screen identifies SLC25A39 in mitochondrial glutathione transport linking iron homeostasis to OXPHOS. *Nat. Commun.* **13**, 2483 (2022).

105. K. M. Flynn, T. F. Cooper, F. B.-G. Moore, V. S. Cooper, The environment affects epistatic interactions to alter the topology of an empirical fitness landscape. *PLoS Genet.* **9**, e1003426 (2013).

106. M. Costanzo, J. Hou, V. Messier, J. Nelson, M. Rahman, B. VanderSluis, W. Wang, C. Pons, C. Ross, M. Ušaj, B.-J. San Luis, E. Shuteriqi, E. N. Koch, P. Aloy, C. L. Myers, C. Boone, B. Andrews, Environmental robustness of the global yeast genetic interaction network. *Science* **372** (2021).

107. Y. Sun, Q. Yang, J. Shen, T. Wei, W. Shen, N. Zhang, P. Luo, J. Zhang, 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* **9**, 745859 (2021).

108. C. A. Ortmann, D. G. Kent, J. Nangalia, Y. Silber, D. C. Wedge, J. Grinfeld, E. J. Baxter, C. E. Massie, E. Papaemmanuil, S. Menon, A. L. Godfrey, D. Dimitropoulou, P. Guglielmelli, B.

Bellosillo, C. Besses, K. Döhner, C. N. Harrison, G. S. Vassiliou, A. Vannucchi, P. J. Campbell, A. R. Green, Effect of mutation order on myeloproliferative neoplasms. *N. Engl. J. Med.* **372**, 601–612 (2015).

109. M. Boström, E. Larsson, Somatic mutation distribution across tumour cohorts provides a signal for positive selection in cancer. *Nat. Commun.* **13**, 7023 (2022).

110. K. L. Bolton, R. N. Ptashkin, T. Gao, L. Braunstein, S. M. Devlin, D. Kelly, M. Patel, A. Berthon, A. Syed, M. Yabe, C. C. Coombs, N. M. Caltabellotta, M. Walsh, K. Offit, Z. Stadler, D. Mandelker, J. Schulman, A. Patel, J. Philip, E. Bernard, G. Gundem, J. E. A. Ossa, M. Levine, J. S. M. Martinez, N. Farnoud, D. Glodzik, S. Li, M. E. Robson, C. Lee, P. D. P. Pharoah, K. H. Stopsack, B. Spitzer, S. Mantha, J. Fagin, L. Boucai, C. J. Gibson, B. L. Ebert, A. L. Young, T. Druley, K. Takahashi, N. Gillis, M. Ball, E. Padron, D. M. Hyman, J. Baselga, L. Norton, S. Gardos, V. M. Klimek, H. Scher, D. Bajorin, E. Paraiso, R. Benayed, M. E. Arcila, M. Ladanyi, D. B. Solit, M. F. Berger, M. Tallman, M. Garcia-Closas, N. Chatterjee, L. A. Diaz Jr, R. L. Levine, L. M. Morton, A. Zehir, E. Papaemmanuil, Cancer therapy shapes the fitness landscape of clonal hematopoiesis. *Nat. Genet.* **52**, 1219–1226 (2020).

111. S. Salehi, F. Kabeer, N. Ceglia, M. Andronescu, M. J. Williams, K. R. Campbell, T. Masud, B. Wang, J. Biele, J. Brimhall, D. Gee, H. Lee, J. Ting, A. W. Zhang, H. Tran, C. O'Flanagan, F. Dorri, N. Rusk, T. R. de Algara, S. R. Lee, B. Y. C. Cheng, P. Eirew, T. Kono, J. Pham, D. Grewal, D. Lai, R. Moore, A. J. Mungall, M. A. Marra, IMAXT Consortium, A. McPherson, A. Bouchard-Côté, S. Aparicio, S. P. Shah, Clonal fitness inferred from time-series modelling of single-cell cancer genomes. *Nature* **595**, 585–590 (2021).

112. L. A. Liggett, J. DeGregori, Changing mutational and adaptive landscapes and the genesis of cancer. *Biochim. Biophys. Acta Rev. Cancer* **1867**, 84–94 (2017).

113. T.-K. Hsu, J. Asmussen, A. Koire, B.-K. Choi, M. A. Gadhikar, E. Huh, C.-H. Lin, D. M. Konecki, Y. W. Kim, C. R. Pickering, M. Kimmel, L. A. Donehower, M. J. Frederick, J. N. Myers, P. Katsonis, O. Lichtarge, A general calculus of fitness landscapes finds genes under selection in cancers. *Genome Res.* **32**, 916–929 (2022).

114. N. Misra, E. Szczurek, M. Vingron, Inferring the paths of somatic evolution in cancer. *Bioinformatics* **30**, 2456–2463 (2014).

115. S. M. Ernst, J. M. Mankor, J. van Riet, J. H. von der Thüsen, H. J. Dubbink, A. Jgjv, A. J. de Langen, E. F. Smit, A. C. Dingemans, K. Monkhorst, Tobacco Smoking-Related Mutational Signatures in Classifying Smoking-Associated and Nonsmoking-Associated NSCLC. *J. Thorac. Oncol.* **18** (2023).

116. S. D. Hobbs, A. B. Wilmink, D. J. Adam, A. W. Bradbury, Assessment of smoking status in patients with peripheral arterial disease. *J. Vasc. Surg.* **41** (2005).

117. J. Hald, J. Overgaard, C. Grau, 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.* **42** (2003).

118. P. Brennan, P. A. Buffler, P. Reynolds, A. H. Wu, H. E. Wichmann, A. Agudo, G. Pershagen, K. H. Jöckel, S. Benhamou, R. S. Greenberg, F. Merletti, C. Winck, E. T. Fontham, M. Kreuzer, S. C. Darby, F. Forastiere, L. Simonato, P. Boffetta, Secondhand smoke exposure in adulthood and risk of lung cancer among never smokers: a pooled analysis of two large studies. *International journal of*

cancer **109** (2004).

119. S. M. Morris, A role for p53 in the frequency and mechanism of mutation. *Mutat. Res.* **511** (2002).

120. J. Zámborszky, B. Szikriszt, J. Z. Gervai, O. Pipek, Á. Póti, M. Krzystanek, D. Ribli, J. M. Szalai-Gindl, I. Csabai, Z. Szallasi, C. Swanton, A. L. Richardson, D. Szüts, Loss of BRCA1 or BRCA2 markedly increases the rate of base substitution mutagenesis and has distinct effects on genomic deletions. *Oncogene* **36**, 746–755 (2016).

121. R. Beroukhim, C. H. Mermel, D. Porter, G. Wei, S. Raychaudhuri, J. Donovan, J. Barretina, J. S. Boehm, J. Dobson, M. Urashima, K. T. Mc Henry, R. M. Pinchback, A. H. Ligon, Y.-J. Cho, L. Haery, H. Greulich, M. Reich, W. Winckler, M. S. Lawrence, B. A. Weir, K. E. Tanaka, D. Y. Chiang, A. J. Bass, A. Loo, C. Hoffman, J. Prensner, T. Liefeld, Q. Gao, D. Yecies, S. Signoretti, E. Maher, F. J. Kaye, H. Sasaki, J. E. Tepper, J. A. Fletcher, J. Tabernero, J. Baselga, M.-S. Tsao, F. Demichelis, M. A. Rubin, P. A. Janne, M. J. Daly, C. Nucera, R. L. Levine, B. L. Ebert, S. Gabriel, A. K. Rustgi, C. R. Antonescu, M. Ladanyi, A. Letai, L. A. Garraway, M. Loda, D. G. Beer, L. D. True, A. Okamoto, S. L. Pomeroy, S. Singer, T. R. Golub, E. S. Lander, G. Getz, W. R. Sellers, M. Meyerson, The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).

122. E. Besedina, F. Supek, Copy number losses of oncogenes and gains of tumor suppressor genes generate common driver mutations. *Nat. Commun.* **15**, 6139 (2024).

123. I. Tomanek, C. C. Guet, Adaptation dynamics between copy-number and point mutations. *Elife* **11** (2022).

124. G. Ciriello, E. Cerami, C. Sander, N. Schultz, Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res.* **22**, 398–406 (2012).

125. J. Chen, H. Yang, A. S. M. Teo, L. B. Amer, F. G. Sherbaf, C. Q. Tan, J. J. S. Alvarez, B. Lu, J. Q. Lim, A. Takano, R. Nahar, Y. Y. Lee, C. Z. J. Phua, K. P. Chua, L. Suteja, P. J. Chen, M. M. Chang, T. P. T. Koh, B.-H. Ong, D. Anantham, A. A. L. Hsu, A. Gogna, C. W. Too, Z. W. Aung, Y. F. Lee, L. Wang, T. K. H. Lim, A. Wilm, P. S. Choi, P. Y. Ng, C. K. Toh, W.-T. Lim, S. Ma, B. Lim, J. Liu, W. L. Tam, A. J. Skanderup, J. P. S. Yeong, E.-H. Tan, C. L. Creasy, D. S. W. Tan, A. M. Hillmer, W. Zhai, Genomic landscape of lung adenocarcinoma in East Asians. *Nat. Genet.* **52**, 177–186 (2020).

126. M. Imielinski, A. H. Berger, P. S. Hammerman, B. Hernandez, T. J. Pugh, E. Hodis, J. Cho, J. Suh, M. Capelletti, A. Sivachenko, C. Sougnez, D. Auclair, M. S. Lawrence, P. Stojanov, K. Cibulskis, K. Choi, L. de Waal, T. Sharifnia, A. Brooks, H. Greulich, S. Banerji, T. Zander, D. Seidel, F. Leenders, S. Ansén, C. Ludwig, W. Engel-Riedel, E. Stoelben, J. Wolf, C. Goparju, K. Thompson, W. Winckler, D. Kwiatkowski, B. E. Johnson, P. A. Jänne, V. A. Miller, W. Pao, W. D. Travis, H. I. Pass, S. B. Gabriel, E. S. Lander, R. K. Thomas, L. A. Garraway, G. Getz, M. Meyerson, Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150**, 1107–1120 (2012).

127. M. Jamal-Hanjani, G. A. Wilson, N. McGranahan, N. J. Birkbak, T. B. K. Watkins, S. Veeriah, S. Shafi, D. H. Johnson, R. Mitter, R. Rosenthal, M. Salm, S. Horswell, M. Escudero, N. Matthews, A. Rowan, T. Chambers, D. A. Moore, S. Turajlic, H. Xu, S.-M. Lee, M. D. Forster, T. Ahmad, C. T. Hiley, C. Abbosh, M. Falzon, E. Borg, T. Marafioti, D. Lawrence, M. Hayward, S. Kolvekar, N. Panagiotopoulos, S. M. Janes, R. Thakrar, A. Ahmed, F. Blackhall, Y. Summers, R. Shah, L. Joseph, A. M. Quinn, P. A. Crosbie, B. Naidu, G. Middleton, G. Langman, S. Trotter, M. Nicolson,

H. Remmen, K. Kerr, M. Chetty, L. Gomersall, D. A. Fennell, A. Nakas, S. Rathinam, G. Anand, S. Khan, P. Russell, V. Ezhil, B. Ismail, M. Irvin-Sellers, V. Prakash, J. F. Lester, M. Kornaszewska, R. Attanoos, H. Adams, H. Davies, S. Dentro, P. Taniere, B. O'Sullivan, H. L. Lowe, J. A. Hartley, N. Iles, H. Bell, Y. Ngai, J. A. Shaw, J. Herrero, Z. Szallasi, R. F. Schwarz, A. Stewart, S. A. Quezada, J. Le Quesne, P. Van Loo, C. Dive, A. Hackshaw, C. Swanton, TRACERx Consortium, Tracking the Evolution of Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **376**, 2109–2121 (2017).

128. E. J. Jordan, H. R. Kim, M. E. Arcila, D. Barron, D. Chakravarty, J. Gao, M. T. Chang, A. Ni, R. Kundra, P. Jonsson, G. Jayakumaran, S. P. Gao, H. C. Johnsen, A. J. Hanrahan, A. Zehir, N. Rekhtman, M. S. Ginsberg, B. T. Li, H. A. Yu, P. K. Paik, A. Drilon, M. D. Hellmann, D. N. Reales, R. Benayed, V. W. Rusch, M. G. Kris, J. E. Chaft, J. Baselga, B. S. Taylor, N. Schultz, C. M. Rudin, D. M. Hyman, M. F. Berger, D. B. Solit, M. Ladanyi, G. J. Riely, Prospective Comprehensive Molecular Characterization of Lung Adenocarcinomas for Efficient Patient Matching to Approved and Emerging Therapies. *Cancer Discov.* **7**, 596–609 (2017).

129. N. A. Rizvi, M. D. Hellmann, A. Snyder, P. Kvistborg, V. Makarov, J. J. Havel, W. Lee, J. Yuan, P. Wong, T. S. Ho, M. L. Miller, N. Rekhtman, A. L. Moreira, F. Ibrahim, C. Bruggeman, B. Gasmi, R. Zappasodi, Y. Maeda, C. Sander, E. B. Garon, T. Merghoub, J. D. Wolchok, T. N. Schumacher, T. A. Chan, Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124–128 (2015).

130. H. Rizvi, F. Sanchez-Vega, K. La, W. Chatila, P. Jonsson, D. Halpenny, A. Plodkowski, N. Long, J. L. Sauter, N. Rekhtman, T. Hollmann, K. A. Schalper, J. F. Gainor, R. Shen, A. Ni, K. C. Arbour, T. Merghoub, J. Wolchok, A. Snyder, J. E. Chaft, M. G. Kris, C. M. Rudin, N. D. Socci, M. F. Berger, B. S. Taylor, A. Zehir, D. B. Solit, M. E. Arcila, M. Ladanyi, G. J. Riely, N. Schultz, M. D. Hellmann, 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.* **36**, 633–641 (2018).

131. T. Zhang, P. Joubert, N. Ansari-Pour, W. Zhao, P. H. Hoang, R. Lokanga, A. L. Moye, J. Rosenbaum, A. Gonzalez-Perez, F. Martínez-Jiménez, A. Castro, L. A. Muscarella, P. Hofman, D. Consonni, A. C. Pesatori, M. Kebede, M. Li, B. E. Gould Rothberg, I. Peneva, M. B. Schabath, M. L. Poeta, M. Costantini, D. Hirsch, K. Heselmeyer-Haddad, A. Hutchinson, M. Olanich, S. M. Lawrence, P. Lenz, M. Duggan, P. M. S. Bhawsar, J. Sang, J. Kim, L. Mendoza, N. Saini, L. J. Klimczak, S. M. Ashiqul Islam, B. Otlu, A. Khandekar, N. Cole, D. R. Stewart, J. Choi, K. Brown, N. E. Caporaso, S. H. Wilson, Y. Pommier, Q. Lan, N. Rothman, J. S. Almeida, H. Carter, T. Ried, C. F. Kim, N. Lopez-Bigas, M. Garcia-Closas, J. Shi, Y. Bossé, B. Zhu, D. A. Gordenin, L. B. Alexandrov, S. J. Chanock, D. C. Wedge, M. T. Landi, Genomic and evolutionary classification of lung cancer in never smokers. *Nat. Genet.* **53**, 1348 (2021).

132. M. A. Gillette, S. Satpathy, S. Cao, S. M. Dhanasekaran, S. V. Vasaikar, K. Krug, F. Petralia, Y. Li, W.-W. Liang, B. Reva, A. Krek, J. Ji, X. Song, W. Liu, R. Hong, L. Yao, L. Blumenberg, S. R. Savage, M. C. Wendl, B. Wen, K. Li, L. C. Tang, M. A. MacMullan, S. C. Avanessian, M. H. Kane, C. J. Newton, M. Cornwell, R. B. Kothadia, W. Ma, S. Yoo, R. Mannan, P. Vats, C. Kumar-Sinha, E. A. Kawaler, T. Omelchenko, A. Colaprico, Y. Geffen, Y. E. Maruvka, F. da Veiga Leprevost, M. Wiznerowicz, Z. H. Gümüş, R. R. Veluswamy, G. Hostetter, D. I. Heiman, M. A. Wyczalkowski, T. Hiltke, M. Mesri, C. R. Kinsinger, E. S. Boja, G. S. Omenn, A. M. Chinnaiyan, H. Rodriguez, Q. K. Li, S. D. Jewell, M. Thiagarajan, G. Getz, B. Zhang, D. Fenyö, K. V. Ruggles, M. P. Cieslik, A. I. Robles, K. R. Clauser, R. Govindan, P. Wang, A. I. Nesvizhskii, L. Ding, D. R. Mani, S. A. Carr, Clinical Proteomic Tumor Analysis Consortium, Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. *Cell* **182**, 200–225.e35 (2020).

133. J. Gao, B. A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S. O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, E. Cerami, C. Sander, N. Schultz, Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, l1 (2013).

134. Cancer Genome Atlas Research Network, Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).

135. R. J. Hartmaier, L. A. Albacker, J. Chmielecki, M. Bailey, J. He, M. E. Goldberg, S. Ramkissoon, J. Suh, J. A. Elvin, S. Chiacchia, G. M. Frampton, J. S. Ross, V. Miller, P. J. Stephens, D. Lipson, High-Throughput Genomic Profiling of Adult Solid Tumors Reveals Novel Insights into Cancer Pathogenesis. *Cancer Res.* **77**, 2464–2475 (2017).

136. The AACR Project GENIE Consortium, AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer Discov.* **7**, 818 (2017).

137. H. Kadara, M. Choi, J. Zhang, E. R. Parra, J. Rodriguez-Canales, S. G. Gaffney, Z. Zhao, C. Behrens, J. Fujimoto, C. Chow, Y. Yoo, N. Kalhor, C. Moran, D. Rimm, S. Swisher, D. L. Gibbons, J. Heymach, E. Kaftan, J. P. Townsend, T. J. Lynch, J. Schlessinger, J. Lee, R. P. Lifton, I. I. Wistuba, R. S. Herbst, Whole-exome sequencing and immune profiling of early-stage lung adenocarcinoma with fully annotated clinical follow-up. *Ann. Oncol.* **28**, 75–82 (2017).

138. M. Lawrence, R. Gentleman, V. Carey, rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841–1842 (2009).

139. F. Manders, A. M. Brandsma, J. de Kanter, M. Verheul, R. Oka, M. J. van Roosmalen, B. van der Roest, A. van Hoeck, E. Cuppen, R. van Boxtel, MutationalPatterns: the one stop shop for the analysis of mutational processes. *BMC Genomics* **23**, 134 (2022).

140. L. B. Alexandrov, J. Kim, N. J. Haradhvala, M. N. Huang, A. W. Tian Ng, Y. Wu, A. Boot, K. R. Covington, D. A. Gordenin, E. N. Bergstrom, S. M. A. Islam, N. Lopez-Bigas, L. J. Klimczak, J. R. McPherson, S. Morganella, R. Sabarinathan, D. A. Wheeler, V. Mustonen, G. Getz, S. G. Rozen, M. R. Stratton, The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).

141. R. Rosenthal, N. McGranahan, J. Herrero, B. S. Taylor, C. Swanton, deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, 1–11 (2016).

142. S. Devarakonda, D. Morgensztern, R. Govindan, Genomic alterations in lung adenocarcinoma. *Lancet Oncol.* **16** (2015).

143. H. Greulich, The Genomics of Lung Adenocarcinoma: Opportunities for Targeted Therapies. *Genes Cancer* **1**, 1200 (2010).

144. J. S. Seo, Y. S. Ju, W. C. Lee, J. Y. Shin, J. K. Lee, T. Bleazard, J. Lee, Y. J. Jung, J. O. Kim, J. Y. Shin, S. B. Yu, J. Kim, E. R. Lee, C. H. Kang, I. K. Park, H. Rhee, S. H. Lee, J. I. Kim, J. H. Kang, Y. T. Kim, The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res.* **22** (2012).

145. S. Li, Y. L. Choi, Z. Gong, X. Liu, M. Lira, Z. Kan, E. Oh, J. Wang, J. C. Ting, X. Ye, C. Reinhart, X. Liu, Y. Pei, W. Zhou, R. Chen, S. Fu, G. Jin, A. Jiang, J. Fernandez, J. Hardwick, M. W. Kang, I H, H. Zheng, J. Kim, M. Mao, Comprehensive Characterization of Oncogenic Drivers in Asian Lung Adenocarcinoma. *J. Thorac. Oncol.* **11** (2016).

146. O. Abril-Pla, V. Andreani, C. Carroll, L. Dong, C. J. Fonnesbeck, M. Kochurov, R. Kumar, J. Lao, C. C. Luhmann, O. A. Martin, M. Osthege, R. Vieira, T. Wiecki, R. Zinkov, PyMC: a modern, and comprehensive probabilistic programming framework in Python. *PeerJ Comput. Sci.* **9**, e1516 (2023).

147. S. S. Wilks, The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *aoms* **9**, 60–62 (1938).

Acknowledgements

We thank Jeff Mandell for his advice on mutational signature attribution and mutation rate estimation.

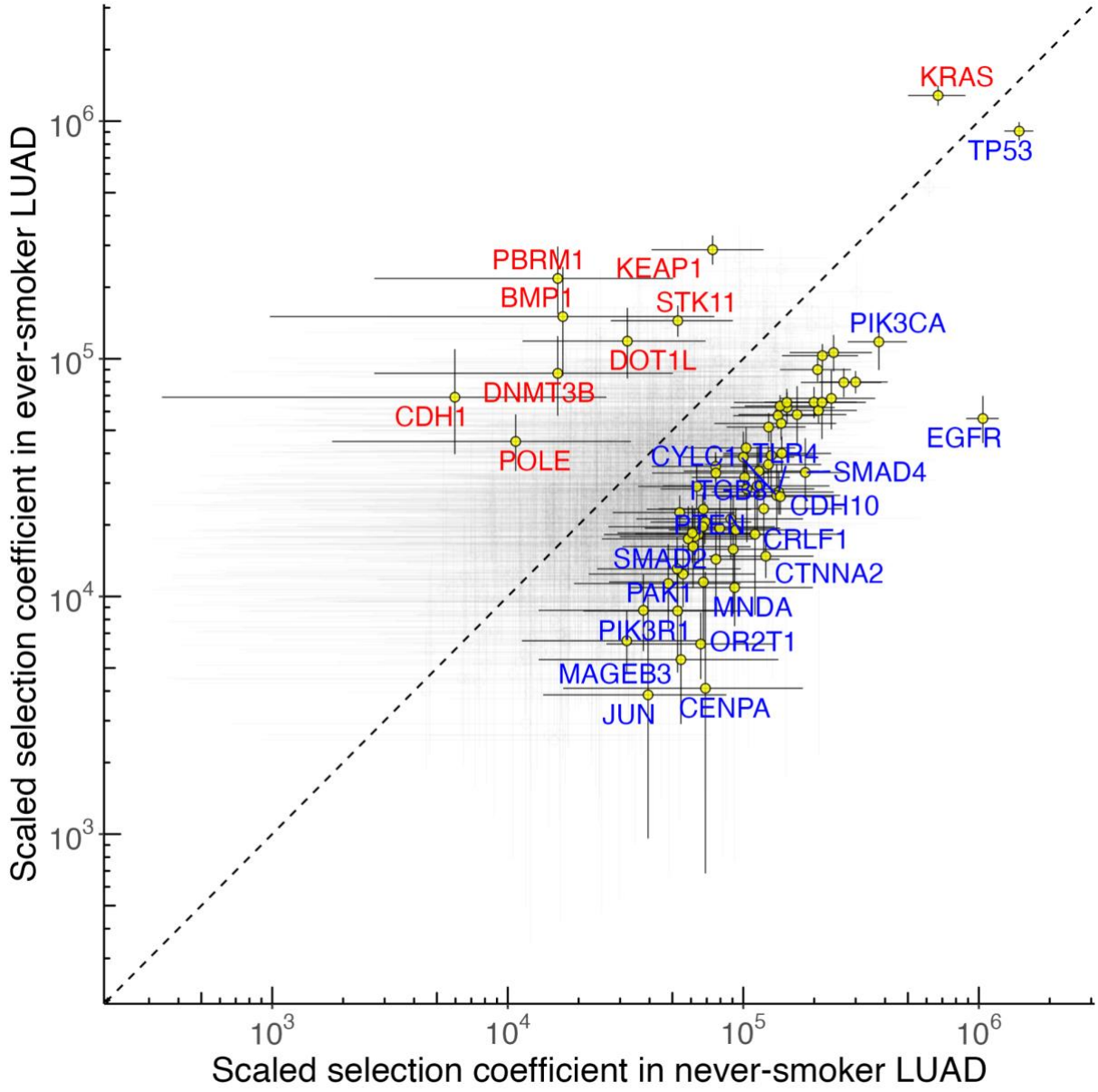
Funding: Yale College First-Year Summer Research Fellowship in the Sciences and Engineering (KD)

Author contributions: Conceptualization: JPT, JAA-M, KD; Data curation: KD; Formal analysis: KD, JAA-M, Investigation: KD, JAA-M, JPT, Methodology: KD, JAA-M, Project administration: JPT, Software: KD, JAA-M, Supervision: JPT, JAA-M; Visualization: KD, JAA-M; Writing – original draft: KD; Writing – review & editing: KD, JPT, JAA-M

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: All data and code underlying this study are available in the main text or the supplementary materials. Particularly, they can be accessed in the project Github repository: https://github.com/Townsend-Lab-Yale/lung-smoking.

Supplementary Materials



Supplementary Figure 1. Strengths of somatic selection on 1200 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).

Table S1. Characteristics of datasets included in study.

Study	Institution/project	Smoking status classified	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	Yes	Whole-exome sequencing	3 / 567[*]	567
Jordan et al., 2017	Memorial Sloan Kettering Cancer Center (MSKCC)	Yes	Targeted sequencing (MSK-IMPACT 341, MSK-IMPACT 410)	125 / 372[†]	372
Chen et al., 2020	OncoSG	Yes	Whole-exome sequencing	At least 86 / 300[‡, §]	300
Zhang et al., 2021	National Cancer Institute	Yes	Whole-genome sequencing	None	187
Rizvi et al., 2018	MSKCC	Yes	Targeted sequencing (MSK-IMPACT 341, MSK-IMPACT 410, MSK-IMPACT 468)	186 / 186[‖]	186
Imielinski et al., 2012.	Broad Institute	Yes	Whole-exome sequencing, whole-genome sequencing, or both	None	180
Gillette et al., 2020	Clinical Proteomic Tumor Analysis Consortium	Yes	Whole-exome sequencing	None	108
Kadara et al., 2017	Yale University	Yes	Whole-exome sequencing	Not available	108
Jamal-Hanjani et al., 2017	TracerX	Yes	Whole-exome sequencing	None	61
Rizvi et al., 2015	MSKCC	Yes	Whole-exome sequencing	28 / 28[¶]	28
AACR Project GENIE v9.0.0	AACR Project GENIE	No	Targeted sequencing (multiple panels used)	Not available	5996
Hartmaier et al., 2017	Foundation Medicine Adult Cancer Clinical Dataset	No	Targeted sequencing (FoundationOne CDx)	Not available	1205

Ding et al., 2009	Tumor Sequencing Project	No	Targeted sequencing (custom panel of 623 genes)	Not available	159

[*] Treatment not indicated
[†] Combination of targeted therapy, immunotherapy, and/or chemotherapy
[‡] Treatment information is incomplete for 92 samples
[§] Tyrosine-kinase inhibitor and/or chemotherapy
[‖] Anti–PD-(L)1 therapy, sometimes combined with anti–CTLA-4 therapy
[¶] Pembrolizumab