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

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

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

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

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

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

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

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

Abstract

The mutagenic effects of tobacco smoke are well-reported. However, tobacco smoke also likely alters the somatic selective pressures that drive lung cancer evolution, and these altered pressures have never been quantified nor compared to the oncogenic effects of tobacco-induced mutagenesis. This comparison is necessary for prediction of targeted therapy efficacy given exposure history. Here, we estimated background mutation rates, somatic selection, and selective epistasis for 21 driver genes in ever- and never-smoker lung adenocarcinoma (ES- and NS-LUAD).

As expected, mutation rates were gene-specifically elevated in ES-LUAD. However, variant prevalence differences between ES- and NS- LUAD could not be explained by differential mutagenesis. *KRAS*, *KEAP1*, and *STK11* mutations were more strongly selected in ES-LUAD. *EGFR*, *PIK3CA*, and *SMAD4* mutations were more strongly selected in NS-LUAD, wherein *EGFR* mutations were associated with upregulation of epithelial-mesenchymal transition—a functional consequence specific to exposure history.

Epistasis was pervasive and context-dependent: ES-LUAD exhibited more synergy and less antagonism, implying a broader, more navigable adaptive landscape than NS-LUAD. We also performed the first systematic quantification of higher-order-than-pairwise epistasis, revealing sub-additive and emergent synergies that shape cancer evolution. Our results disambiguate the mutagenic and selective effects of tobacco smoke, demonstrating how environmental insults reshape cancer evolution and enabling prediction of therapeutic efficacy based on smoking history and somatic genotype.

Author Summary

Smoking is a well-known cause of lung cancer because it increases the number of mutations in lung cells. But tobacco smoke also causes inflammation and other physiological changes that might influence how tumors evolve. Using DNA from lung tumors, we asked whether these changes make certain mutations more helpful to cancer growth and thereby better targets for therapy. We also asked whether some mutations were more helpful to cancer growth in those who haven't smoked. To do so, we estimated not only how often mutations occurred, but also how strongly they were favored during cancer development. Focusing on 21 key lung cancer genes, including *KRAS*, *EGFR,* and *TP53*, we found that mutations in some, such as *KRAS* and *KEAP1* were more strongly favored in smokers. Mutations in genes such as *EGFR* and *PIK3CA* were more strongly favored in non-smokers, where they were also linked to non-smoker-specific changes in gene activity. We also found that mutations in tumors interact differently depending on smoking history: smokers' tumors showed more cooperation between mutations, which may help them evolve in more flexible ways. This study shows how smoking reshapes the genetic paths lung cancers take—and why a patient's smoking history matters when choosing targeted treatments. Understanding these evolutionary patterns can help improve treatment decisions and lead to more personalized care for lung cancer patients.

Introduction

Tobacco smoke is well known to cause cancer, in part through its mutagenic effects [1]. Carcinogen exposure from tobacco generates distinct patterns of mutations [2], a subset of which then contribute to oncogenesis [3–9]. In addition to this mutagenic effect, tobacco smoke induces physiological changes in the lung—inflammation, oxidative stress, immune dysregulation, and tissue damage [1,10–12]—that can alter the lung microenvironment and thereby the selective pressures acting on specific somatic mutations. These microenvironmental

68 disruptions could modulate the fitness advantages conferred by specific mutations, thereby
69 influencing their likelihood of contributing to cancer development. Indeed, several studies have
70 suggested that, in principle, endogenous or exogenous alterations to the microenvironment
71 caused by factors such as aging or tobacco exposure likely shape the trajectories of cancer
72 evolution [13–17]. However, the extent to which such physiological effects alter the adaptive
73 landscape of cancer has not previously been quantified. Therefore, quantification of the
74 adaptive impacts of a highly disruptive exogenous insult such as tobacco smoke is crucial to
75 understanding how environmental exposures shape the evolution of individual cancer
76 evolutionary trajectories.

77 Among lung cancer subtypes, LUAD is the most common and accounts for the highest
78 proportion of cases in never-smokers [18]. As of 2002, nearly 23% of LUADs in the United
79 States occurred in never-smokers [19], a proportion that has been rising as cigarette usage
80 declines [20]. Numerous studies have reported that LUAD exhibits distinct somatic variant
81 profiles in never-smokers compared to ever-smokers, which has been taken as evidence of their
82 divergent genetic vulnerabilities [21–25]. However, differences in variant prevalence alone are
83 confounded as to whether they are a consequence of effects of tobacco-induced mutagenesis
84 or a consequence of tobacco-altered selection dynamics. Therefore, quantification of the
85 relative oncogenic roles of the mutagenic and physiological effects of smoking in LUAD will be
86 increasingly crucial to determination of optimal use of targeted therapies for patients with and
87 without a prior smoking history.

88 Computational models now enable inference of baseline somatic mutation rates from
89 synonymous mutations observed in tumor genomes—implicitly [6] or explicitly [26] accounting
90 for carcinogen-induced mutagenesis, such as that caused by tobacco smoke. Comparing these
91 baseline mutation rates to the observed prevalence of nonsynonymous mutations enables
92 estimation of somatic selection acting on specific substitutions [8,27,28]. This framework

93 provides a foundation for disentangling the oncogenic contributions of the mutagenic and
94 selective effects of smoking. Moreover, selection on each mutation is likely influenced not only
95 by the microenvironmental effects of tobacco smoke, but also by pairwise [28–30] and higher-
96 order [28] selective epistasis between mutations. In selective epistasis, mutations affect the
97 selection operating on other mutations. These interactions between the selective effects of
98 mutations shape the adaptive landscape of cancer evolution [5,28,31,32]. Understanding how
99 tobacco smoke influences not just mutation and selection, but also epistasis, is essential for a
100 complete picture of LUAD evolution.

101 Because knowledge of epistasis is crucial to inference of cancer-driving mutation modules
102 and synthetic vulnerabilities, a number of algorithms have been developed that attempt to detect
103 epistasis between mutations or sets of mutations by quantifying over- and under-
104 representation [33,34] in the context of heterogeneous mutation rates [30,35–37]. However, with
105 the exception of Coselens [30], these methods do not enable meaningful quantitative
106 comparisons of the effects of mutations in distinct somatic genotypes. This quantification is a
107 crucial step for prioritization and translation of knowledge of epistasis into effective therapeutic
108 strategies. Moreover, all of these approaches are restricted to pairwise inferences [30,34] or to
109 inference of discrete modules [33,35–37], and are not well suited to the quantitative inference of
110 higher-order epistasis (involving more than two mutations). When present, higher-order
111 epistasis substantially influences the adaptive landscape and shapes evolutionary trajectories
112 [38–43], including the degree to which selective epistasis diminishes along the adaptive
113 trajectory [44–47]. Systematic characterization of the presence and biological significance of
114 higher-order epistasis to cancer evolution has been challenging. Recent theory facilitates
115 understanding of higher-order epistasis in cancer by providing the selection strength of
116 mutations for each somatic genotype [28]. This novel approach has not yet been applied in the
117 context of environmental risk factors that could also affect selection, such as tobacco smoke.

118 Accordingly, analysis of the effect of smoke on the adaptive landscape of LUAD will benefit from
119 consideration of pairwise and higher-order epistatic effects and potential differences in epistasis
120 between microenvironments.

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

133 # Results

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

136 To disentangle the relative contributions of mutation and selection to observed mutation
137 frequencies in LUAD, we deconvolved the prevalence of somatic variants in our curated set of
138 21 well-supported driver genes into estimates of the background oncogenic mutation rates and
139 scaled selection coefficients. Across all smoking-status-classified LUAD samples in our dataset
140 (n = 1,722), *TP53* and *KRAS* were the most frequently mutated genes, followed distantly by
141 *EGFR*, *KEAP1*, and *STK11* (**Fig. 1A**). Background oncogenic mutation rates for the majority of
142 these genes spanned slightly more than an order of magnitude—between 7.6×10^{-8} and

143 8.7 × 10^{-7} (**Fig. 1B**). When compared to the observed variant prevalences, this variation in

144 background oncogenic mutation rate frequently implies substantial and variable impacts of

145 somatic selection on the prevalence of each variant.

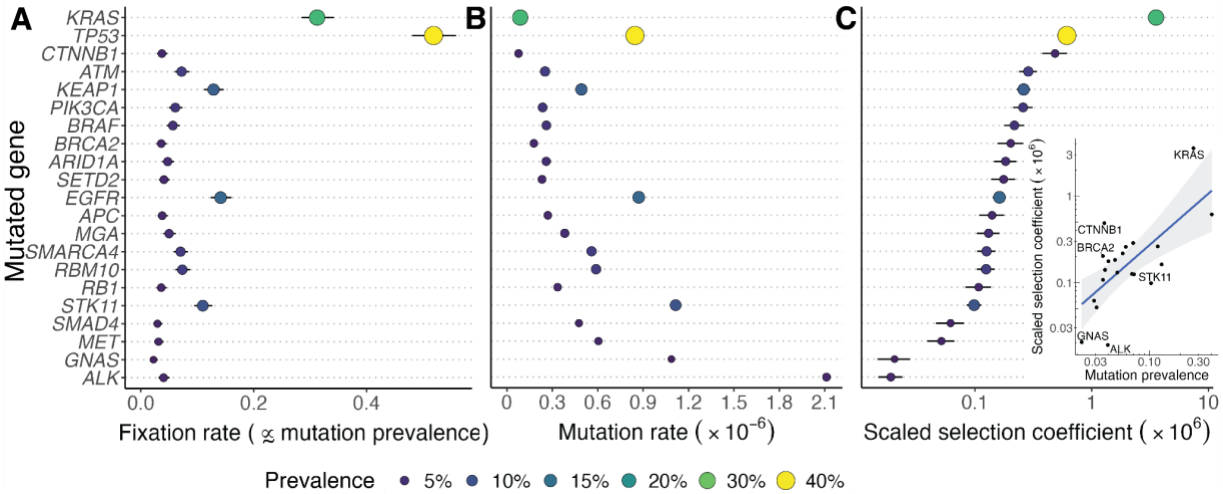


Figure 1. Fixation rates and scaled somatic selection coefficients for 21 known and potential driver mutations among 1,722 lung adenocarcinomas. (**A**) Poisson-corrected fixation rates (rates at which each mutation occurs and becomes fixed within the population). (**B**) Background rate at which oncogenic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

146

147 Scaled selection coefficients for oncogenic mutations in each gene exhibited a long-tailed

148 distribution (**Fig. 1C**), with mutations in *KRAS* experiencing the strongest selection, followed

149 distantly by mutations of *TP53* and *CTNNB1*, then mutations in *ATM*, *KEAP1*, and *PIK3CA*, and

150 then other genes that experienced weaker but still substantial selection (**Fig. 1C**). Notably,

151 *CTNNB1* mutations were present in only 3.7% of tumors. Nevertheless, knowledge of the low

152 underlying oncogenic mutation rate of *CTNNB1* indicates that these infrequent mutations

153 experienced stronger selection than did the more frequent mutations in *STK11, RBM10*, and

154 *SMARCA4*. Generally, selection and mutation prevalence were correlated ($r = 0.68$, $P < 0.001$),

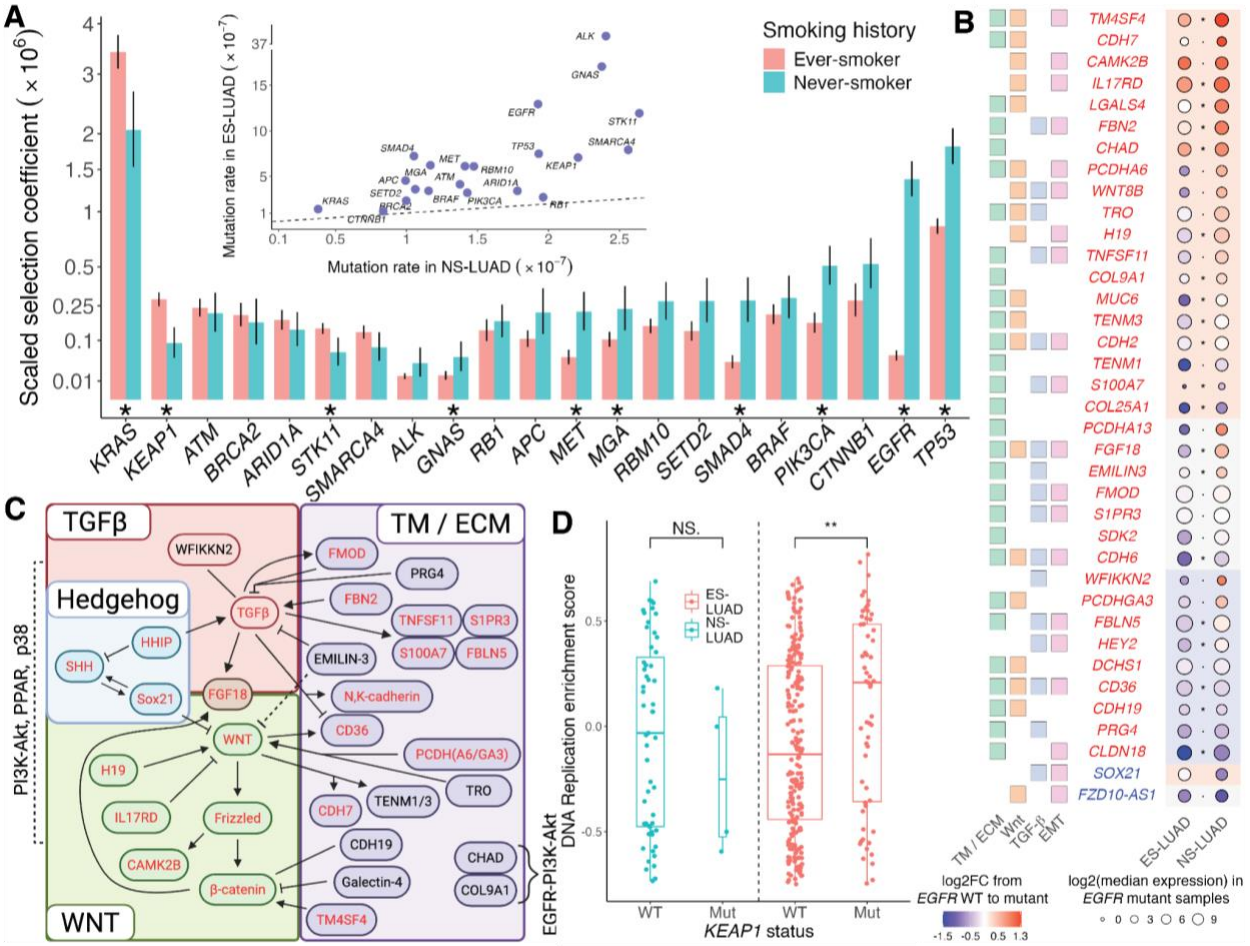
155 but several discrepancies were present between the prevalence of a mutation and the relative

156 intensity of selection it experienced (**Fig. 1C,** inset). These discrepancies highlight the

157 importance of reporting selection as the effect size of mutations, rather than reporting solely
158 prevalence, which is only moderately associated with strength of selection.
159 The oncogenic mutation rates and selection coefficients presented above were quantified
160 based on all tumors, regardless of smoking status. However, both the mutagenic processes and
161 selective advantages of mutations were plausibly influenced by the genotoxic and physiological
162 effects of tobacco smoke. To assess these distinct effects, we stratified tumors by smoking
163 status and performed separate analyses of somatic selection in ES- and NS-LUAD tumors.
164

165 **Mutation and selection differ substantially between ever-smoker and never-**
166 **smoker LUAD**

167 Stratification of LUAD tumors by smoking status based on mutational signature attributions
168 and clinical annotation yielded 1,066 ES-LUADs and 656 NS-LUADs. In each group, we
169 independently estimated oncogenic mutation rates and selection coefficients for the 21 well-
170 established driver genes (**Fig. 2**). Both mutation rates and selection pressures differed markedly
171 between ES- and NS-LUAD. Oncogenic mutation rates were elevated in ES-LUAD across all 21
172 genes. Increases ranged from modest (39%–43% for *RB1* and *CTNNB1*) to extreme (571–
173 1474% for *EGFR*, *SMAD4*, *GNAS*, and *ALK*; **Fig. 2A inset**). These elevated mutation rates
174 demonstrate the intense mutagenic effect of tobacco smoking on the peripheral tissue of the
175 lung.



176

Figure 2. Mutational and adaptive landscapes of ever-smoker lung adenocarcinoma (ES-LUAD) and never-smoker lung adenocarcinoma (NS-LUAD). (**A**) Strengths of somatic selection (*y* axis is square-root scaled; *: non-overlapping 95% confidence intervals) for mutations of 21 known and potential driver genes in NS-LUAD (teal bars; n = 656) and ES-LUAD (salmon bars; n = 1,066), averaged across somatic genotypes. Inset: Background oncogenic mutation rates for these genes in ES-LUAD and NS-LUAD (dashed line: equality). (**B**) Differential expression of genes between *EGFR* wild-type (WT) and *EGFR* mutant samples in ES-LUAD (n = 273) and NS-LUAD (n = 59; log2FC: $\log_2$ of fold change in expression) for genes deemed to have a significant gene expression interaction between *EGFR* allelic status and smoking status. Color of each gene name denotes whether the change in expression was more positive (red) or more negative (blue) in NS-LUAD compared to ES-LUAD (*: $P < 0.05$, •: $P < 0.1$, *SOX21*: $P = 0.116$; background color: change in gene expression from normal (n = 33) to LUAD (n = 332) is increased [pink], decreased [light blue], or insignificant [gray]). Leftward squares designate whether the gene produces a transmembrane (TM) or extracellular matrix (ECM) protein as well as whether it facilitates or interacts with the Wnt, and TGF-β and the epithelial-mesenchymal transition (EMT) signaling pathways. (**C**) Interactions between proteins (red: implicated in EMT) whose *EGFR*-mutation-associated expression changes differ between ES-LUAD and NS-LUAD. (**D**) Sample-specific DNA replication expression scores for *KEAP1* WT and mutant ES- and NS-LUAD (**: $P < 0.01$).

177 We found that reported differences between ES- and NS-LUAD in the prevalences of key
178 driver mutations [21,22,48] were often driven by differential selection, rather than solely by
179 differential mutagenesis (**Fig. 2A**). Among the 21 driver genes analyzed, *TP53*, *EGFR*, *PIK3CA*,
180 *SMAD4*, *MET*, *MGA*, and *GNAS* mutations experienced stronger positive selection in NS-LUAD.
181 *STK11*, *KEAP1*, and *KRAS* mutations experienced stronger positive selection in ES-LUAD.
182 *RB1*, *BRAF*, and *BRCA2* mutations experienced similar selection in both cancer types.
183 Most mutations that experienced significantly different selection between ES- and NS-
184 LUAD were strongly selected in one context and weakly selected in the other, demonstrating
185 pronounced context-dependence in LUAD evolution. In ES-LUAD, *KEAP1* mutations
186 experienced the third strongest selection—after *TP53* and *KRAS*—and experienced over three
187 times stronger selection than in NS-LUAD. Likewise, *STK11* mutations experienced over twice
188 as much selection in ES-LUAD. Indeed, both *KEAP1* and *STK11* mutations are enriched in
189 ever-smoker populations [23,49,50]. *EGFR* mutations exhibited the most dramatic difference:
190 they were more than twenty times more strongly selected in NS- than ES-LUAD, and were also
191 the third most strongly selected mutations in NS-LUAD after *TP53* and *KRAS*. *EGFR* mutations
192 are much more common in NS-LUAD than in ES-LUAD [21,51]. Similarly, *PIK3CA*, *SMAD4*, and
193 *MET* mutations experienced over three times stronger selection in NS- relative to ES-LUAD.
194 To investigate mechanisms underlying the differential selection of *EGFR* mutations, we
195 assessed their differential transcriptional effects in ES- and NS-LUAD using TCGA expression
196 data. In fewer than one percent of genes did the expression fold change from *EGFR* wild type
197 samples compared to *EGFR* mutant samples differ significantly between ES- and NS-LUAD
198 (**Fig. 2B**). However, several of these genes exhibited markedly stronger upregulation or
199 downregulation in EGFR-mutant NS-LUAD compared to ES-LUAD. Genes that were
200 upregulated were frequently components of Wnt, TGF-β, and Hedgehog signaling pathways or
201 of the extracellular matrix (**Fig. 2B–C**). Indeed, in an analysis of gene sets upregulated in NS-

202 versus ES-LUAD, the Hallmark gene set associated with epithelial-mesenchymal transition was
203 the most enriched. On the other hand, *KEAP1* mutations were associated with a substantially
204 higher DNA replication enrichment score in ES-LUAD but not in NS-LUAD (**Fig. 2D**). These
205 examples relate microenvironmental context to modifications of the downstream effects of driver
206 mutations and their accompanying selective advantage.

207 Extending our analysis from 21 to 1,200 LUAD-associated genes revealed a striking
208 asymmetry in selection by smoking status: only six genes—*KRAS*, *KEAP1*, *STK11*, *PBRM1*,
209 *DNMT3B,* and *CDH1*—were significantly more strongly selected in ES- than NS-LUAD, whereas
210 87 experienced stronger selection in NS-LUAD (**Fig. S1**). Among the mutations most strongly
211 favored in NS-LUAD were those in *EGFR*, *SMAD4*, *SMAD2*, *PTEN*, *PIK3R1, PAK1*, *FGF3*,
212 *GATA3*, *JUN*, *CTNNA2*, and *TLR4*. Mutations often experienced weaker selection in ES-LUAD,
213 which is more heavily mutated and frequently possesses multiple known drivers, than in NS-
214 LUAD, which often features a single known driver (**Fig. S1**).

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

222

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

224 To assess how frequently and how strongly the evolutionary trajectory of LUAD is affected
225 by pairwise epistasis, we quantified the effects of driver mutations both in isolation (in tumors
226 that were wild type at a paired gene) and in the presence of each paired gene mutation. For ES-

227 and NS-LUAD, we evaluated all 420 possible gene pairs among the 21 driver genes: pairwise
228 interactions were identified that modulate the selective advantage of mutations, and did so in a
229 smoking-dependent manner.
230

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

233 To investigate directional epistasis between two common LUAD drivers, *TP53* and *RB1*,
234 we first stratified tumors by smoking status and somatic genotype at these two loci, and then
235 estimated the fixation rates of each mutation from each somatic genotype (**Fig. 3A**). In ES-
236 LUAD, *TP53* mutations exhibited a high fixation rate when *RB1* was wild-type but were inferred
237 to exhibit a negligible fixation rate when *RB1* was mutated (***λ*, Fig. 3B**). Conversely, *RB1*
238 mutations exhibited a low fixation rate when *TP53* was wild-type but exhibited a high fixation
239 rate when *TP53* was mutated. These asymmetries indicate the most likely evolutionary
240 sequence in ES-LUAD: mutation of *TP53* followed by mutation of *RB1*. In wild-type tumors, the
241 mutation rate of TP53 was over double the mutation rate of *RB1* ($\mu_{WT\rightarrow TP53} = 7.4 \times 10^{-7}$,
242 $\mu_{WT\rightarrow RB1} = 2.0 \times 10^{-7}$); the same was true of mutant tumors ($\mu_{RB1\rightarrow RB1+TP53} = 8.9 \times 10^{-7}$,
243 $\mu_{TP53\rightarrow RB1+TP53} = 3.6 \times 10^{-7}$; **Fig. 3B**). Accordingly, *RB1* variants in a *TP53*-mutant genotype must
244 have experienced comparable selection to *TP53*-variants in a wild-type genotype (γ, **Fig. 3B**),
245 despite the substantially lower fixation rate of the *RB1* variants. This pattern reflects synergistic
246 epistasis: prior mutation of *TP53* increases selection for *RB1*. In contrast, the reverse order
247 exhibited antagonistic epistasis—*TP53* was less strongly selected in the presence of *RB1*
248 mutation—though this last effect did not achieve statistical significance.

249

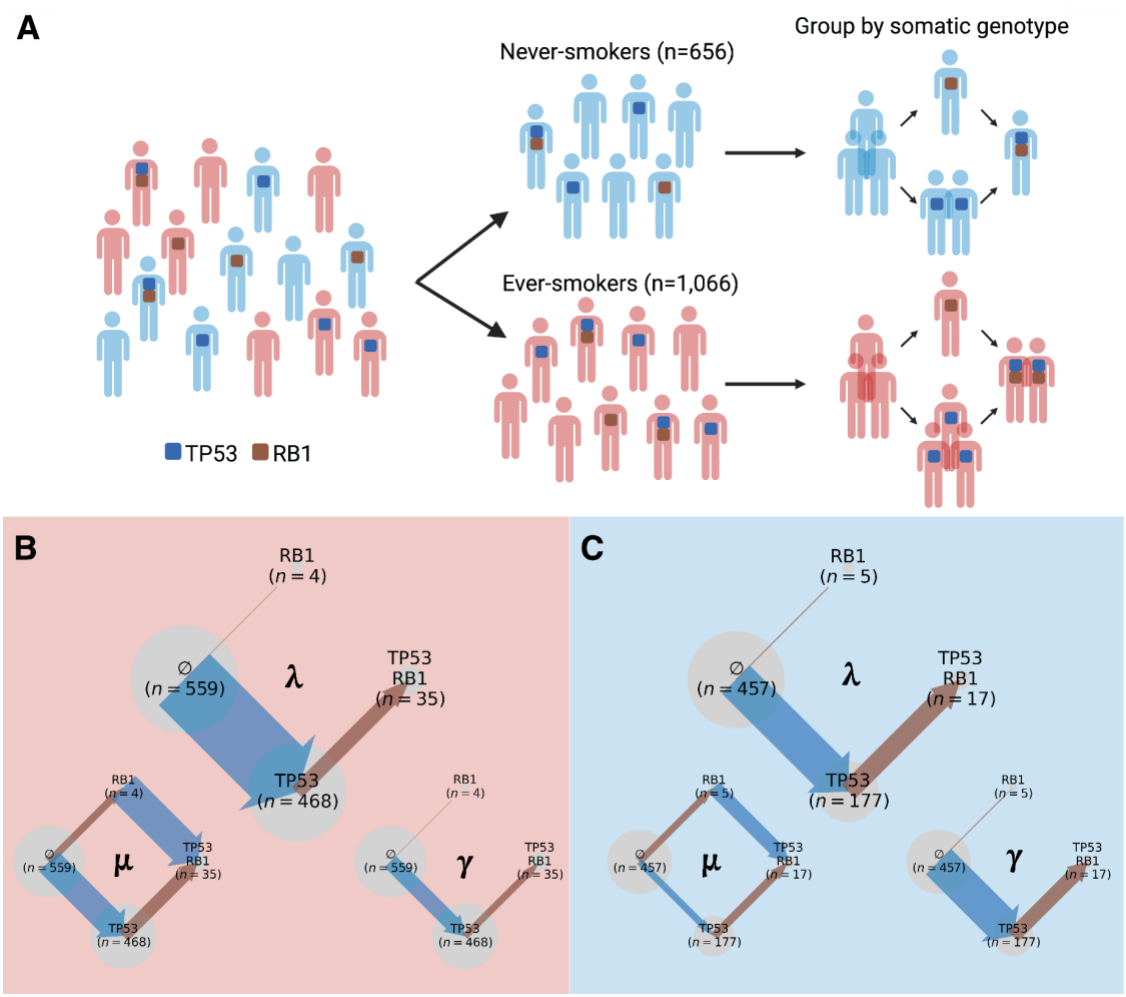


Figure 3. Pairwise selective epistasis inferred within the evolutionary trajectories of lung adenocarcinoma (LUAD), depicted for *TP53* and *RB1*. (**A**) Genomic sequence data from lung adenocarcinomas can be stratified by clinically- and genetically-determined patient smoking status (blue: never-smokers;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

250 In NS-LUAD, the fixation rate for *TP53* mutations from a wild-type genotype was lower

251 than in ES-LUAD, and the fixation rate for *RB1* mutations in a *TP53*-mutant background was

252 higher (λ, **Fig. 3C**). In single-mutant tumors, the mutation rate of *TP53* is nearly double that of

253 *RB1* ($\mu_{RB1\rightarrow RB1+TP53}$ = 4.0 × 10^{-7}, $\mu_{TP53\rightarrow RB1+TP53}$ = 2.2 × 10^{-7}), but in wild-type NS-LUAD tumors,

254 the mutation rate of *TP53* was similar to the mutation rate of *RB1* ($\mu_{WT\rightarrow TP53}$ = 1.48 × 10^{-7},

255 $\mu_{WT\rightarrow RB1}$ = 1.8 × 10^{-7}). This similarity of mutation rates contrasts with the disparity in ES-LUAD.

256 Despite these differences, the inferred evolutionary trajectory mirrored that of ES-LUAD: the

257 most likely (λ) and most adaptive (γ) path involved mutation of *TP53* followed by mutation of

258 *RB1*. As in ES-LUAD, prior *TP53* mutation significantly synergistically enhanced selection for

259 *RB1* mutation, and prior *RB1* mutation antagonized selection for *TP53* mutation, without

260 reaching statistical significance (γ, **Fig. 3C**).

261

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

263 Analysis of all pairwise combinations of the 21 reputed LUAD driver genes revealed

264 frequent synergistic and antagonistic epistasis. In ES-LUAD, synergistic interactions were

265 especially pronounced: for genes whose mutations were under strong positive selection (with

266 the notable exceptions of *TP53* and *KRAS*) synergistic epistasis led to two- to eight-fold

267 increases in selection. For example, selection on mutations in *KEAP1, ATM*, *BRCA2*, *PIK3CA*,

268 and *ARID1A* was markedly enhanced in the presence of prior *TP53* or *KRAS* mutations

269 (**Fig. 4A**). Additional strong synergistic interactions included those affecting mutants of *ARID1A*

270 in the context of somatic genotypes with *ATM* or *RBM10* mutations; mutants of *STK11* in the

271 context of mutated *KEAP1*; and mutants of *ATM* in the context of mutated *STK11*. Mutations in

272 the non-*TP53* tumor suppressors *BRCA2*, *SETD2*, *RBM10*, *RB1*, *MGA*, and *APC*, as well as in

273 the chromatin-remodeling genes *ARID1A* and *SMARCA4*, experienced moderate baseline

274 selection that was frequently enhanced by the presence of diverse prior mutations. These genes

275 never experienced antagonistic epistasis with each other and rarely with other mutations.

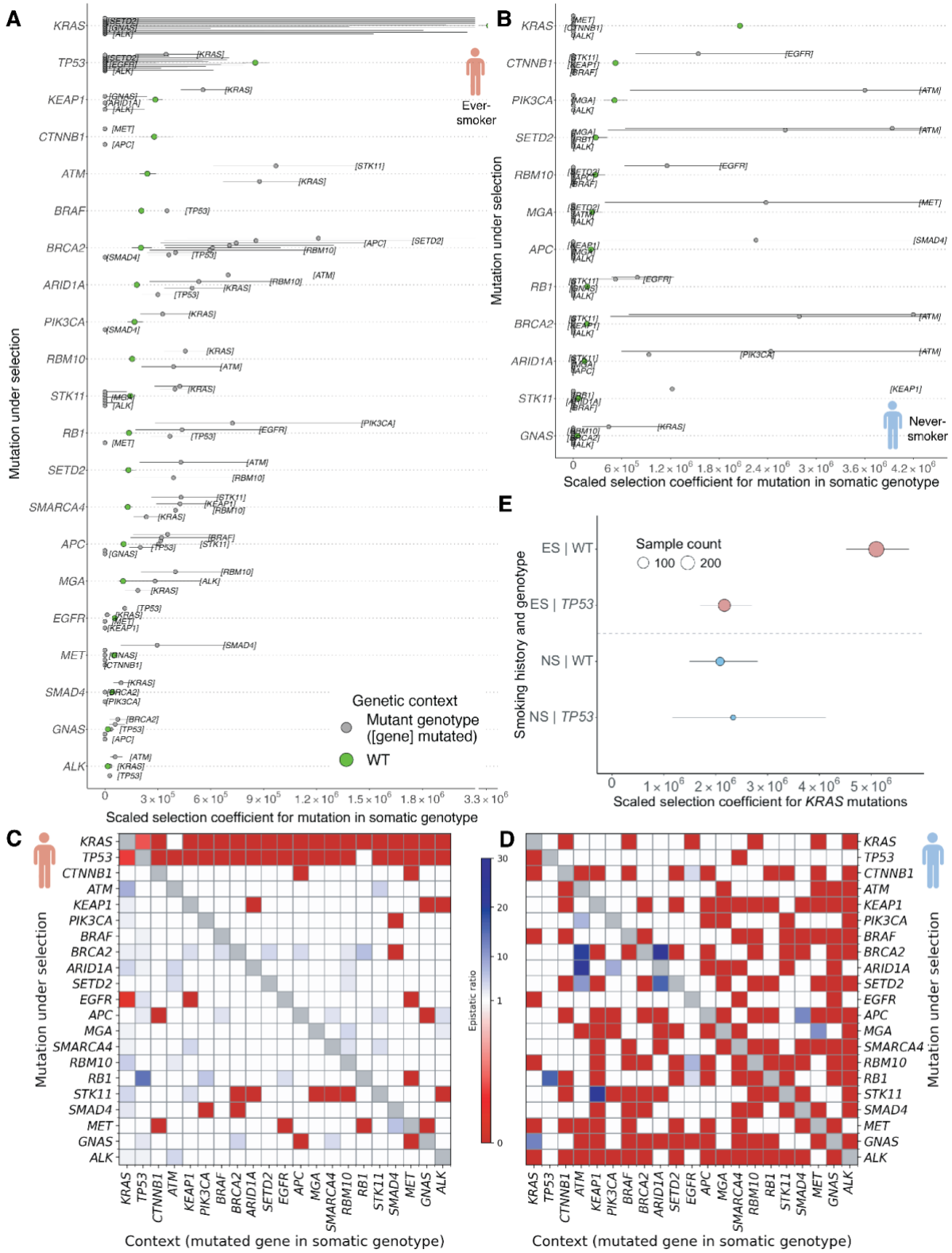

A
Mutation under selection
Ever-smoker
Genetic context
Mutant genotype ([gene] mutated)
WT
Scaled selection coefficient for mutation in somatic genotype
B
Mutation under selection
Never-smoker
Scaled selection coefficient for mutation in somatic genotype
E
Smoking history and genotype
ES | WT
ES | TP53
NS | WT
NS | TP53
Sample count
100
200
Scaled selection coefficient for KRAS mutations
C
Mutation under selection
Context (mutated gene in somatic genotype)
D
Epistatic ratio
Mutation under selection
Context (mutated gene in somatic genotype)

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

276

277 Patterns of epistasis involving oncogenes were heterogeneous. Notably, *KRAS* and
278 *CTNNB1* mutations showed no elevation of selection in the presence of any prior mutations in
279 the 21-gene set. Conversely, selection for mutations of *GNAS* and *ALK* was enhanced by three
280 prior mutations (**Fig. 4A**). For many oncogenes, selection was enhanced by prior *TP53*
281 mutation, but was antagonized by mutations in other oncogenes such as *MET*. Across all tested
282 genes, the median effect size of statistically significant synergistic epistasis was a 3.6× increase
283 in selection, and the magnitude of this synergy did not correlate with the baseline strength of
284 selection for either the first or the second mutation.

285

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

288 In ES-LUAD, we identified eight gene pairs exhibiting statistically significant antagonistic
289 sign epistasis, in which selection for a driver mutation shifted from positive to negative in the
290 presence of another mutated gene. Mutations in *CTNNB1* and *GNAS* became selectively
291 disadvantageous in the presence of *MET* and *APC* mutations. Similarly, selection for *SMAD4*
292 mutations was antagonized by prior *BRCA2* and *PIK3CA* mutations, and *MET* mutations were
293 antagonized by mutations of *RB1*, *EGFR*, *CTNNB1*, and *GNAS*. In each of these eight cases,
294 the reciprocal mutation order also induced negative selection (e.g., mutated *CTNNB1* induced

295 negative selection for *MET* mutations). For over a hundred other ordered pairs, antagonistic
296 sign epistasis was estimated, without reaching statistical significance. Antagonistic shifts in the
297 magnitude of selection, in which selection for a mutation remained positive but was substantially
298 reduced, were indicated in 33 cases, of which three were significant: the presence of *KRAS*
299 mutations caused a 64% reduction in selection for *TP53* mutations and an 81% reduction in
300 selection for *EGFR* mutations, and *TP53* mutations caused a 58% reduction in selection for
301 *KRAS* mutations (**Fig. 4A**). Across all analyses of ES- and NS-LUAD, new mutations in *TP53*
302 and *KRAS* were never inferred to experience significantly stronger selection in the context of
303 another driver mutation.
304

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

306 Compared to ES-LUAD, NS-LUAD exhibited a higher frequency of antagonistic epistasis
307 and fewer instances of synergistic epistasis (**Fig. 4B**). Of the 65 gene pairs exhibiting significant
308 antagonistic epistasis in ES-LUAD, 25 also reached significance in NS-LUAD, and an additional
309 36 exhibited antagonistic trends without achieving significance (**Fig. 4C–D**). Thus, instances of
310 selective antagonism inferred in ES-LUAD were nearly always present in NS-LUAD.
311 This consistency contrasted with a substantial inconsistency in selective synergies
312 between ES- and NS-LUAD. Of the 42 gene pairs whose mutations exhibited significant
313 synergistic epistasis in ES-LUAD, only five were significantly synergistic in NS-LUAD. An
314 additional 25 trended synergistic without reaching significance. Strikingly, 16 of these 42
315 instances were inferred to be significantly antagonistic in NS-LUAD, including many involving
316 *BRCA2*. For example, *KEAP1* mutation enhanced selection for *BRCA2* mutation in ES-LUAD
317 but antagonized it in NS-LUAD (**Fig. 4C–D**). These reversals reveal distinct adaptive constraints
318 imposed by the smoking-altered tumor microenvironment. Overall, antagonistic interactions

319 were relatively stable across smoking status, whereas synergistic epistasis was highly
320 contingent on smoking status.
321 Mutations in these 21 driver genes were subject to a broad range of epistatic and
322 environmental interactions, as well as combinations thereof (**Fig. S2**). For example, *TP53*
323 mutations antagonized *KRAS* mutations in ES-LUAD but had no substantial effect on *KRAS*
324 mutations in NS-LUAD (**Fig. 4E**). These context-dependent epistatic effects substantially
325 contribute to the divergent trajectories of ES- and NS-LUAD.
326 Fewer examples of synergistic epistasis were observed in NS-LUAD compared to ES-
327 LUAD. Nevertheless, synergistic epistasis contributed to a median twelve-fold increase in
328 selection on secondary mutations. Notable examples include enhanced selection for *CTNNB1*,
329 *RB1*, and *RBM10* mutations after *EGFR* mutation, for *PIK3CA*, *BRCA2*, *SETD2*, and *ARID1A*
330 mutations after *ATM* mutation, and for *SETD2* mutation after *ARID1A* mutation (**Fig. 4B**). Even
331 in relatively less mutated NS-LUAD, cooperative genetic interactions dramatically reshape the
332 fitness landscape of evolving tumors.
333
334 *Pairwise epistasis can be asymmetric in both sign and magnitude*
335 In both ES- and NS-LUAD, several gene pairs exhibited asymmetric epistasis, in which the
336 direction or strength of interaction depended on the order of mutation acquisition. For instance,
337 in ES-LUAD, mutation of either *KEAP1* or *STK11* enhanced selection on mutations of the other,
338 but with markedly different magnitudes: primary *STK11* mutation was estimated to increase
339 selection for a secondary *KEAP1* mutation by 140%, whereas primary *KEAP1* mutation
340 significantly increased selection for a secondary mutation of *STK11* by 360%. In other cases,
341 the sign of epistasis reversed depending on mutational order. Most notably in ES-LUAD,
342 primary *KRAS* mutation synergistically increased selection for a secondary *KEAP1* mutation, but
343 primary *KEAP1* mutation strongly antagonized selection for a secondary *KRAS* mutation

344 (**Fig. 4A**). This asymmetry means that a mutation in *KRAS* followed by a mutation in *KEAP1*
345 was advantageous, but a mutation in *KRAS* was disadvantageous or lethal when preceded by a
346 mutation in *KEAP1*. This dramatic directional asymmetry and eighteen other such instances
347 exemplify how somatic evolutionary trajectories can be constrained or potentiated by the
348 specific sequence in which mutations occur.
349

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

351

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

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

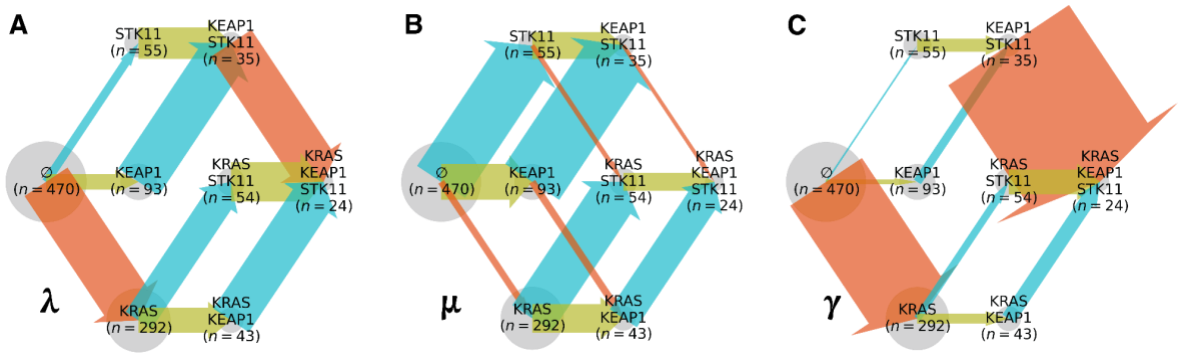


Figure 5. Evolutionary trajectory for *KRAS*, *KEAP1*, and *STK11* mutations in ever-smoker lung adenocarcinoma. (**A**) Fixation rates, (**B**) oncogenic mutation rates, and (**C**) scaled selection coefficients for mutations of *KRAS* (orange arrows; arrows width proportional to rate or coefficient),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

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

377 *STK11* mutations experienced significantly stronger selection in the presence of either
378 *KRAS* or *KEAP1* mutations. If the synergistic effects of *KRAS* and *KEAP1* mutations on *STK11*
379 compounded additively, then *STK11* mutations would experience selection equivalent to the
380 three-way product of the individual synergistic effects and the baseline selection for *STK11*
381 mutations. However, the upper bound for selection for *STK11* mutations in a *KRAS-KEAP1*
382 double-mutant genotype was less than this product (**Fig. 5C**). Accordingly, the compound
383 synergistic effect of concurrent *KEAP1* and *KRAS* mutations represents an example of sub-
384 additive higher-order epistasis.
385
386 *Most pairwise epistatic effects compound without strong higher-order interaction*
387 The sub-additive epistasis affecting mutations of *STK11* demonstrates that assessments
388 of higher-order epistasis will be necessary to thoroughly understand cancer evolution. However,
389 most effects of driver mutations compounded in a manner consistent with minimal higher-order
390 epistasis. Examining all 1330 triads out of the 21 genes, we compared selection for a new
391 mutation in single- versus double-mutant backgrounds. In the vast majority of cases, a
392 combination of pairwise epistatic effects was sufficient to explain significant differences in
393 selection for a new mutation between single- and double-mutant genotypes. Generally, when
394 selection for a secondary mutation in one gene was epistatically altered by presence of a
395 primary mutation in a second gene and was unaffected by a primary mutation in a third gene,
396 selection on the secondary mutation remained commensurately epistatically altered when both
397 other variants of the triad were concurrently present. For example, in ES-LUAD, selection for
398 *SMARCA4* mutations was synergistically enhanced by *STK11* mutations but unaffected by
399 *TP53* mutations, and correspondingly, selection for mutations in *SMARCA4* was synergistically
400 increased in a *STK11-TP53*-mutant genotype (**Fig. 6A,B**). Similarly, in NS-LUAD, selection for
401 *RBM10* mutations was synergistically increased in both a *TP53*-mutant and *TP53-EGFR*-mutant

402 genotype (**Fig. 6C**). Overall, pairwise interactions typically dominate the adaptive landscape of

403 LUAD, but higher-order epistasis can have substantial influence on specific evolutionary

404 trajectories.

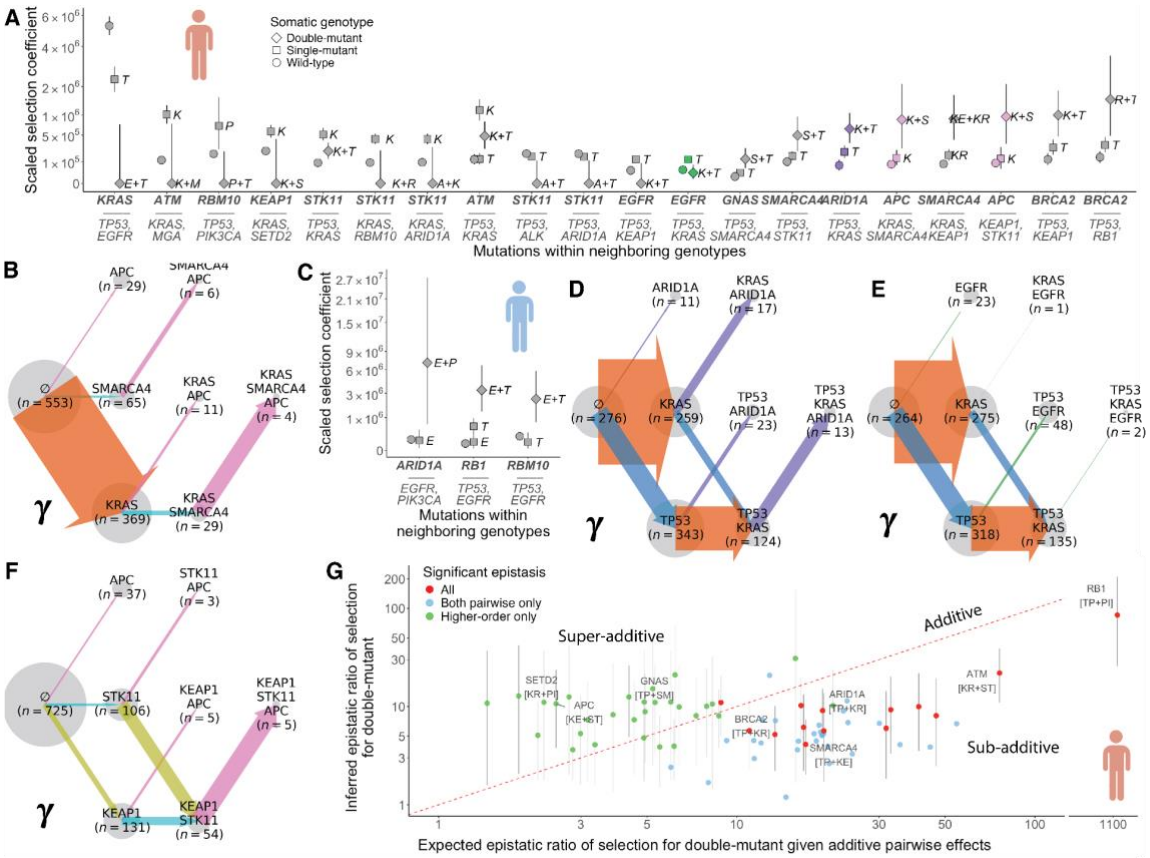


Figure 6. Higher-order selective epistasis in lung adenocarcinoma (LUAD). (**A**) Selection within gene triads (*y* axis is square-root scaled), for a mutation (bold font) in a context of no co-occurring mutations (circle), one co-occurring mutation (square; labeled with first letter(s) of the gene(s) indicated in the tick label that are mutated), and two co-occurring mutations (diamond) in ever-smoker (ES) LUAD [pink: visualized in panels B and F**;** purple: visualized in panel D; green: visualized in panel E]. (**B**) In ES-LUAD, the synergistic effect of *SMARCA4* mutations (light blue arrows) on selection for *APC* mutations (pink arrows; arrow width proportional to scaled selection coefficient; circle radius proportional to somatic genotype prevalence) does not diminish in the presence of a neutrally-interacting *KRAS* mutation (orange arrows). (**C**) Selection within gene triads in never-smoker (NS) LUAD; *y* axis is square-root scaled, symbols are as in panel A]. (**D**) Pairwise and higher-order epistasis in ES-LUAD: co-mutation of *TP53* (blue) and *KRAS* (orange), each of which independently increase selection for *ARID1A* mutations (purple), produces an even stronger synergy for *ARID1A* mutations; and (**E**) antagonistic effect of *KRAS* mutations (orange) on selection for *EGFR* mutations (green) outweighs the synergistic effect of *TP53* mutations (blue). (**F**) In ES-LUAD, co-mutation of *KEAP1* (light blue) and *STK11* (yellow) synergistically increases selection for *APC* mutations (pink) despite neutral pairwise effects. (**G**) Comparison of the epistatic effect of double-mutant-genotype (label: bottom row, two-letter abbreviation) on selection for new mutation (label: top row) to the "expected" product of pairwise epistatic effects (blue points: both pairwise epistatic effects are statistically significant, but not the higher order effect; green: the higher-order effect is significant, but neither pairwise effect is; red: all effects are significant)

When two co-occurring mutations each independently enhanced selection for a third mutation, the synergistic effects frequently compounded. For instance, *ARID1A* mutations experienced increased selection in both KRAS- and *TP53*-mutant genotype. These synergistic effects compounded, resulting in mutations of *ARID1A* experiencing three times stronger selection in a *KRAS*-*TP53*-mutant genotype than in a *TP53*-mutant genotype and nine times stronger selection than in a wild-type genotype (**Fig. 6D**). Similar patterns of compounded synergy were evident for *ATM* mutations in *KRAS*-*STK11*-mutant genotypes, as well as for *SMARCA4* mutations in a *KEAP1*-*STK11*-mutant background. Across cases, compounded synergy frequently drove selection to levels up to an order of magnitude higher than in genotypes lacking either partner mutation. This substantial compounding constitutes a crucial element of how driver mutations accumulate in cancer genomes, positively reinforcing progression of cancer toward more aggressive forms.

418

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

421 In the case of compounding synergistic effects, it is also possible to observe super-
422 additive higher-order epistasis, in which selection for a mutation in the presence of two
423 independently synergistic mutations is even larger than what would be expected from the
424 product of their independent synergistic effects. Indeed, some cases are consistent with either
425 additive synergy or super-additive epistasis, such as mutations of *SMARCA4* in a *KRAS-*
426 *KEAP1*-mutant genotype in ES-LUAD (**Fig. 6A**). More frequently, cases are consistent with
427 either additive synergy or sub-additive epistasis, in which selection for a mutation in the
428 presence of two independently synergistic mutations is smaller than expected from the sum of
429 synergistic effects. Indeed, in 43 of 45 cases wherein a new mutation arose in a genotype of two
430 significantly synergistic mutations, estimates of selection for the new mutation were lower than
431 would be expected from purely additive synergistic epistasis. Such widespread sub-additivity
432 pervades the adaptive landscape, tempering rather than amplifying selective pressures, even in
433 the context of strong pairwise synergy.

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

443 Another form of higher-order epistasis is the emergence of synergy or antagonism from a
444 combination of mutations, each of which does not individually exhibit pairwise epistasis with the
445 affected gene. We observed several instances in which selection for a mutation was significantly
446 increased by the co-occurrence of two mutations that did not individually have significant
447 synergistic effects: in ES-LUAD, selection for mutations of *APC* increased by an order of
448 magnitude in a *KEAP1*-*STK11*-mutant genotype despite not being significantly altered when
449 only one of *KEAP1* or *STK11* was mutated (**Fig. 6A**,**F**). In this instance, significance was
450 reached for synergy when compared to wild-type or a *KEAP1*-mutant genotype, but not when
451 compared to the *STK11*-mutant genotypes, which was itself epistatically neutral. We detected
452 multiple similar instances of emergent synergy, such as for *ARID1A* mutation in an *EGFR*-
453 *PIK3CA*-mutant genotype in NS-LUAD (**Fig. 6G**). We also observed one instance of emergent
454 antagonism, in which *STK11* mutations were antagonized in a *TP53*-*EGFR*-mutant genotype
455 but not in either single-mutant genotype alone.
456 Overall, we observed two major classes of higher-order epistasis: sub-additive synergy
457 and emergent—and consequently super-additive—synergy. In particular, when both component
458 epistatic effects were strongly synergistic, their combination was almost always sub-additively
459 synergistic (**Fig. 6G**). On the other hand, when both component epistatic effects were neutral or
460 insignificant, a super-additive synergistic effect occasionally emerged from their combination
461 (**Fig. 6G**). Interestingly, as the strength of the component epistatic effects increases, the inferred
462 epistatic effect of the double-mutant combination rapidly plateaus at an approximately ten to
463 twenty-fold increase in selection.
464 In this survey of the adaptive landscape of LUAD, we discovered several cases of higher-
465 order epistasis that substantially influence the selective advantage of oncogenic mutations.
466 Over a third of LUAD samples in ever-smokers possessed mutations in three or more of 21
467 tested genes—a proportion that would likely rise in an analysis of a greater number of driver

468 genes. In these patients, higher-order epistasis may play a substantial role in the evolutionary
469 trajectories of their tumors and their responses to targeted therapies. Indeed, we found that a
470 wide variety of adaptive trajectories ensued from the combined epistatic effects of multiple
471 mutations. Some of these trajectories cannot be predicted from pairwise analyses. However, our
472 large-scale analysis of the adaptive landscape of 21 mutations has significantly supported only
473 a subset of the many potential instances of higher-order epistasis between oncogenic mutations
474 in LUAD. Indeed, multiple other instances of higher-order epistasis were estimated to be
475 present. Their effect would be clarified with greater sample sizes.
476
477

478 Discussion

479
480 Here we have characterized the distinct adaptive landscapes of lung adenocarcinoma in
481 ever-smokers and never-smokers by first estimating gene-specific background mutation rates in
482 ES- and NS-LUAD, then comparing the intensity of somatic selection and selective epistasis
483 acting on those mutations between the two groups. Consistent with the established mutagenic
484 effects of tobacco carcinogens [4], we detected elevated oncogenic mutation rates in all tested
485 genes in ES- compared to NS-LUAD. Accounting for these background rates revealed that
486 mutations in genes such as *EGFR*, *PIK3CA*, and *MET* provide a stronger selective advantage in
487 NS-LUAD, while mutations in *KRAS*, *KEAP1*, and *STK11* provide a stronger selective
488 advantage in ES-LUAD. Focusing on *EGFR* and *KEAP1*, we found that these adaptive
489 differences were at least partially explained by the differential effects of these mutations on
490 epithelial-mesenchymal transition and DNA replication transcriptional program activity in ES-
491 and NS-LUAD.
492 By quantifying the selective advantage of these mutations within specific somatic
493 genotypes, our analysis recovered known epistatic interactions—such as the enhancement of

494 selection on *RB1* and *EGFR* mutations by prior *TP53* mutations—and revealed context-specific
495 interactions, such as the synergistic effect of *EGFR* mutations on *CTNNB1* mutations uniquely
496 in NS-LUAD. We also detected substantial asymmetries: *CTNNB1* mutations did not impact
497 *EGFR* mutations in ES- or NS-LUAD. Notably, we found more frequent synergy in ES-LUAD
498 and substantially more pervasive antagonism in NS-LUAD, constituting fundamental differences
499 in the evolutionary navigability of their adaptive landscapes. Finally, we showed that higher-
500 order epistasis—particularly sub-additive synergistic and emergent synergistic epistasis—is
501 frequent in and substantially guides the evolutionary trajectories of cancer. These findings of
502 differential selection and epistasis reveal that exposure to tobacco smoke reshapes the adaptive
503 landscape of LUAD not only by elevating mutation burden, but also by altering the selective
504 value of mutations and their epistatic interactions—highlighting opportunities to leverage these
505 adaptive differences for biomarker development and therapeutic stratification.
506 The prevalences of driver mutations have been reported to substantially differ between
507 smokers and never-smokers in LUAD [21,22] as well as in non-small cell lung cancer [NSCLC;
508 23,24]. However, previous studies have either refrained from explaining these differences or
509 attributed them primarily to tobacco-induced mutagenesis [22,23]. Our results show that they
510 often cannot be explained by differential mutagenesis alone. For example, *KRAS* variants are
511 over five times more frequent in ES- than NS-LUAD tumors, but the background, neutral rate at
512 which *KRAS* mutations occur is only about three times higher in ES-LUAD. To resolve this
513 discrepancy, *KRAS* mutations must be nearly twice as oncogenic in ES- compared to NS-
514 LUAD. In general, the differential prevalence of these driver mutations is often not solely
515 attributable to tobacco-induced mutagenesis, but also due to their altered selective advantage in
516 the tobacco-altered lung environment. This physiological-adaptive effect of smoking is crucial to
517 account for: changes in the selective advantage of a mutation suggest altered functional impact

518 within a given tissue environment, with direct implications for the efficacy of targeted therapies
519 and the interpretation of biomarkers in smokers versus non-smokers.

520 Multiple perspectives have proposed that the adaptive landscape of oncogenesis is
521 altered by physiological insults from endogenous and exogenous factors [14,17,52,53].
522 However, data-driven evidence for this hypothesis has remained limited [cf. 54]. Here, we
523 provide compelling evidence that canonical driver mutations confer altered selective advantages
524 in the smoking-affected lung microenvironment of ever-smokers with LUAD. These shifts in
525 selection are linked to context-dependent phenotypic effects of mutations. For example, *EGFR*
526 mutations are associated with greater activity of the epithelial-mesenchymal transition (EMT)
527 transcriptional program in NS-LUAD, but not in ES-LUAD—likely explaining their stronger
528 selective advantage in never-smokers. Such divergent phenotypic consequences can be
529 attributed to smoking-induced alterations to transcriptional program activity [55–57], tumor
530 microenvironment composition [58], and cell fitness. These findings provide a mechanistic basis
531 for how exogenous exposures can reshape the evolutionary trajectory of cancer by altering the
532 adaptive value of oncogenic mutations.

533 In the case of *TP53*, Rodin and Rodin [13] reported comparable background mutation
534 rates between smokers and nonsmokers with lung cancer, and attributed the higher frequency
535 of *TP53* mutations in smokers to increased selection in the smoking-altered microenvironment.
536 By contrast, our analysis—leveraging a substantially larger dataset, incorporating tissue- and
537 context-specific covariates, and applying advanced statistical methodologies to estimate
538 background mutation rates [6]—identified a markedly elevated *TP53* mutation rate in ES-LUAD.
539 With this substantial differential mutagenesis accounted for, our analysis revealed that *TP53*
540 mutations are actually under weaker positive selection in ever-smokers than in never-smokers.
541 This finding challenges earlier conclusions and underscores the importance of accurately

542 resolving both mutational processes and selection when interpreting variant prevalence across
543 environmental exposures.
544 Despite mounting evidence of the substantial biological impact of tobacco smoke, clinical
545 trials for lung cancer therapies often do not assess the influence of smoking status—and
546 frequently do not even collect usage history [59], perhaps due to a conception that tobacco
547 smoke has a negligible effect on outcomes [60]. However, our findings reveal that tobacco
548 smoke profoundly reshapes the evolutionary trajectory of lung adenocarcinoma, necessitating
549 more systematic incorporation of smoking history into trial design and treatment decision-
550 making, particularly for targeted therapies. For example, the substantially stronger selection for
551 *EGFR* mutations in NS-LUAD implies that EGFR tyrosine kinase inhibitors (TKIs) should yield
552 greater clinical benefit in never-smokers. Indeed, a retrospective analysis of patients with
553 *EGFR*-mutant LUAD found that the objective response rate (ORR) of patients receiving EGFR
554 TKIs declined from 73% among never-smokers to 46% among individuals with over 10 pack-
555 years of smoking history [61]. Similar patterns have been reported in additional retrospective
556 studies of LUAD [62] and across NSCLC more broadly [63,64]. This concordance between
557 selection-based predictions and observed outcomes highlights the translational value of
558 selection coefficients as predictors of targeted therapy efficacy. By identifying likely responders
559 based on environmental exposures that shape the tumor microenvironment, exposure-informed
560 selection coefficients can improve clinical trial stratification, enhance statistical power, and refine
561 treatment personalization in precision oncology.
562 Prior studies have questioned whether computational analyses can reliably detect
563 epistasis, citing concerns about confounding by subtype-specific mutation profiles, variation in
564 mutation burden, and limited statistical power [31,65]. To address these concerns, we analyzed
565 ES- and NS-LUAD independently and comparatively, modeled subtype- and gene-specific
566 background mutation rates and effect sizes of mutations in specific somatic genotypes instead

567 of suboptimal measures of co-occurrence or mutual exclusivity, and integrated data from
568 multiple cohorts to ensure sufficient statistical power. Consequently, we were able to make
569 robust inferences of both pairwise and higher-order selective epistasis, many of which align with
570 experimental and clinical evidence. For example, consistent with our finding that *KRAS*
571 mutations synergistically enhance selection for *KEAP1* and *STK11* mutations, prior studies
572 show that *KEAP1* loss accelerates LUAD progression in *KRAS*-mutant mouse tumors [66,67]
573 and compensates for the deleterious effects of *KRAS*-*STK11*-co-mutation [68], and mutations of
574 *KEAP1* and/or *STK11* worsen survival and treatment response in *KRAS*-mutant LUAD [69,70]
575 and NSCLC [71–76]. Furthermore, in accordance with the inferred mutually synergistic effects of
576 mutations of *KEAP1* and *STK11* in ES-LUAD, co-mutation of *KEAP1* and *STK11* promote cell
577 proliferation [77] and worsen survival in NSCLC independently of *KRAS* [78,79].
578 Our analyses also identified synergies between *EGFR* and other drivers: *RBM10* and
579 *CTNNB1* in NS-LUAD, and *RB1* in both NS- and ES-LUAD. These interactions are corroborated
580 by experimental and clinical studies showing that *RBM10* mutations promote *EGFR*-mutant
581 LUAD [80] and confer resistance to EGFR TKIs [81], that *CTNNB1* mutations are enriched in
582 advanced *EGFR*-mutant NSCLC and associated with metastasis and TKI resistance [82], and
583 that *RB1* mutations in *EGFR*-mutant LUAD mediate histologic transformation to small-cell lung
584 cancer [83,84]. We also revealed that mutations of *TP53* and *KRAS* were widely antagonized by
585 other driver mutations, consistent with their early, clonal emergence during LUAD evolution
586 [85,86] and supporting their frequently primary role in tumor initiation. This panoply of
587 alignments between our results and experimental and clinical evidence encourages deeper
588 investigation into the many epistatic relationships we have revealed that remain untested in the
589 laboratory or clinic. They represent fertile ground for mechanistic investigation and promise new
590 therapeutic targets and biomarker candidates.

591 Previous studies have shown that epistatic interaction networks vary across cancer types
592 in humans [87] and across environmental contexts in other organisms [88–93]. However,
593 extensive differences in selective epistasis within a single cancer type exposed to distinct tissue
594 environments have not previously been demonstrated. Here, we have shown that tobacco
595 smoke exposure markedly reshapes the selective epistatic interactions underlying LUAD
596 evolution. Broadly, synergistic epistasis was more frequent in ES-LUAD, whereas antagonistic
597 epistasis was more frequent in NS-LUAD. Indeed, in some cases, epistasis between driver
598 genes was synergistic in ES-LUAD, but antagonistic in NS-LUAD. In other cases, the magnitude
599 of synergy differed across exposures: synergy for mutation of *TP53* in *EGFR*-mutant tumors
600 was stronger in ES- than in NS-LUAD. The higher frequency of synergy and lower frequency of
601 antagonism in ES-LUAD suggests a more navigable adaptive landscape—characterized by a
602 higher frequency of beneficial mutational combinations and fewer evolutionary barriers—than in
603 NS-LUAD. This increased accessibility enables ES-LUADs to develop along a greater diversity
604 of evolutionary trajectories, contributing to increased interpatient heterogeneity. Moreover, it
605 may also help to explain why tobacco smoke elevates the risk of lung cancer to such an
606 exceptional degree. In NS-LUAD, widespread antagonistic epistasis may reflect a greater
607 fitness cost of mutation accumulation, possibly due to a more immunologically reactive tumor
608 microenvironment [12,48,58,94] or due to a lower tolerance for mutation accumulation in the
609 non-dysregulated lung tissue cells of never-smokers. Taken together, these findings reveal that
610 environmental exposures modulate selective epistasis within a cancer type, suggesting the
611 importance of exposure history to predicting tumor evolutionary trajectories and to guiding
612 epistasis-informed treatment strategies.
613 It has often been assumed that the adaptive landscape of cancer can be adequately
614 described without epistasis [6,95–99] or with only pairwise epistasis [31,32,100]. However, our
615 results demonstrate that not only pairwise but also higher-order epistasis—particularly sub-

616 additive and emergent epistasis—play a crucial role in modulating the adaptive value of cancer
617 drivers. For instance, in ES-LUAD, selection for *APC* mutations increased by an order of
618 magnitude in an *KEAP1*-*STK11*-mutant genotype, despite showing no significant pairwise
619 synergy with either mutation alone. This emergent epistasis, arising only in the triadic context,
620 would be missed by analyses restricted to pairwise comparisons, necessitating consideration of
621 multi-mutant genotypes in both computational and experimental studies. Moreover, almost all
622 significant pairwise synergistic effects combined sub-additively in triads, rapidly reaching a
623 plateau of synergistic benefit. This pattern aligns with the diminishing-returns epistasis observed
624 in asexual populations [44–47]. Overall, higher-order synergistic epistasis biases tumors toward
625 some evolutionary trajectories, while higher-order antagonism prunes most trajectories. Both
626 make tumor evolution more predictable. Thus, incorporation of higher-order epistasis into
627 models of tumorigenesis is essential to understanding and guiding the evolutionary paths that
628 tumors take under therapy.

629 Current NSCLC treatment guidelines for EGFR TKIs do not account for co-occurring
630 mutations. However, our findings suggest that concurrent mutations could serve as biomarkers
631 of treatment response. Antagonistic epistasis between *EGFR* mutations and mutations in *KRAS*
632 or *KEAP1* implies that the efficacy of EGFR TKIs may be reduced in tumors harboring these
633 concurrent alterations. This prediction aligns with clinical observations: *KRAS* mutations
634 decrease survival in patients with *EGFR*-mutant LUAD patients treated with EGFR TKIs [101],
635 and *KEAP1* mutations reduced EGFR TKI efficacy in *EGFR*-mutant lung cancer [102,103]. More
636 broadly, antagonistic epistasis is a signal of synthetic lethality [32]. Our results also demonstrate
637 mutual negative selection between *BRCA2* and *SMAD4* loss-of-function mutations—suggesting
638 that pharmacological inhibition of *SMAD4* could be therapeutically beneficial in *BRCA2*-mutant
639 tumors, in analogy to usage of PARP inhibitors in *BRCA1*/*2* deficient tumors. In general,
640 selective epistasis can clarify the frequent heterogeneity of response to targeted therapies

641 [37,104], and can improve stratification of responders and nonresponders to targeted therapy in
642 clinical trials [105,106], enhancing trial power and therapeutic precision.

643 Because *TP53* mutations synergize with mutations of *EGFR*, we would expect that in
644 tumors with concurrent *EGFR* and *TP53* mutations, EGFR TKIs would be especially effective
645 because they would both eliminate the baseline effect of the *EGFR* mutation and diminish the
646 effect of the concurrent *TP53* mutation that is mediated through the *EGFR* mutation. However,
647 *TP53* mutations are consistently associated with resistance to EGFR TKIs in *EGFR*-mutant
648 LUAD [107–109] and NSCLC [110–114]. This contradiction may be explained by the pleiotropic
649 effects of *TP53* loss, including increased somatic nucleotide mutation rates [109], widespread
650 copy-number alterations [107,115], and dysregulation of gene expression programs linked to
651 resistance [116] and progression [115]. These pleiotropic effects provide greater genetic
652 variation that facilitates the evolution of resistance independently of the selective synergy
653 between *TP53* loss and *EGFR* mutation.

654 Previous computational studies of pairwise interactions have demonstrated asymmetries
655 in the magnitude of pairwise epistasis [cf. 30]. Our analysis has revealed not only epistasis of
656 asymmetric magnitudes, but also asymmetric signs of selection, in which one ordering of the
657 mutations is advantageous while the reverse is disadvantageous. These directional
658 asymmetries likely arise from cellular or stable transcriptional transitions initiated by the primary
659 mutation that unidirectionally alter the selective context for subsequent alterations—an effect not
660 captured by conventional analyses of mutual exclusivity or co-occurrence. Indeed, the presence
661 of such asymmetries of epistasis are evident in established genetic models of tumor evolution,
662 such as the *APC* → *KRAS* → *TP53* mutation sequence in colorectal cancer [117,118], whose
663 strongly preferential ordering requires directional epistasis. Asymmetric epistasis has also been
664 supported by genetic and computational genomic studies of *TP53* and *CCNE1* mutations [32],
665 *TP53* and *RB1* mutations [30], and *JAK2* and *TET2* mutations [119]. The presence of directional

666 epistasis implies preferential orders of mutation acquisition that offer new opportunities to
667 predict and potentially redirect cancer evolution.

668 This study was subject to several limitations. First, its retrospective nature introduces the
669 possibility that covariates of smoking behavior—such as comorbid health states, environmental
670 exposures, inherited genetics, and sex biases—may contribute to the observed differences in
671 the adaptive landscapes of ES-LUAD and NS-LUAD. However, none of these factors are known
672 to exert a physiological impact on the lung that is comparable to smoking tobacco. Accordingly,
673 the selective differences we report are most parsimoniously attributed to the direct and
674 downstream effects of smoking. Second, smoking status in our study was primarily inferred from
675 mutational signatures in tumor sequences, which can be susceptible to misclassification.
676 However, this method has been validated for classification of smoking status in NSCLC and is
677 believed to outperform self-reported smoking histories [21,25,120]; self-reported status has
678 been found to be unreliable [121,122] and often does not account for the risk conferred by
679 secondhand smoke [123]. Furthermore, because our approach explicitly quantifies the smoking-
680 associated mutational burden in the lung periphery—where LUAD develops—it also implicitly
681 identifies patients who have experienced significant injury from smoking in the regions where
682 LUAD originates. Finally, our analysis was restricted to somatic SNVs and did not consider
683 indels, copy number alterations (CNAs), or other structural variants (SVs). CNAs are common in
684 cancer [124] and pairwise selective epistasis has been indicated between SNVs and CNAs and
685 other SVs [125,126] or between SVs [37,87,127]. Unfortunately, accurate background mutation
686 rates are currently unavailable for SVs. As such rates become calculable, integrated analysis of
687 SNVs and SVs will reveal new epistatic relationships and clarify the somatic adaptive landscape
688 of cancer.

689 Overall, we have demonstrated that the adaptive landscapes of ES- and NS-LUAD diverge
690 substantially, implicating tobacco smoking not only as a mutagen but also as a modulator of

691 somatic selection. Our analysis also shows that symmetric and asymmetric selective epistasis
692 among driver mutations is both substantial and pervasive in lung adenocarcinoma and
693 generates the distinct evolutionary trajectories of ES-LUAD and NS-LUAD. Furthermore, we
694 provide a systematic characterization of higher-order selective epistasis in LUAD, revealing its
695 underrecognized role in shaping tumor evolution. Together, these results show that the
696 contribution of mutations to tumor growth is profoundly context-dependent—modulated by both
697 tumor somatic genetics and the environmental exposures it has experienced. This insight
698 illuminates new opportunities for precision oncology through the integration of somatic genotype
699 and exposure history into treatment strategies. Continued investigation into epistatic,
700 environmental, and even gene-by-gene-by-environment interactions in lung adenocarcinoma
701 and other cancers will be essential for the advancement of our understanding of cancer
702 evolution and the improvement of patient-specific therapeutic strategies.
703

704 # Methods

705 ## Design overview

706 The aim of this study was to test for differences in background mutation rates, subsequent
707 somatic selection, and gene-by-gene epistatic effects between ES- and NS-LUAD. To do so, we
708 aggregated mutation data of 2,097 LUAD tumor samples from multiple institutional sources. We
709 classified tumors as either ES- or NS-LUAD either by detection of the smoking-associated
710 mutational signature SBS4 in exome or genome sequences or, in the case of targeted panel
711 sequence data, by relying on the patient's self-reported smoking history. We applied established
712 approaches to estimate background mutation rates, scaled selection coefficients, and pairwise
713 and higher-order selective epistasis within the ES- and NS-LUAD cohorts. To assess the
714 differential transcriptional effects of driver mutations, we compared ES- and NS-LUAD tumors

715 with and without specific mutations. Differences in selection were conservatively evaluated
716 based on non-overlapping 95% confidence intervals.
717

718 **Data sources**

719 We sourced clinicogenomic data from (i) eight published datasets hosted by cBioPortal
720 [85,128–135], (ii) The Cancer Genome Atlas dataset hosted on Genomic Data Commons [136],
721 and (iii) 108 samples sequenced at Yale University [137]. All datasets reported sample sizes,
722 sequencing methods, and clinical covariates, and were incorporated into our analysis
723 accordingly (**Supplementary Table 1**). We only included samples from primary LUAD tumors,
724 and we did not include any datasets from studies that preselected patients based on their
725 mutational profiles. When multiple samples from a single patient were available, one was
726 selected at random. In cases where patients appeared in more than one dataset, we retained
727 the sample with the highest-quality sequencing and most complete clinical annotations.
728 For consistency across datasets, we standardized all variant genomic coordinates to the
729 hg19 reference genome using the liftOver function of rtracklayer v1.62.0 [138]. We used
730 cancereffectsizeR v2.10.1 to remove variants that failed liftover, were incorrectly classified as
731 single-nucleotide variants (SNVs), or did not meet quality criteria for downstream analysis. For
732 panel-sequenced samples, we inferred the genomic regions covered by each panel from the
733 gene list provided for each panel using the corresponding hg19 coordinates.
734

735 **Mutational signature attributions and smoking-status classification**

736 We performed mutational signature deconvolution for tumor samples with whole-exome
737 or whole-genome sequence data, but not for panel-sequenced samples, due to the low reliability
738 of signature attribution when few silent somatic variants are present. We conducted signature
739 attribution using cancereffectsizeR with MutationalPatterns v3.12.0 [139], applying COSMIC

740 v3.2 single-base substitution signatures [26,140]. We excluded COSMIC signatures that were
741 unlikely to have been observed in lung adenocarcinomas [140], and we excluded treatment-
742 associated signatures when deconvolving samples known to be treatment-naive. Samples in
743 which the smoking-associated SBS4 signature contributed more than the MutationalPatterns
744 default threshold of 5% of the total signature burden were classified as ever-smoker ES-LUAD,
745 and otherwise were classified as never-smoker NS-LUAD. Zhang et al. detected no signature of
746 SBS4 in the samples used in their study; therefore we directly classified these samples as NS-
747 LUAD. As in Rosenthal et al. [141], we excluded samples with ≤50 single-nucleotide variants in
748 whole-exome data from classification due to insufficient mutation burden for reliable signature
749 deconvolution. We compared signature-based smoking assignments to available clinical
750 annotation of smoking status across datasets, validating that there was 67–90% concordance.
751 Discrepancies were likely due to underreporting of smoking, second-hand smoke exposure,
752 minimal mutational consequences in light smokers, or misclassification in clinical records. The
753 selection and epistasis results produced from either classification method were 99.4%
754 consistent with each other. We classified panel-sequenced samples according to clinical
755 annotations of smoking status when available; we excluded samples annotated as exposed to
756 “former light” smoking from all smoking-stratified analyses due to ambiguity in exposure and
757 effect.
758

759 **Estimation of oncogenic mutation rates**

760 We estimated gene-specific oncogenic mutation rates separately for the ES- and NS-
761 LUAD cohorts following the approach of Alfaro-Murillo and Townsend [28]. The oncogenic
762 mutation rate for a gene was defined as the neutral rate at which oncogenic variants occur
763 within the gene. We defined oncogenic variants as variants that were observed at least once in

764 any tumor sample in the dataset. The background oncogenic mutation rate of each gene was
765 computed as the sum of the background rates of all oncogenic variants within the gene. For a
766 specific variant derived from a trinucleotide mutation type m (*e.g.* CCT → CAT) at a site i within
767 gene g in a tumor sample cohort c, the somatic background mutation rate

768 $$\mu_{i,g,c}^{m} = \mu_{g,c} \cdot \mathrm{Prob}\left\{T_m|c\right\} \cdot \mathrm{Prob}\left\{S_{i,m_{\mathrm{WT}},g}\right\},$$

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

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

779 $$\mu_{i,g,c}^{m} = \frac{\mu_{g,c} \cdot n_m^c}{S_{m_{\mathrm{WT}}}^{g} \cdot N_c}.$$

780 The background oncogenic mutation rate of gene g is the sum of the background rates of all
781 oncogenic variants within g:

782 $$\hat{\mu}_{g,c} = \sum_{(i,m)\in\mathcal{S}_g} \mu_{i,g,c}^{m}$$

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

788 Finally, the oncogenic mutation rates per somatic genotypes were computed by rescaling
789 the per-gene background rates using the tumor mutation burden (TMB) observed in each

790 genotype. Specifically, for a given cohort *c*, the mutation rate for a jump between the somatic
791 genotype **x** the somatic genotype **y** is given by

792 $$\mu_{\mathbf{x}\to\mathbf{y},c} = \frac{\mathrm{TMB}_{\{\mathbf{x},\mathbf{y}\}}}{\mathrm{TMB}_{\mathrm{all}}}\hat{\mu}_{g,c},$$

793 where $\mathrm{TMB}_{\{\mathbf{x},\mathbf{y}\},c}$ is the average TMB in tumors with the somatic genotypes **x** or **y**, $\mathrm{TMB}_{\mathrm{all}}$ is
794 average TMB for all tumors, and *g* is the gene that is mutated in the somatic genotype **y** but not
795 in **x** (the mutation gained by the jump). In cases where one genotype has no samples, the
796 average is taken only from the observed partner. If both endpoints have zero samples, the rate
797 defaults to the baseline per-gene rate $\mu_{\mathbf{x}\to\mathbf{y},c} = \hat{\mu}_{g,c}$, avoiding spurious rescaling in unsupported
798 genotype classes. We considered TMB only from our largest dataset (TCGA) because after
799 comparing the datasets with the Kolmogorov–Smirnov test, all of them proved to come from
800 different distributions.
801

802 **Estimation of scaled selection coefficients and selective epistasis**

803 To quantify the growth and survival advantages conferred by somatic SNVs in LUAD, we
804 calculated scaled selection coefficients for recurrent mutations in a curated set of 1,200 genes
805 [136,142–145]. These coefficients roughly equate to the ratio of the fixation rate for the
806 oncogenic mutant (the rate at which the mutant is generated and subsequently selected to a
807 high enough frequency to be observed as a somatic variant within the cancer cell population) to
808 its background mutation rate. Precisely, we computed the fixation rate by analytical
809 maximization of the binomial likelihood of observing at least one non-silent SNV in the gene
810 given the number of samples with and without a non-silent SNV in the gene. Scaled selection
811 coefficients for each mutated gene were estimated separately for ES- and NS-LUAD, as well as
812 for a combined cohort. Within each group, we included samples with whole-exome, whole-
813 genome, or targeted gene panel sequences (TGS). Results were consistent when excluding

814 TGS samples, indicating robustness to sequencing strategy. Inclusion of TGS samples provided
815 greater statistical power to detect significant differences in selection between cohorts. Results
816 were also robust to cohort-specific normalization: applying median-scaling of coefficients prior to
817 between-cohort comparison yielded consistent outcomes.

818 From the curated set of 1,200 genes, we selected 21 for detailed analysis of selective
819 epistasis based on strong prior evidence as LUAD drivers [136,143] and their elevated average
820 strength of selection across somatic genotypes. We quantified epistasis by calculating scaled
821 selection coefficients conditional on the somatic genotype as in Alfaro-Murillo and Townsend
822 [28]: for each pair or triplet of genes, we calculated the fixation rate of mutations in one gene
823 conditional on the presence or absence of somatic variants in the others. We performed this
824 calculation by maximizing the multinomial likelihood for the number of samples with each
825 somatic genotype using PyMC [146]. The probability of a tumor being in a specific somatic
826 genotype during the somatic evolutionary trajectory from organogenesis to biopsy was
827 computed via a continuous-time Markov chain model [28]. We computed asymmetric 95%
828 confidence intervals using Wilks' theorem [147]. To obtain scaled selection coefficients for each
829 mutation occurring within each somatic genotype, we divided the respective Poisson-corrected
830 fixation rate [148] by the background oncogenic mutation rate of the gene.

831 To systematically assess selective epistasis in both the ever-smoker and the never-
832 smoker groups, we evaluated 3,080 models representing all possible two- and three-gene
833 combinations from the 21 LUAD driver genes. For each model, somatic genotypes were defined
834 exclusively by the variant status (mutant or wildtype) of the genes included in each combination.
835 To ensure complete genotype classification, we only included panel-sequenced samples in a
836 given model if the sequencing panel covered all of the genes under analysis.
837

838 **Classification of synergistic and antagonistic selective epistasis**

839 To classify a mutant somatic genotype as interacting synergistically or antagonistically with
840 a certain mutation, we compared the scaled selection coefficients for a mutation when it
841 occurred in the mutant somatic genotype to the scaled selection coefficients obtained for the
842 same mutation in a somatic genotype lacking mutations in the genes included in the model. If
843 the scaled selection coefficient for a mutation in gene *A* when it occurred in a wild-type
844 genotype ($\varnothing \rightarrow A$) was less than the scaled selection coefficient for a mutation in gene *A* when it
845 occurred in a somatic genotype with a mutation in gene *B* ($B \rightarrow A + B$), then we deemed the
846 epistatic effect of mutant gene *B* on mutant gene *A* to be synergistic. If the scaled selection
847 coefficient for $\varnothing \rightarrow A$ was greater than the scaled selection coefficient for $B \rightarrow A + B$, then we
848 deemed the epistatic effect of mutant gene *B* on mutant gene *A* to be antagonistic. We further
849 distinguished between two scales of antagonism: antagonistic sign epistasis when there was a
850 reversal of the direction of selection—from positive (coefficient $\gamma > 1$) to negative selection
851 (coefficient $\gamma < 1$)—in the antagonistic genotype, and antagonistic magnitude epistasis when
852 there was a reduction in strength of selection but selection remained positive in the antagonistic
853 genotype.
854

855 **Differential expression analysis**

856 To evaluate the transcriptional impact of key driver mutations across smoking-stratified
857 LUAD tumors, we performed differential expression analysis using RNA-seq data from the
858 TCGA-LUAD project. We obtained unstranded raw read counts for 59 normal lung and 539
859 primary LUAD tumor samples using TCGAbiolinks v2.30.0. We excluded formalin-fixed paraffin-
860 embedded samples and tumors which received neoadjuvant treatment to minimize technical

861 and biological confounders. We retained solid-tissue normal samples as controls. We excluded
862 genes with fewer than 10 total counts across all samples. For each sample, we assigned
863 smoking status based on SBS4 signature weight from the matched MAF. The genes *EGFR*,
864 *KRAS*, and *KEAP1* were considered mutated if at least one non-silent SNV, insertion, or
865 deletion was present in the gene (excluding intronic regions). After this filtering and annotation,
866 we validated that a principal-component analysis revealed no evidence of batch effects,
867 confirming the suitability of the dataset for differential expression analysis.

868 We performed differential expression analysis with DESeq2 v1.42.1 [149] using two
869 designs. Differential transcriptional effects of *EGFR* mutations in ES- and NS-LUAD were
870 quantified as

871 $$K_i = \beta_0 + \beta_1 x_t + \beta_2 x_s + \beta_3 x_e + \beta_4 x_r + \beta_5 (x_s * x_e) + \beta_6 (x_s * x_r),$$

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

878 $$K_i = \beta_0 + \beta_1 x_t + \beta_2 x_s + \beta_3 x_k + \beta_4 (x_s * x_k),$$

879 where x_k is the *KEAP1* mutation status.

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

884

885 **Statistical Analysis**

886 Differences in selection coefficients were assessed conservatively by non-overlapping
887 95% confidence intervals for each estimate. Differential gene expression was analyzed using
888 two-tailed Wald tests, and statistical significance was determined based on Benjamini-
889 Hochberg–adjusted $P < 0.05$. Gene set enrichment analyses were likewise considered
890 significant at a Benjamini-Hochberg–adjusted $P < 0.05$. Analysis was performed in Python
891 (3.9.5) and R (4.3.0).

892
893

894 ## Data and code availability

895 All code, data, and instructions necessary to reproduce the results and figures in the current
896 study is available in a public Zenodo repository, https://zenodo.org/records/16380037 [150] with
897 no access restrictions.
898
899

900 **Acknowledgements**

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

903 **Authors' contributions**

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

908 **Competing interests**

909 The authors declare that they have no competing interests.

Materials and correspondence

Correspondence and material requests should be addressed to Jeffrey Townsend.

Funding

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

References

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

2. Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, et al. Mutational signatures associated with tobacco smoking in human cancer. Science. 2016;354: 618–622.

3. Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, Hainaut P. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. Oncogene. 2002;21: 7435–7451.

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

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

6. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal patterns of selection in cancer and somatic tissues. Cell. 2017;171: 1029–1041.e21.

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

8. Tan C, Mandell JD, Dasari K, Cannataro VL, Alfaro-Murillo JA, Townsend JP. Heavy mutagenesis by tobacco leads to lung adenocarcinoma tumors with KRAS G12 mutations other than G12D, leading KRAS G12D tumors-on average-to exhibit a lower mutation burden. Lung Cancer. 2022;166: 265–269.

9. Mochizuki A, Shiraishi K, Honda T, Higashiyama RI, Sunami K, Matsuda M, et al. Passive smoking-induced Mutagenesis as a promoter of lung carcinogenesis. J Thorac Oncol. 2024;19: 984–994.

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

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

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

13. Rodin SN, Rodin AS. Human lung cancer and p53: the interplay between mutagenesis and selection. Proc Natl Acad Sci U S A. 2000;97: 12244–12249.

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

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

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

17. Stratton MR, Humphreys L, Alexandrov LB, Balmain A, Brennan P, Campbell PJ, et al. Implementing mutational epidemiology on a global scale: Lessons from mutographs. Cancer Discov. 2025;15: 22–27.

18. Okamoto T, Suzuki Y, Fujishita T, Kitahara H, Shimamatsu S, Kohno M, et al. The prognostic impact of the amount of tobacco smoking in non-small cell lung cancer--differences between adenocarcinoma and squamous cell carcinoma. Lung Cancer. 2014;85: 125–130.

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

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

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

22. Moorthi S, Paguirigan A, Itagi P, Ko M, Pettinger M, Hoge AC, et al. The genomic landscape of lung cancer in never-smokers from the Women's Health Initiative. JCI Insight. 2024;9. doi:10.1172/jci.insight.174643

23. Wang X, Ricciuti B, Nguyen T, Li X, Rabin MS, Awad MM, et al. Association between smoking history and tumor mutation burden in advanced non-small cell lung cancer. Cancer Res. 2021;81: 2566–2573.

24. Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell. 2012;150: 1121–1134.

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

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

27. Mandell JD, Cannataro VL, Townsend JP. Estimation of Neutral Mutation Rates and Quantification of Somatic Variant Selection Using cancereffectsizeR. Cancer Res. 2023;83: 500–505.

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

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

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

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

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

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

33. Zhang Z, Yang Y, Zhou Y, Fang H, Yuan M, Sasser K, et al. A forward selection algorithm to identify mutually exclusive alterations in cancer studies. J Hum Genet. 2021;66: 509–518.

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

35. Liu S, Liu J, Xie Y, Zhai T, Hinderer EW, Stromberg AJ, et al. MEScan: a powerful statistical framework for genome-scale mutual exclusivity analysis of cancer mutations. Bioinformatics. 2021;37: 1189–1197.

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

37. Mina M, Raynaud F, Tavernari D, Battistello E, Sungalee S, Saghafinia S, et al. Conditional Selection of Genomic Alterations Dictates Cancer Evolution and Oncogenic Dependencies. Cancer Cell. 2017;32: 155–168.e6.

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

39. Weinreich DM, Lan Y, Wylie CS, Heckendorn RB. Should evolutionary geneticists worry about higher-order epistasis? Curr Opin Genet Dev. 2013;23: 700–707.

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

41. Weinreich DM, Lan Y, Jaffe J, Heckendorn RB. The influence of higher-order epistasis on biological fitness landscape topography. J Stat Phys. 2018;172: 208–225.

42. Domingo J, Diss G, Lehner B. Pairwise and higher-order genetic interactions during the evolution of a tRNA. Nature. 2018;558: 117–121.

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

44. Wei X, Zhang J. Patterns and Mechanisms of Diminishing Returns from Beneficial Mutations. Mol Biol Evol. 2019;36: 1008–1021.

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

46. Chou H-H, Chiu H-C, Delaney NF, Segrè D, Marx CJ. Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science. 2011;332: 1190–1192.

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

48. Sui Q, Liang J, Hu Z, Chen Z, Bi G, Huang Y, et al. Genetic and microenvironmental differences in non-smoking lung adenocarcinoma patients compared with smoking patients. Transl Lung Cancer Res. 2020;9: 1407–1421.

49. Frank R, Scheffler M, Merkelbach-Bruse S, Ihle MA, Kron A, Rauer M, et al. Clinical and Pathological Characteristics of KEAP1- and NFE2L2-Mutated Non-Small Cell Lung Carcinoma (NSCLC). Clin Cancer Res. 2018;24. doi:10.1158/1078-0432.CCR-17-3416

50. Nadal E, Palmero R, Muñoz-Pinedo C. Mutations in the Antioxidant KEAP1/NRF2 Pathway Define an Aggressive Subset of NSCLC Resistant to Conventional Treatments. J Thorac Oncol. 2019;14. doi:10.1016/j.jtho.2019.08.005

51. Subramanian J, Govindan R. Lung Cancer in Never Smokers: A Review. J Clin Oncol. 2016 [cited 25 Apr 2024]. doi:10.1200/JCO.2006.06.8015

52. Scott J, Marusyk A. Somatic clonal evolution: A selection-centric perspective. Biochim Biophys Acta. 2017;1867: 139–150.

53. Fleenor CJ, Marusyk A, DeGregori J. Ionizing radiation and hematopoietic malignancies: altering the adaptive landscape. Cell Cycle. 2010;9: 3005–3011.

54. Hill W, Lim EL, Weeden CE, Lee C, Augustine M, Chen K, et al. Lung adenocarcinoma promotion by air pollutants. Nature. 2023;616: 159–167.

55. Bossé Y, Postma DS, Sin DD, Lamontagne M, Couture C, Gaudreault N, et al. Molecular signature of smoking in human lung tissues. Cancer Res. 2012;72: 3753–3763.

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

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

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

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

60. Peters EN, Warren GW, Sloan JA, Marshall JR. Tobacco assessment in completed lung cancer treatment trials: Pooling Project: Smoking and Lung Cancer. Cancer. 2016;122: 3260–3262.

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

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

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

64. Igawa S, Sasaki J, Otani S, Ishihara M, Takakura A, Katagiri M, et al. Impact of smoking history on the efficacy of gefitinib in patients with non-small cell lung cancer harboring activating epidermal growth factor receptor mutations. Oncology. 2015;89: 275–280.

65. van de Haar J, Canisius S, Yu MK, Voest EE, Wessels LFA, Ideker T. Identifying epistasis in cancer genomes: A delicate affair. Cell. 2019;177: 1375–1383.

66. Romero R, Sayin VI, Davidson SM, Bauer MR, Singh SX, LeBoeuf SE, et al. Keap1 loss promotes Kras-driven lung cancer and results in dependence on glutaminolysis. Nat Med. 2017;23: 1362–1368.

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

68. Galan-Cobo A, Sitthideatphaiboon P, Qu X, Poteete A, Pisegna MA, Tong P, et al. LKB1 and KEAP1/NRF2 pathways cooperatively promote metabolic reprogramming with enhanced glutamine dependence in KRAS-mutant lung adenocarcinoma. Cancer Res. 2019;79: 3251–3267.

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

70. Kwack WG, Shin SY, Lee SH. Primary resistance to immune checkpoint blockade in an STK11/TP53/KRAS-mutant lung adenocarcinoma with high PD-L1 expression. Onco Targets Ther. 2020;13: 8901–8905.

71. Arbour KC, Jordan E, Kim HR, Dienstag J, Yu HA, Sanchez-Vega F, et al. Effects of Co-occurring Genomic Alterations on Outcomes in Patients with KRAS-Mutant Non–Small Cell Lung Cancer. Clin Cancer Res. 2018;24: 334–340.

72. Proulx-Rocray F, Routy B, Nassabein R, Belkaid W, Tran-Thanh D, Malo J, et al. The

prognostic impact of KRAS, TP53, STK11 and KEAP1 mutations and their influence on the NLR in NSCLC patients treated with immunotherapy. Cancer Treat Res Commun. 2023;37: 100767.

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

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

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

76. Aredo JV, Padda SK, Kunder CA, Han SS, Neal JW, Shrager JB, et al. Impact of KRAS mutation subtype and concurrent pathogenic mutations on non-small cell lung cancer outcomes. Lung Cancer. 2019;133: 144–150.

77. Wohlhieter CA, Richards AL, Uddin F, Hulton CH, Quintanal-Villalonga À, Martin A, et al. Concurrent mutations in STK11 and KEAP1 promote ferroptosis protection and SCD1 dependence in lung cancer. Cell Rep. 2020;33: 108444.

78. Julian C, Pal N, Gershon A, Evangelista M, Purkey H, Lambert P, et al. Overall survival in patients with advanced non-small cell lung cancer with KRAS G12C mutation with or without STK11 and/or KEAP1 mutations in a real-world setting. BMC Cancer. 2023;23. doi:10.1186/s12885-023-10778-6

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

80. Bao Y, Zhang S, Zhang X, Pan Y, Yan Y, Wang N, et al. RBM10 Loss Promotes EGFR-Driven Lung Cancer and Confers Sensitivity to Spliceosome Inhibition. Cancer Res. 2023;83: 1490–1502.

81. Nanjo S, Wu W, Karachaliou N, Blakely CM, Suzuki J, Chou YT, et al. Deficiency of the splicing factor RBM10 limits EGFR inhibitor response in EGFR-mutant lung cancer. The Journal of clinical investigation. 2022;132. doi:10.1172/JCI145099

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

83. Offin M, Chan JM, Tenet M, Rizvi HA, Shen R, Riely GJ, et al. Concurrent RB1 and TP53 alterations define a subset of EGFR-mutant lung cancers at risk for histologic transformation and inferior clinical outcomes. J Thorac Oncol. 2019;14: 1784–1793.

84. Pros E, Saigi M, Alameda D, Gomez-Mariano G, Martinez-Delgado B, Alburquerque-Bejar

JJ, et al. Genome-wide profiling of non-smoking-related lung cancer cells reveals common RB1 rearrangements associated with histopathologic transformation in EGFR-mutant tumors. Ann Oncol. 2020;31: 274–282.

85. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the Evolution of Non-Small-Cell Lung Cancer. N Engl J Med. 2017;376: 2109–2121.

86. Frankell AM, Dietzen M, Al Bakir M, Lim EL, Karasaki T, Ward S, et al. The evolution of lung cancer and impact of subclonal selection in TRACERx. Nature. 2023;616: 525–533.

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

88. Guénolé A, Srivas R, Vreeken K, Wang ZZ, Wang S, Krogan NJ, et al. Dissection of DNA damage responses using multiconditional genetic interaction maps. Mol Cell. 2013;49: 346–358.

89. Harrison R, Papp B, Pál C, Oliver SG, Delneri D. Plasticity of genetic interactions in metabolic networks of yeast. Proc Natl Acad Sci U S A. 2007;104: 2307–2312.

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

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

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

93. Costanzo M, Hou J, Messier V, Nelson J, Rahman M, VanderSluis B, et al. Environmental robustness of the global yeast genetic interaction network. Science. 2021;372. doi:10.1126/science.abf8424

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

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

96. Bolton KL, Ptashkin RN, Gao T, Braunstein L, Devlin SM, Kelly D, et al. Cancer therapy shapes the fitness landscape of clonal hematopoiesis. Nat Genet. 2020;52: 1219–1226.

97. Salehi S, Kabeer F, Ceglia N, Andronescu M, Williams MJ, Campbell KR, et al. Clonal fitness inferred from time-series modelling of single-cell cancer genomes. Nature. 2021;595: 585–590.

98. Liggett LA, DeGregori J. Changing mutational and adaptive landscapes and the genesis of

cancer. Biochim Biophys Acta Rev Cancer. 2017;1867: 84–94.

99. Hsu T-K, Asmussen J, Koire A, Choi B-K, Gadhikar MA, Huh E, et al. A general calculus of fitness landscapes finds genes under selection in cancers. Genome Res. 2022;32: 916–929.

100. Misra N, Szczurek E, Vingron M. Inferring the paths of somatic evolution in cancer. Bioinformatics. 2014;30: 2456–2463.

101. Marchetti A, Milella M, Felicioni L, Cappuzzo F, Irtelli L, Del Grammastro M, et al. Clinical implications of KRAS mutations in lung cancer patients treated with tyrosine kinase inhibitors: an important role for mutations in minor clones. Neoplasia. 2009;11: 1084–1092.

102. Hellyer JA, Stehr H, Das M, Padda SK, Ramchandran K, Neal JW, et al. Impact of KEAP1/NFE2L2/CUL3 mutations on duration of response to EGFR tyrosine kinase inhibitors in EGFR mutated non-small cell lung cancer. Lung Cancer. 2019;134: 42–45.

103. Krall EB, Wang B, Munoz DM, Ilic N, Raghavan S, Niederst MJ, et al. KEAP1 loss modulates sensitivity to kinase targeted therapy in lung cancer. Elife. 2017;6. doi:10.7554/eLife.18970

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

105. Weigelt B, Reis-Filho JS. Epistatic interactions and drug response. J Pathol. 2014;232: 255–263.

106. Wilkins JF, Cannataro VL, Shuch B, Townsend JP. Analysis of mutation, selection, and epistasis: an informed approach to cancer clinical trials. Oncotarget. 2018;9: 22243–22253.

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

108. Bria E, Pilotto S, Amato E, Fassan M, Novello S, Peretti U, et al. Molecular heterogeneity assessment by next-generation sequencing and response to gefitinib of EGFR mutant advanced lung adenocarcinoma. Oncotarget. 2015;6: 12783–12795.

109. Vokes NI, Chambers E, Nguyen T, Coolidge A, Lydon CA, Le X, et al. Concurrent TP53 Mutations Facilitate Resistance Evolution in EGFR-Mutant Lung Adenocarcinoma. J Thorac Oncol. 2022;17: 779–792.

110. Canale M, Petracci E, Delmonte A, Chiadini E, Dazzi C, Papi M, et al. Impact of TP53 Mutations on Outcome in EGFR-Mutated Patients Treated with First-Line Tyrosine Kinase Inhibitors. Clin Cancer Res. 2017;23: 2195–2202.

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

112. Canale M, Petracci E, Delmonte A, Bronte G, Chiadini E, Ludovini V, et al. Concomitant TP53 Mutation Confers Worse Prognosis in EGFR-Mutated Non-Small Cell Lung Cancer

Patients Treated with TKIs. J Clin Med Res. 2020;9. doi:10.3390/jcm9041047

113. Choudhury NJ, Lavery JA, Brown S, de Bruijn I, Jee J, Tran TN, et al. The GENIE BPC NSCLC cohort: A real-world repository integrating standardized clinical and genomic data for 1,846 patients with non-small cell lung cancer. Clin Cancer Res. 2023;29: 3418–3428.

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

115. Donehower LA, Soussi T, Korkut A, Liu Y, Schultz A, Cardenas M, et al. Integrated analysis of TP53 gene and pathway alterations in The Cancer Genome Atlas. Cell Rep. 2019;28: 1370–1384.e5.

116. Rho JK, Choi YJ, Ryoo B-Y, Na III, Yang SH, Kim CH, et al. p53 enhances gefitinib-induced growth inhibition and apoptosis by regulation of Fas in non-small cell lung cancer. Cancer Res. 2007;67: 1163–1169.

117. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990;61: 759–767.

118. Paterson C, Clevers H, Bozic I. Mathematical model of colorectal cancer initiation. Proc Natl Acad Sci U S A. 2020;117: 20681–20688.

119. Ortmann CA, Kent DG, Nangalia J, Silber Y, Wedge DC, Grinfeld J, et al. Effect of mutation order on myeloproliferative neoplasms. N Engl J Med. 2015;372: 601–612.

120. Ernst SM, Mankor JM, van Riet J, von der Thüsen JH, Dubbink HJ, Jgjv A, et al. Tobacco Smoking-Related Mutational Signatures in Classifying Smoking-Associated and Nonsmoking-Associated NSCLC. J Thorac Oncol. 2023;18. doi:10.1016/j.jtho.2022.11.030

121. Hobbs SD, Wilmink AB, Adam DJ, Bradbury AW. Assessment of smoking status in patients with peripheral arterial disease. J Vasc Surg. 2005;41. doi:10.1016/j.jvs.2004.12.039

122. Hald J, Overgaard J, Grau C. Evaluation of objective measures of smoking status--a prospective clinical study in a group of head and neck cancer patients treated with radiotherapy. Acta Oncol. 2003;42. doi:10.1080/02841860310005020

123. Brennan P, Buffler PA, Reynolds P, Wu AH, Wichmann HE, Agudo A, et al. Secondhand smoke exposure in adulthood and risk of lung cancer among never smokers: a pooled analysis of two large studies. International journal of cancer. 2004;109. doi:10.1002/ijc.11682

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

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

126. Tomanek I, Guet CC. Adaptation dynamics between copy-number and point mutations.

Elife. 2022;11. doi:10.7554/eLife.82240

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

128. Chen J, Yang H, Teo ASM, Amer LB, Sherbaf FG, Tan CQ, et al. Genomic landscape of lung adenocarcinoma in East Asians. Nat Genet. 2020;52: 177–186.

129. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell. 2012;150: 1107–1120.

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

131. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science. 2015;348: 124–128.

132. Rizvi H, Sanchez-Vega F, La K, Chatila W, Jonsson P, Halpenny D, et al. Molecular Determinants of Response to Anti-Programmed Cell Death (PD)-1 and Anti-Programmed Death-Ligand 1 (PD-L1) Blockade in Patients With Non-Small-Cell Lung Cancer Profiled With Targeted Next-Generation Sequencing. J Clin Oncol. 2018;36: 633–641.

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

134. Gillette MA, Satpathy S, Cao S, Dhanasekaran SM, Vasaikar SV, Krug K, et al. Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. Cell. 2020;182: 200–225.e35.

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

136. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014;511: 543–550.

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

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

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

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

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

142. Devarakonda S, Morgensztern D, Govindan R. Genomic alterations in lung adenocarcinoma. Lancet Oncol. 2015;16. doi:10.1016/S1470-2045(15)00077-7

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

144. Seo JS, Ju YS, Lee WC, Shin JY, Lee JK, Bleazard T, et al. The transcriptional landscape and mutational profile of lung adenocarcinoma. Genome Res. 2012;22. doi:10.1101/gr.145144.112

145. Li S, Choi YL, Gong Z, Liu X, Lira M, Kan Z, et al. Comprehensive Characterization of Oncogenic Drivers in Asian Lung Adenocarcinoma. J Thorac Oncol. 2016;11. doi:10.1016/j.jtho.2016.08.142

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

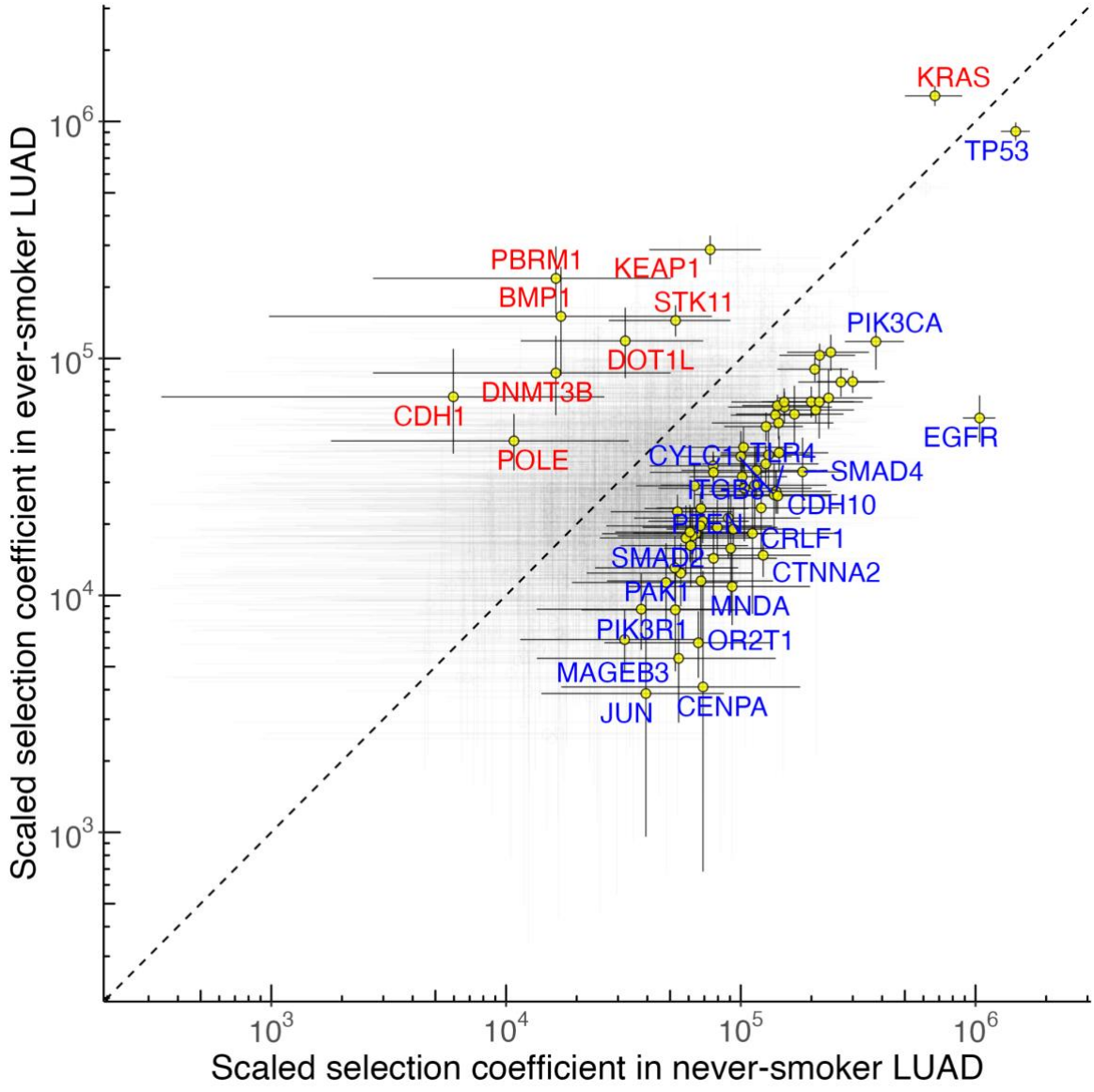
147. Wilks SS. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. aoms. 1938;9: 60–62.

148. Cannataro VL, Gaffney SG, Townsend JP. Effect sizes of somatic mutations in cancer. J Natl Cancer Inst. 2018;110: 1171–1177.

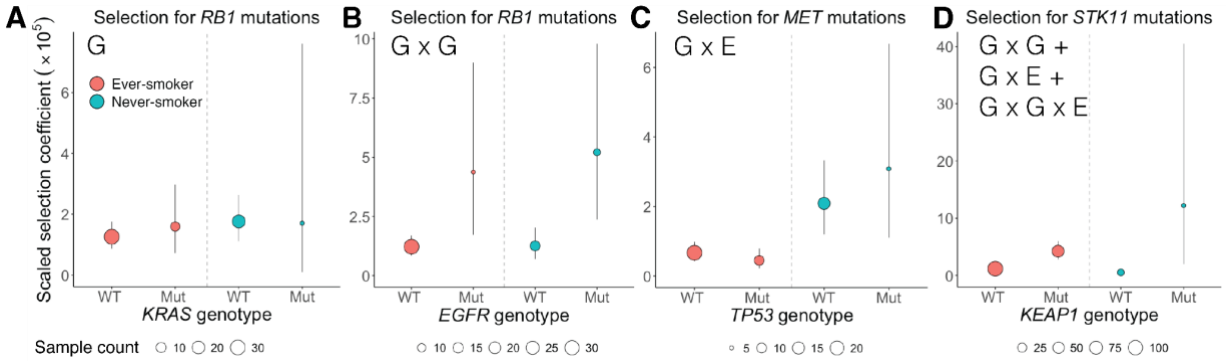
149. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15: 550.

150. Dasari K. Townsend-Lab-Yale/lung-smoking: Software for “Tobacco smoke alters the trajectory of lung adenocarcinoma evolution via effects on somatic selection and epistasis.” [cited 23 Jul 2025]. doi:10.5281/zenodo.16379015

Supporting information



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



Supplementary Figure 2. Four patterns of the influences of genetic and environmental interactions in ever-smoker and never-smoker lung adenocarcinoma. (**A**) *RB1* mutations with wild-type (WT) or mutant *KRAS*, (**B**) *RB1* mutations with wild-type or mutant *EGFR*, © MET mutations with wild-type or mutant *TP53*, and (**D**) *STk11* mutations with wild-type or mutant *KEAP1*.

Table S1. Characteristics of datasets included in study.

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

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

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