Early *NOTCH1* mutation is positively selected but epistatically suppresses evolution of later esophageal squamous-cell carcinoma drivers

Kira A. Glasmacher[1,2] (0009-0004-6300-3812), Jeffrey D. Mandell[2] (0000-0002-3839-2543), Mia Jackson[2] (0009-0005-4111-6098), Nic Fisk[3] (0000-0002-1940-393X), Vincent L. Cannataro[1†] (0000-0002-6364-7747), Jeffrey P. Townsend[2,4,5,6†] (0000-0002-9890-3907)

[1]Department of Biology, Emmanuel College, Boston, MA, [2]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, [3]Department of Cell and Molecular Biology, University of Rhode Island, Kingston, RI, [4]Department of Biostatistics, Yale School of Public Health, New Haven, CT, [5]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, [6]Yale Cancer Center, Yale University, New Haven, CT

[†]To whom correspondence should be addressed: Vincent Cannataro (617 975-9105; cannatarov@emmanuel.edu), Emmanuel College, 400 Fenway, Boston MA 02115; and Jeffrey Townsend (203 737-7042; jeffrey.townsend@yale.edu), Department of Biostatistics, Yale School of Public Health, 135 College Street, New Haven, CT 06510. The authors declare no potential conflicts of interest.

Keywords: Somatic evolution, Esophageal squamous-cell carcinoma, Selection, Epistasis, *NOTCH1* mutation

Running Title: *NOTCH1* early selection and epistasis in esophageal tissues

Abstract

Background

Somatic mutations commonly accumulate in histologically normal tissues and contribute to cancer development. However, many somatic variants found at high frequencies in normal tissues are also found at high frequencies in cancers arising from the same tissue types. Consequently, the roles of these variants in cancer development remain poorly understood. To address this, we assessed the selective pressures across steps of tumorigenesis, as well as epistasis between driver mutations.

Methods

We employed evolutionary modeling to quantify the strength of selection on somatic variants across two key steps of esophageal development: from organogenesis to clonal histologically normal epithelium, and from there to esophageal squamous-cell carcinoma. Analyzing sequence data from 2171 samples, our first-of-its kind model revealed the stepwise contributions of somatic mutations to increased cellular division and survival and their pairwise epistatic effects on driver mutation selection.

Results

NOTCH1 substitutions were strongly selected along the step from organogenesis to clonal histologically normal esophageal tissue, explaining their high prevalence in samples of this tissue. However, for the first time, we show that there is little to no positive selection for *NOTCH1, NOTCH2,* and *FAT1* mutations during progression from clonal histologically normal esophageal tissue to esophageal squamous-cell carcinoma in humans, leading to a conclusion that these substitutions promote clonal growth in normal tissue, but do not drive tumorigenesis

from extant clonal tissue growths. Moreover, we provide a somatic genetic basis for this differential role: we demonstrate for the first time that mutations in *NOTCH1* exhibit antagonistic epistasis with mutations of well-known tumor suppressor genes *TP53* and *RB1*, reducing selection for these progressive mutations in tumorigenesis.

Conclusions

Our findings demonstrate that early positively selected mutations in *NOTCH1* and other genes can shape evolutionary trajectories in ways that ultimately constrain malignant progression. Quantification of step-specific selection and identification of antagonistic epistatic interactions with key tumor suppressors reveals that somatic evolution is often context-dependent—what promotes clonal expansion in normal tissue may later impede growth or survival in tumors. These insights underscore the need for precision strategies that account for the shifting fitness landscape across premalignant and malignant stages, informing early detection, prevention, and therapeutic prioritization.

Background

Somatic mutations accumulate as our cells divide and as they are exposed to mutagens. Many of these mutations have little to no effect on the rates that cells proliferate and survive; others can lead to benign or malignant growths via increasing cell division and survival [1]. The effects of such mutations on the development of cancer have been characterized on the levels of the genome [2], the gene [3], and even individual nucleotides or amino acids [4]. Sequencing of normal or histologically precancerous tissues has revealed that many of the mutations at high prevalence in cancerous tissue can also appear at high prevalence in their presumed precursors—for example, in atypical hyperplasia preceding uterine endometrioid carcinoma [5]. However, construction of trajectories from precancerous to cancerous tissue cannot be comprehensively performed solely on the basis of relative substitution prevalence in precancerous and cancerous tissue, because substitution prevalence encompasses both the rate at which mutations arise in cell lineages and the strength of positive selection within cell populations. Accurate inference of selection requires disentangling mutational and selective processes that underlie evolutionary dynamics. Accordingly, calculation of the scaled selection coefficients early in oncogenesis should be performed by stepwise deconvolution of mutation rates across successive stages of tumor evolution such as histologically normal tissues and corresponding cancers.

Detection of somatic variants in normal tissues has historically been challenging due to their low variant allele frequencies and the heterogeneous stem-cell composition of normal tissues. Recent advances, including deep sequencing of small biopsies containing clones [6–9] and sequencing elements from microscopically distinct clonal units [10,11], have enabled the

identification of substitutions that are present in only a small fraction of cells. To infer the contribution of driver mutations to tumorigenesis, some studies have relied on variant prevalence among samples [10,12]. Prevalence alone, however, conflates the distinct influences of mutation rate and strength of selection: gene substitutions may be prevalent in normal or tumor tissue not because they are functionally important for cell survival or proliferation, but simply because they mutate more frequently in individual progenitor cells. Varying prevalences of somatic substitutions can arise from quantitative differences in underlying mutation rates that are unrelated to biological function in oncogenesis. Statistical models for driver detection have striven to account for these underlying differences across the genome by martialing both gene-level synonymous mutation rates and gene-level covariates [13]. Techniques using synonymous substitutions can quantify an underlying "neutral" mutation rate, examine their frequency relative to nonsynonymous substitutions to detect drivers [13] and quantify gene-level selection [3,8]. Indeed, these methods have revealed recurrent substitutions in cancer-associated genes within physiologically normal skin, esophageal, colorectal, and endometrial epithelial cells [6,9,10,14], as well as in clonal hematopoiesis in blood [15,16].

Application of methods that detect somatic selection in the skin epidermis has revealed strong positive selection for mutations in cancer-associated genes such as *NOTCH1* and *TP53* [17]. However, the presence of positively selected mutations in normal tissues does not necessarily imply that these mutations contribute to the development of cancer. Some variants may expand within healthy cell populations due to selective advantages unrelated to malignancy, then "hitchhike" on subsequent, independent tumorigenic mutations to become somewhat prevalent in cancerous tissues. The functional impact of a mutation can change over the course of

tumor evolution due to epistatic interactions with other mutations. Epistatic interactions altering selective impact of other mutations along an evolutionary trajectory may be common [18], in turn affecting the cancerous effects of mutations at one stage of development compared to another. Recognition and understanding of this dynamic context of selection is essential to elucidation of the precise roles of somatic mutations in tumorigenesis, and has crucial implications for informing therapy and for opening new avenues for effective drug discovery and translation [19–21].

Our understanding of the differential evolutionary trajectories in humans that lead to benign clonal expansions instead of malignant cancers is limited. A conventional understanding of clonal evolution and selection suggests that substitutions in driver genes should increase in prevalence as tissues progress from benign to premalignant to malignant neoplasms [22,23]. However, this assumption is increasingly questioned. For example, BRAF V600E substitutions have been reported in 70–88% of benign nevi but only 40–45% of melanomas, and HER2 expression is more common in ductal carcinoma in situ than in invasive breast cancer [24]. These "inverted" differences suggest that some mutations may promote clonal expansion without advancing malignancy—and in some cases, may even inhibit it.

Higa and DeGregori [25] proposed that *NOTCH1* mutant cells in the esophagus might have a lower likelihood of accumulating mutations that drive malignant progression because they have reached distinct adaptive fitness peaks associated with clonal expansions of the normal esophageal epithelium. Specifically, mutations that result in a reduction or loss of NOTCH1 function are thought to increase cell fitness within the normal epithelium, where nonsense and essential splice mutations are under especially strong positive selection [26,27]. However, this

early selective benefit may diminish in the tumor context, reducing the competitiveness of *NOTCH1*-mutant cells [28]. Indeed, a recent study in mice demonstrated that NOTCH1 mutations can simultaneously drive clonal expansion in normal mouse esophageal tissue and impair mouse tumor formation [29]. Quantification of selection under an evolutionary model that incorporates variation in underlying mutation rates would clarify the contribution of mutations in *NOTCH1* and other genes in clonal expansions within normal and tumor tissue of the human esophagus.

Here we investigate the selective effects of somatic single-nucleotide mutations in clonal histologically normal human esophageal (CHNE) tissue [generally considered to be "normal" tissue; 7,8,30] and esophageal squamous-cell carcinoma (ESCC). Deriving a novel evolutionary model enabling step-specific inference of selective effects, we quantify mutation rates and scaled selection coefficients for mutations of ESCC-associated genes in CHNE tissues and malignant tissues. To reveal how interactions between mutations influence these steps of tumor evolution, we analyzed pairwise selective epistatic effects among frequently mutated genes, including *NOTCH1* and *TP53.* This integrative approach provides insights into how context-dependent selection and genetic interactions shape early and late human tumor evolution.

Methods

Data acquisition and preprocessing

We assembled a dataset of 2757 esophageal samples, including 1016 CHNE tissues and 1741 ESCC tumors. This dataset comprised sequencing data from 624 whole-genome and 879 whole-exome samples from a combined cohort of ESCC data spanning 17 different datasets,

including TCGA, ICGC, and other published studies [31]; 25 ESCC samples with whole-exome sequencing from Liu et al. [32]; as well as 34 CHNE and 10 ESCC whole-exome samples from Yuan et al. [30] (a selection that excluded 8 samples labelled as "normal" that exhibited histological evidence of severe dysplasia or carcinoma in situ and shared mutations with matched tumors). Targeted sequencing data were obtained for 137 ESCC samples from cBioPortal [33–36], 142 CHNE and 66 ESCC samples from Yokoyama et al. [7], and 832 CHNE samples from Martincorena et al. [8]. To ensure data integrity and avoid redundancy, we excluded samples that we identified as clonally related through shared mutations or spatial proximity in the Yokoyama and Martincorena datasets: biopsies from adjacent regions were retained only if they didn't share any variants. To ensure robustness and generalizability, we included samples from a wide range of populations (**Supplementary Table S1).** We also used datasets that were generated through various data processing pipelines (**Supplementary Table S2**), minimizing bias from any single methodology.

We compiled a list of ESCC-associated genes based on previous studies of the somatic variant landscapes of both CHNE epithelium and ESCC [37–43]. To ensure statistical power and population-wide relevance, we limited our analyses to genes with substitutions observed in 25 or more samples across the combined dataset. For inference of selection, we focused on single-nucleotide variants (SNVs) and excluded nucleotide substitutions occurring within one or two base pairs of one another in the same tumor sample, as these may represent clustered or multi-nucleotide substitutions with distinct rates of occurrence from independent SNVs in the same tumor. Reanalysis of the data without exclusion of these possible multi-nucleotide events

did not result in any substantial differences to our conclusions.

Stepwise mutation rates

To model mutation accumulation across sequential steps of esophageal tumorigenesis, we adopted a stepwise framework representing the tissue progression from esophageal organogenesis to CHNE tissue, and in some cases to ESCC. We discretized this continuous process into two evolutionary steps determined by pathology: Step 0→1 from esophageal organogenesis to CHNE tissue, and Step 1→2 from CHNE tissue to ESCC. During these steps, genes could be subject to differing mutation rates. To estimate gene-specific proportions of early and late mutations separately for CHNE and ESCC samples, we used dNdScv [3], incorporating all whole-exome and whole-genome samples and diverse tissue-specific covariates [4]. We evaluated the fits of the negative binomial model underlying dNdScv by consideration of the estimated overdispersion parameter θ. In the ESCC cohort, $\theta = 5.23$ indicated a suitable fit, with a variance below the mean expected mutation rate. In contrast, the CHNE cohort yielded $\theta < 1$, suggesting overdispersion and greater-than-expected variance [3]. To assess the robustness of estimates obtained over subsets of the data, we performed a jackknife procedure by iteratively omitting individual samples and re-estimating model parameters. These jackknifed subsets yielded a satisfactory mean estimate of 1.86 for the overdispersion parameter. Moreover, analysis of all jackknifed subsets yielded results consistent with the conclusions of our study.

For each pathologically-defined tissue set, we used dNdScv to obtain expected numbers of synonymous substitutions (*exp_syn_cv*) for each gene using its covariate-informed model. We normalized these values by dividing by the number of samples in each cohort and then by the

number of synonymous sites per gene, yielding expected rates of mutation per nucleotide for each gene *g*, $E_{0\rightarrow1,g}$ and $E_{0\rightarrow2,g}$. We assessed the proportion of mutations accumulated along the step from esophageal organogenesis to CHNE tissue (Step 0→1) for each gene, computing $p_{0\rightarrow1,g} = E_{0\rightarrow1,g}/E_{0\rightarrow2,g}$. Similarly, the proportion of mutations accumulated along the step from CHNE tissue to ESCC (Step 1→2) was assessed as $p_{1\rightarrow2,g} = 1 - p_{0\rightarrow1,g}$. These step-specific proportions were used to model mutation and selection separately across the two discrete steps of somatic evolution and tumor progression.

We used cancereffectsizeR v. 2.10.2 [44] to calculate sample-specific mutation rates for designated sets of genomic sites and to model stepwise selection within the dataset. For each tumor *i* and designated set of sites in a gene *g,* cancereffectsizeR calculated the mutation rate $\mu_{0\rightarrow2,g,i}$ summed over the designated sites, informed by synonymous substitution rates of genes ($E_{0\rightarrow2,g}$) and by sample-specific mutational spectra derived from mutational signature analysis [4]. We partitioned the mutation rates for designated sets of sites across two evolutionary steps: $\mu_{0\rightarrow1,g,i} = p_{0\rightarrow1,g} \times \mu_{0\rightarrow2,g,i}$, representing the mutation rate from esophageal organogenesis to CHNE tissue (Step 0→1), and $\mu_{1\rightarrow2,g,i} = p_{1\rightarrow2,g} \times \mu_{0\rightarrow2,g,i}$, representing the rate during the progression from CHNE tissue to ESCC (Step 1→2). These rates were crucial for the estimation of selection coefficients within each step of tumorigenesis.

To designate the set of relevant variants for distinct estimates of selection, we identified recurrently mutated positions for each gene. For the oncogenes *PIK3CA* and *NFE2L2,* sites were designated as putatively selected if they were recurrently mutated. For the tumor-suppressor genes *FAT1*, *NOTCH1*, *NOTCH2*, *TP53*, *FBXW7*, and *RB1*, the designated set included recurrently mutated sites and all nonsense/stop-loss variants [44]. Gene classification as

oncogenes or tumor suppressors followed Chakravarty et al. [45]. All reported mutation rates and scaled selection coefficients were calculated over these designated sets of putatively selected variants.

Stepwise scaled selection calculations

In Mandell et al. [44], likelihood for the strength of selection γ was calculated for any single cohort of *N* tumors by maximizing

$$\mathcal{L}(\gamma|\mu_1,\ldots,\mu_n,\ldots,\mu_N) = \prod_{i=1}^{n} \mathrm{e}^{-\mu_i\gamma} \times \prod_{i=n+1}^{N} 1-\mathrm{e}^{-\mu_i\gamma}, \qquad \textbf{(1)}$$

where μ_i, $1 \leq i \leq N$, is the mutation rate of a defined set of sites in sample i, and where $n < N$ are defined such that any variant in the defined set of sites is absent in n samples and is present in samples $n+1\ldots N$, and where the subscript *g* denoting gene identity is suppressed in **Eq. 1** and later equations for clarity of presentation.

To estimate step-specific scaled selection coefficients within a two-step sequential model of tumorigenesis, we elaborated the likelihood formulated by **Eq. 1** to differentiate Step 0→1 and Step 1→2:

$$\mathcal{L}(\gamma_{0\to1},\gamma_{1\to2}|\mu_{0\to1,i}\forall i\in 1\ldots M+N,\mu_{1\to2,i}\forall i\in 1\ldots N)=$$
$$\prod_{i=1}^{n}(\mathrm{e}^{-\mu_{0\to1,i}\times\gamma_{0\to1}}\times\mathrm{e}^{-\mu_{1\to2,i}\times\gamma_{1\to2}})\times$$
$$\prod_{i=n+1}^{N}(1-\mathrm{e}^{-\mu_{0\to1,i}\times\gamma_{0\to1}}\times\mathrm{e}^{-\mu_{1\to2,i}\times\gamma_{1\to2}})\times$$
$$\prod_{i=N+1}^{N+m}\mathrm{e}^{-\mu_{0\to1,i}\times\gamma_{0\to1}}\times$$
$$\prod_{i=N+m+1}^{N+M}(1-\mathrm{e}^{-\mu_{0\to1,i}\times\gamma_{0\to1}}) \quad , \qquad \textbf{(2)}$$

for a dataset with M CHNE samples, of which $M-m$ exhibit a substitution and m do not, and where there are N ESCC samples in total, of which $N-n$ exhibit a substitution and n do not. The integer variable *i* above is used to index CHNE samples $i=1+N,\ldots,m+N,\ldots,M+N$ and ESCC samples $i=1,\ldots,n,\ldots,N$. The first product term in **Eq. 2** represents ESCC samples where all variants in the designated set of variants are absent. The second product term represents ESCC samples where any variant in the designated set is present. The third product represents CHNE samples where all variants in the designated set of variants are absent. The fourth product represents CHNE samples with any of the variants in the designated set present.

The likelihood formulation in **Eq. 2** can be generalized to accommodate any number of sequential steps of tumorigenesis. For any number of steps S,

$$\mathcal{L}(\gamma_{0\to1},\ldots,\gamma_{S-1\to S}|\mu_{s-1\to s,i}\forall i\in 1\cdots\sum_{a=s}^{S}N_a)=$$
$$\prod_{h=1}^{S}\left(\prod_{i=(\sum_{a=h+1}^{S}N_a)+1}^{(\sum_{a=h+1}^{S}N_a)+n_h}\left(\prod_{j=1}^{h}\mathrm{e}^{-\mu_{j-1\to j,i}\times\gamma_{j-1\to j}}\right)\times\prod_{i=(\sum_{a=h+1}^{S}N_a)+n_h+1}^{\sum_{a=h}^{S}N_a}\left(1-\prod_{j=1}^{h}\mathrm{e}^{-\mu_{j-1\to j,i}\times\gamma_{j-1\to j}}\right)\right), \textbf{(3)}$$

where $s = 1, \ldots, S$ denotes the step of progression toward cancer, and S corresponds to the last step present in the data. For each step s, there are N_s samples: n_s without a variant and $N_s - n_s$ with a variant in the designated set of variants. The product index *h* specifies the step of a sample, *i* specifies the sample index, and *j* functions as an index that breaks down all *h* steps of a sample's trajectory into the distinct stepwise processes. For each step, the samples are indexed by

$$i_s = 1 + \sum_{a=s+1}^{S} N_a, \ldots, n_s + \sum_{a=s+1}^{S} N_a, \ldots, N_s + \sum_{a=s+1}^{S} N_a$$

resulting in

$$i = 1, \ldots, \sum_{a=1}^{S} N_s$$

across all steps. By multiplying over all transitions in the trajectory, this likelihood captures the step-specific selection acting across a multi-step evolutionary process. This formulation enables flexible inference of selection coefficients in settings where data from intermediate steps—such as precancerous lesions—are available.

To determine whether scaled selection coefficients differed significantly across the steps of tumorigenesis, we performed a likelihood ratio test for each gene. Specifically, we compared the fit of the stepwise model—which estimates separate selection coefficients for each step—to a nested model that assumes a single, constant selection coefficient across all steps. *P* values were calculated based on a X^2 distribution for the likelihood ratio test statistic with one degree of freedom ($df = s - 1$, where *s* represents the total number of steps in the stepwise model) under

the null hypothesis of uniform selection. For reporting purposes, all tests yielding $P < 0.001$ were denoted as such, without additional decimal precision.

Mutation rates for epistatic selection

To maximize statistical power and to capture epistatic interactions across the full developmental trajectory, we conducted an independent analysis using the combined set of tissue sequences from all samples—both CHNE and ESCC. Within a standard cancereffectsizeR workflow [44], we calculated sample-specific site mutation rates without partitioning the mutation rate into discrete steps. These sample-specific site mutation rates were then used to calculate expected substitution probabilities for modeling pairwise genetic interactions.

Epistatic scaled selection calculation

To improve on mutual exclusivity or co-occurrence analyses commonly applied in cancer genomics [e.g. 48,49], we conducted a likelihood-based analysis of epistasis. This method accounts for commonalities of underlying mutation rates for some driver gene pairs that are driven by common underlying mutational processes [47]. We first assessed the likelihood of the data under a null model of no selective epistasis. In this model, the Poisson rate of substitution for mutation A in each tumor i was estimated as $\lambda_{A_i} = \mu_{A_i}\gamma_A$ [47], where μ_{A_i} is the sample-specific neutral mutation rate, γ_A is the scaled selection coefficient for mutation A, and substitution is defined as acquisition of a mutation in a cell followed by its fixation in all or nearly all of the cells in a population. In the null no-epistasis model, the scaled selection coefficient γ_A was assumed to be constant, regardless of the presence or absence of a second

variant B in another gene. Similarly, the selection coefficient for mutation B, γ_B, was inferred independently of the mutational status of A. In the absence of selective epistasis, substitutions are acquired independently with distinct selection coefficients γ_A and γ_B, which are therefore calculated separately as in **Eq. 1**.

For comparison, we analyzed the likelihood of the data under a model of pairwise epistatic selection. In this model, each tumor is characterized by four estimated rates of acquisition for each pair of variants: $\lambda_{A|\emptyset}$ and $\lambda_{B|\emptyset}$ for mutations A and B in the absence of the other substitution, and $\lambda_{A|B}$ and $\lambda_{B|A}$ for acquisition of A and B in the presence of the other substitution (for brevity, omitting the tumor index i arising from tumor-specific μ_{A_i}.) These substitution rates were modeled as the product of sample-specific mutation rates and epistatic scaled selection coefficients. For this analysis, the underlying mutation rate is assumed to be constant for each gene across the full evolutionary trajectory, while selection on any variant may depend on the mutational background of another gene. We estimated selection by maximizing the likelihood function—multiplying over distinct probabilities of samples having only variant A, only variant B, neither variant, or both variants present—as in the pairwise epistasis approach of Alfaro-Murillo and Townsend [46]. We also used this model of pairwise epistatic selection to infer the order of mutation fixation through a continuous-time Markov chain that captures both possible directions of mutation fixation [46].

We performed likelihood ratio tests to compare the fit of the epistatic model versus the no-epistasis models for each gene pair, and classified selective epistatic relationships as synergistic or antagonistic based on the statistical significance of the epistatic model fit parameterized by underlying gene and trinucleotide mutation rates. A relationship was

considered synergistic if the epistatic model predicted co-substitution patterns significantly better than the no-epistasis model, and predicted more co-occurrences than the no-epistasis model. Conversely, a relationship was classified as antagonistic when the epistatic model predicted co-substitution patterns significantly better than the no-epistasis model and predicted fewer co-occurrences than were expected in the no-epistasis model.

Results

Cell-lineage expected mutational burdens differ among normal and tumor tissues

Across all whole-exome sequenced samples, the mean number of somatic substitutions was significantly higher in ESCC tumors than in CHNE tissues ($P < 0.001$). For each gene analyzed, the expected mutational burden per sample per synonymous site along the evolutionary trajectory from esophageal organogenesis to CHNE tissue ($E_{0\rightarrow 1,g}$) was lower than the corresponding expected burden along the evolutionary trajectory from esophageal organogenesis to tumor tissue ($E_{0\rightarrow 2,g}$; **Fig. 1**). On average, the expected mutational burden in ESCC in the absence of selection was 38 times greater than the expected burden in CHNE tissues. The distribution of these ratios was skewed, with a number of genes exhibiting substantially higher ratios. Among ESCC-associated genes, *NOTCH1* exhibited the greatest differential: an expected burden for ESCC tumors that is 116 times greater than the expected burden in CHNE tissues in the absence of selection.

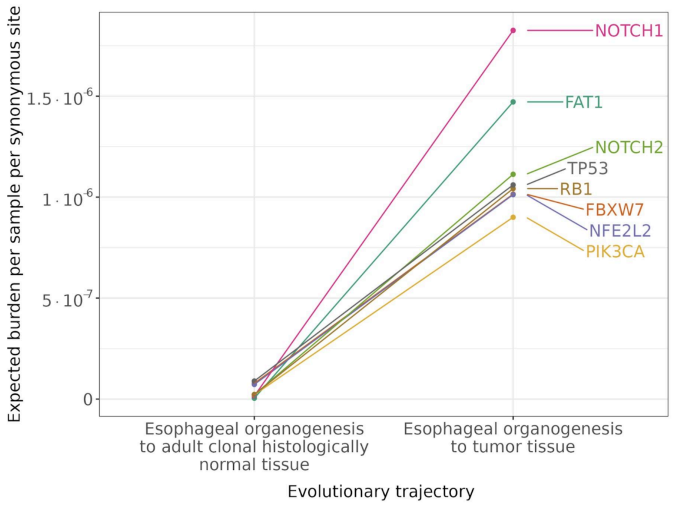

Figure 1. Step-specific expected mutation burden in genes associated with esophageal squamous-cell carcinoma across evolutionary trajectories. Gene-specific expected burden per sample per synonymous site along the evolutionary trajectory from esophageal organogenesis to adult clonal histologically normal tissue and from esophageal organogenesis to esophageal squamous-cell carcinoma, for a selection of esophageal squamous-cell carcinoma-associated genes.

Somatic mutations in NOTCH1, FAT1, and NOTCH2 exhibit strong selection in the transition to adult clonal histologically normal tissue, but are not selected in the transition to malignancy.

In terms of variant prevalence, *NOTCH1* substitutions were present in 61% of the sequences of CHNE tissue. In contrast, they were only present in 23% of the sequences of tumor samples. One explanation for a difference in prevalence is a difference in underlying cellular mutation rate. However, the expected mutational burden per sample per site for *NOTCH1* during the trajectory from esophageal organogenesis to CHNE biopsy is low (1.57×10^{-8}) relative to the trajectory from esophageal organogenesis to ESCC resection (1.83×10^{-6}). The higher prevalence of *NOTCH1* substitutions in CHNE contrasts with the lower underlying expected mutational burden, a result consistent with a previous report [8]. Our analysis demonstrates, furthermore, that this discrepancy is the consequence of strong selection for *NOTCH1* driver substitutions along the step from organogenesis to CHNE tissue, along with selection against, no selection for, or very weak selection for *NOTCH1* driver substitutions along the step from CHNE tissue to ESCC (95% CI $0 \leq \gamma \sim 0 < 5.1$; $0 < \gamma < 1$ corresponds to selection against fixation; **Fig. 2**).

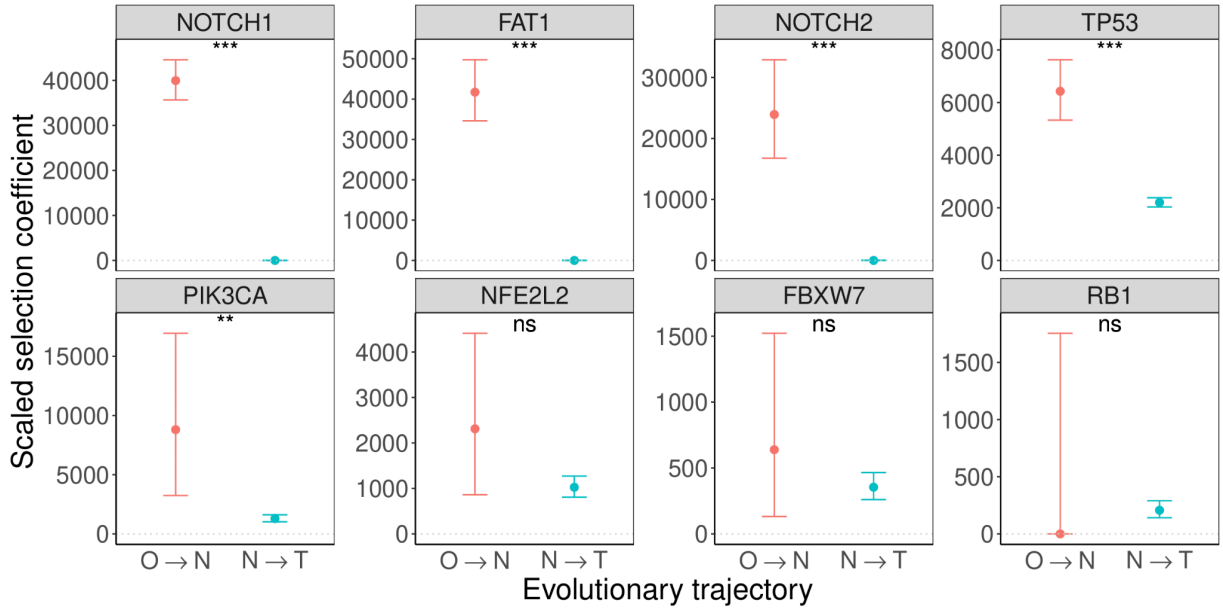

Figure 2. Step-specific selection on genes associated with esophageal squamous-cell carcinoma along the trajectory toward ESCC. Scaled selection coefficients of mutations in genes previously reported as ESCC-drivers *NOTCH1*, *FAT1*, *NOTCH2*, *TP53, PIK3CA*, *NFE2L2, FBXW7*, and *RB1* in evolutionary steps from esophageal organogenesis to adult clonal histologically normal tissue (O → N) and adult clonal histologically normal tissue to primary esophageal squamous-cell tumor (N → T) with 95% confidence intervals. Genes are ordered by statistical significance, as determined by likelihood ratio tests comparing step-specific to one-step selection models (***: $P \leq 0.001$, **: $P \leq 0.01$, *: $P \leq 0.05$, ns: not significant).

NOTCH1 is not the only gene exhibiting this pattern of low expected mutational burden and yet high prevalence in CHNE tissue, and conversely higher expected mutational burden but low prevalence in ESCC tumors. *FAT1* and *NOTCH2* show similar trends. For *FAT1*, the expected mutational burden is 4.64×10^{-9} in CHNE tissue and 1.47×10^{-6} in tumor tissue. In *NOTCH2*, the expected mutational burden is 2.22×10^{-8} in CHNE tissue and 1.11×10^{-6} in tumor tissue. Despite their higher expected mutational burdens in tumors, *FAT1* and *NOTCH2* mutations are more prevalent in CHNE tissue (12.6% and 7.2%) compared to tumor tissue (7.4% and 2.0%). Our analysis demonstrates that this discrepancy is the consequence of strong

selection for driver substitutions in *FAT1* and *NOTCH2* along the step from organogenesis to CHNE tissue ($34{,}608.1 < \gamma \sim 41{,}718.9 < 49{,}743.2$; $16{,}757.9 < \gamma \sim 23{,}925.3 < 32{,}892.7$), along with selection against, no selection for, or very weak selection for *FAT1* driver substitutions ($0 \leq \gamma \sim 0 < 11.2$) and *NOTCH2* driver substitutions ($0 \leq \gamma \sim 0 < 38.2$) along the step from CHNE tissue to ESCC.

Selection on mutated sites of *TP53*, *PIK3CA*, *NFE2L2*, and *FBXW7* was higher along the step from organogenesis to CHNE tissue than along the step from CHNE tissue to ESCC ($P < 0.001$, $P = 0.006$, $P = 0.200$, and $P = 0.496$, respectively). However, the substitution prevalence is lower in CHNE tissue compared to ESCC tumors for *TP53* (25.8% versus 73.0%), *PIK3CA* (0.9% versus 6.2%), *NFE2L2* (0.7% versus 6.4%), and *FBXW7* (0.7% versus 3.8%). This difference in prevalence can be attributed to continued substantial selection for new driver variants in these genes along the step of malignant growth of tumor tissue. This constancy stands in contrast to *NOTCH1*, *NOTCH2*, and *FAT1*, in which driver-site selection was abrogated along the step from CHNE tissue to ESCC.

Selection on some mutations of cancer-driving genes depends on the presence or absence of other cancer driver substitutions

Analysis of epistatic interactions along the full evolutionary trajectory across both steps from esophageal organogenesis to tumor resection revealed substantial synergistic selective epistasis of *NOTCH1* driver-site mutations with mutations of *NOTCH2* ($P < 0.001$; **Fig. 3A**), meaning that the calculated mutation rates of *NOTCH1* and *NOTCH2* variants in esophageal tissues would lead us to expect a lower frequency of co-occurrence under a no-epistasis model

(5 cases) than were observed (14 cases), and that mutations of *NOTCH2* were substantially more strongly selected in the context of mutated *NOTCH1* than in the context of wildtype *NOTCH1* ($\gamma \sim 2713$ versus $\gamma \sim 302$).

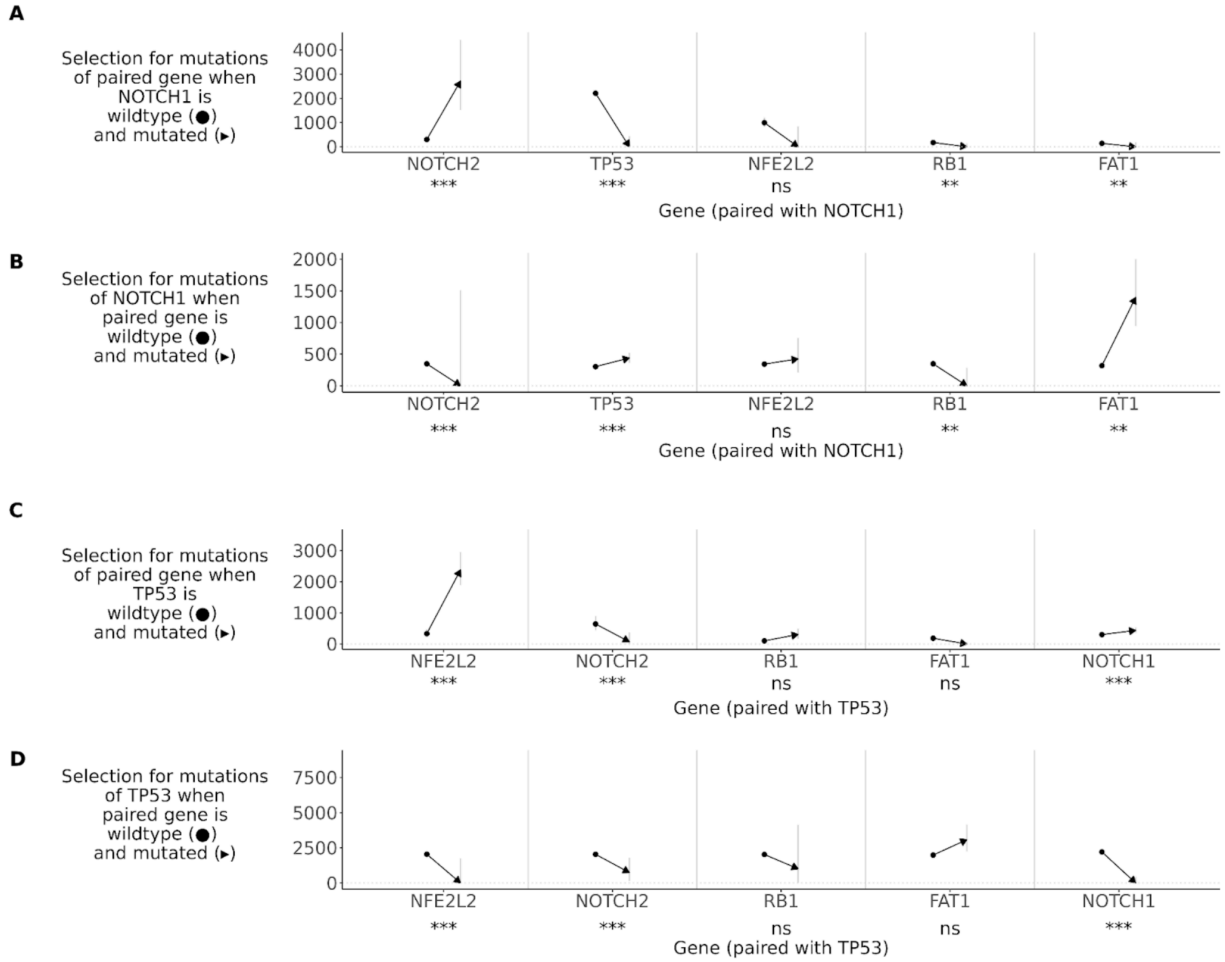

Figure 3. Pairwise selective epistasis of mutations in genes under high individual selection in esophageal tissue development. (**A**) Context-specific (epistatic) scaled selection coefficients for mutations in *NOTCH2*, *TP53, NFE2L2, RB1,* and *FAT1* on *NOTCH1* wildtype and mutant background. (**B**) Context-specific (epistatic) scaled selection coefficients for mutations in *NOTCH1* on wildtype and mutant backgrounds of *NOTCH2*, *TP53, NFE2L2, RB1,* and *FAT1*. (**C**) Context-specific (epistatic) scaled selection coefficients for mutations in *NFE2L2*, *NOTCH2, RB1, FAT1,* and *NOTCH1* on *TP53* wildtype and mutant background. (**D**) Context-specific (epistatic) scaled selection coefficients for mutations in *TP53* on wildtype and mutant backgrounds on *NFE2L2*, *NOTCH2, RB1, FAT1,* and *NOTCH1.*

In contrast to the synergistic interactions, several genes exhibited underlying mutation rates of variants in esophageal tissues would lead us to expect a higher frequency of co-occurrence with *NOTCH1* than is observed. This discrepancy is explained by substantial antagonistic epistasis: a reduction in selection in the context of mutated *NOTCH1* that is evident for mutations of *TP53* ($P < 0.001$), *RB1* ($P < 0.01$) and *FAT1* ($P < 0.01$; **Fig. 3A**). For *TP53*, a major driver of esophageal squamous-cell carcinoma tumorigenesis [7,8,50–53], selection decreased significantly from substantially positively selected in tissue without *NOTCH1* mutation (95% CI $2080 < \gamma \sim 2210 < 2345$) to estimated selection *against* fixation once *NOTCH1* is mutated ($0 \leq \gamma \sim 0 < 434$; $0 < \gamma < 1$ corresponds to selection against fixation). Similarly, mutations of *NOTCH1* exhibited antagonistic epistasis with mutations of *RB1*—another known driver of esophageal squamous-cell carcinoma tumorigenesis [52,53]. Driver-site mutation of *NOTCH1* decreased selection for *RB1* mutation from positively selected in wild-type tissue ($120 < \gamma \sim 175 < 245$) to estimated selection *against* fixation once *NOTCH1* is mutated ($0 \leq \gamma \sim 0 < 139$).

In the case of *NOTCH1* and *RB1*, antagonistic epistasis was mutual and independent of mutation order; mutation of *RB1* similarly results in a substantial decrease from positive selection on driver-site mutations of *NOTCH1* in wildtype ($310 < \gamma \sim 347 < 387$) to estimated selection *against* fixation once *RB1* is mutated ($0 \leq \gamma \sim 0 < 282$). However, in the cases of *NOTCH2*, *TP53*, and *FAT1*, driver mutations in these genes exhibited statistically significant epistasis estimated to be opposite in direction to that observed when *NOTCH1* was mutated first: selection on driver-site mutation of *NOTCH1* increased on a mutant background of *TP53* ($P < 0.001$) or *FAT1* ($P < 0.01$), and decreased on a mutant background of *NOTCH2* ($P < 0.001$; **Fig. 3B**). Whereas selection on driver-site mutations of *TP53* was antagonized by extant

NOTCH1 mutation, selection on driver-site mutations of *NOTCH1* was higher in the context of extant *TP53* driver-site mutation ($364 < \gamma \sim 439 < 524$ versus $262 < \gamma \sim 303 < 348$). Similarly, whereas selection on driver-site mutations of *FAT1* were antagonized by extant *NOTCH1* mutation, selection on driver-site mutations of *NOTCH1* was higher in the context of extant *FAT1* driver-site mutation ($945 < \gamma \sim 1399 < 1999$ versus $281 < \gamma \sim 317 < 356$). Conversely, selection on driver-site mutations of *NOTCH2* was higher in the context of an extant *NOTCH1* mutation, but selection on *NOTCH1* driver-site mutations were trending lower in the context of extant *NOTCH2* mutations ($0 \leq \gamma \sim 0 < 1509$ versus $309 < \gamma \sim 346 < 386$). While the epistatic interaction between *NOTCH1* and *NOTCH2* was identified as significant overall ($P < 0.001$), this significance seems to be primarily driven by the substantial reduction in selection of *NOTCH2* driver-site mutations in the context of extant *NOTCH1* mutations (**Fig. 3A**), explaining the significance level reported for the opposite ordering despite visibly overlapping confidence intervals (**Fig. 3B**).

Analysis of epistatic interactions across the full evolutionary trajectory from esophageal organogenesis to tumor resection also revealed substantial synergistic selective epistasis of *NFE2L2* with mutant *TP53* ($P < 0.001$; **Fig. 3C**)—*NFE2L2* was substantially more strongly selected in the context of mutated *TP53* than in the context of wildtype *TP53* ($1894 < \gamma \sim 2382 < 2950$; versus $224 < \gamma \sim 333 < 472$)—resulting in higher observed co-occurrence than would be expected based on the calculated mutation rates of *NFE2L2* and *TP53* across the evolutionary trajectory. In contrast, a *TP53* mutant background leads to a reduction in selection on mutations in *NOTCH2* ($0 < \gamma \sim 73 < 367$; versus $444 < \gamma \sim 644 < 897$).

For *TP53* and *NOTCH2*, antagonistic epistasis was mutual and independent of the order of mutation, with selection for mutant *TP53* decreasing in the context of an extant *NOTCH2* driver-site mutation ($P < 0.001$; **Fig. 3D**). Conversely to the synergistic epistasis of mutations in *NFE2L2* and *TP53* when *TP53* was mutated first, the selection on mutations in *TP53* decreased to be selected against ($0 \leq \gamma \sim 0 < 1750$) in the context of an extant *NFE2L2* driver-site mutation compared to wildtype ($1929 < \gamma \sim 2048 < 2174$).

Discussion

Here we have developed a framework to quantify selection on somatic variants across sequential steps of cancer evolution, moving beyond static models to capture the dynamic, stepwise forces shaping tumorigenesis. Applying this model and a complementary analysis of pairwise epistasis to somatic DNA sequence data obtained from 430 adult clonal histologically normal human esophageal tissue samples and 1,741 primary esophageal squamous-cell carcinomas, we estimated step-specific scaled selection coefficients for mutations in key cancer-associated genes. These estimates reveal how selective pressures shift markedly from esophageal organogenesis to clonal expansion in normal tissue, and subsequently to malignant transformation. These results resolve a long-standing paradox: why mutations in canonical cancer drivers such as *NOTCH1* and *FAT1* are often more prevalent in premalignant human tissue than in human tumors [7,8]. Rather than manifesting a contradiction, this pattern reflects fundamental, step-specific changes in selection—where early mutations may confer a fitness advantage in normal epithelium, yet become neutral or even deleterious as tumorigenesis proceeds. These findings validate the crucial importance of modeling evolutionary context,

including premalignant states [54], to interpret the somatic mutation landscape [55,56] by illuminating the dynamic forces governing cancer development.

Martincorena et al. [8] first observed strong positive selection on *NOTCH1* mutations in human clonal histologically normal tissue, and speculated that such mutations might lose their selective advantage during malignant progression. Abby et al. [29] further investigated this idea in murine studies, suggesting that early *Notch1*-driven clonal expansions might suppress tumorigenesis in humans. Yet the somatic evolutionary dynamics underlying this paradox—where a putative driver mutation becomes less prevalent during malignant progression—have remained unresolved [24]. Our analysis provides direct evidence from human tissues that clarifies this discrepancy. We show that *NOTCH1* mutations are under strong positive selection during the transition from esophageal organogenesis to clonal histologically normal esophageal tissue, driving early clonal expansion, but are subsequently selected against during progression to ESCC.

This shift in somatic selection can be seen in even starker contrast when it is understood that it occurs despite an increased expected mutation rate as CHNE evolves into ESCC, relative to mutation expectations in the first step from organogenesis to CHNE tissue. Specifically, *NOTCH1* exhibits a low underlying mutation rate along the step from organogenesis to CHNE tissue and a higher rate during later oncogenesis. Accordingly, we show for the first time in humans that the higher prevalence of driver-site substitutions in *NOTCH1* in CHNE samples must be a consequence of strong positive selection in the first step of the evolutionary trajectory and selection against fixation in the subsequent step from CHNE tissue to ESCC. Indeed, despite a higher expected mutational burden in ESCC, the prevalence of *NOTCH1* substitutions in ESCC

samples is lower than in CHNE, providing strong evidence in contradiction of positive selection and likely indicating somatic selection *against* new *NOTCH1* mutations during progression from CHNE to ESCC. This striking reversal in selection across sequential steps during oncogenesis explains prior observations and reframes *NOTCH1* mutation not as a persistent driver of cancer, but as a context-dependent event that is advantageous in normal tissue, yet disadvantageous in the fitness landscape of tumorigenesis that can follow.

We identified strikingly similar patterns of mutation and selection in erstwhile ESCC driver genes *FAT1* and *NOTCH2 [33,57]*. Mutations of both genes are strongly selected along the step from organogenesis to CHNE tissue, yet newly arising mutations in them are selected against during the malignant transition to ESCC—a result that has not previously been reported. These results reinforce the concepts illustrated by *NOTCH1*, demonstrating mutations long considered canonical drivers of cancer may confer a selective advantage only in early, premalignant contexts. More broadly, they challenge the assumption that mutation prevalence directly reflects selective advantage. Instead they reveal how the interplay between mutation and selection is profoundly shaped by developmental timing and somatic genetic context.

Indeed, *NOTCH1* and *TP53* mutations are both common along the developmental trajectory from organogenesis through CHNE to ESCC—driven by moderate to high mutation rates and strong positive selection—yet their co-occurrence in the same tumor is strikingly rare. In the absence of mutational correlations or epistatic effects, one would expect their frequencies to combine proportionally. Their rarity of co-occurrence suggests an evolutionary constraint. In fact, prior experimental studies have demonstrated that interactions between NOTCH signaling and p53 are highly dynamic and depend on cellular context, varying with cell state and tissue

environment [58,59]. Our data-driven evolutionary modeling now provides direct evidence from human tumors that this molecular interaction is reflected in selective antagonistic epistasis: the presence of a *NOTCH1* mutation diminishes the selective advantage of acquiring a *TP53* mutation. This context-dependent suppression likely constrains malignant progression, as mutations of *TP53* are among the strongest-selected drivers of NOTCH1-wild-type tissue toward ESCC. Our results reveal that *NOTCH1* acts not only as a context-specific early driver but as a modulator of subsequent selection, weakening the fitness benefits of additional mutations in genes such as *PIK3CA*, *NFE2L2*, *RB1*, *FBXW7*, and *FAT1*. Significant antagonistic epistasis was present for *RB1* and *FAT1* ($P < 0.01$). This antagonism is consistent with *NOTCH1* mutations being protective against progression to cancer by disrupting the fitness landscape necessary for tumor progression. Yet once *TP53* mutations have occurred and substantially altered the cell state, subsequent mutations of *NOTCH1* may obtain new potential to contribute to cell proliferation and survival. This dependence of selection on cellular state underscores the importance of evolutionary context in interpreting effects of somatic mutations.

In addition to identifying antagonistic epistasis, we have also identified synergistic epistasis, most notably between *TP53* and *NFE2L2*. Their co-occurrence has previously been reported in ESCC [60] and non-small-cell lung cancer [61]. Our study further demonstrates that *TP53* mutations increase selective pressure for subsequent *NFE2L2* mutations along the trajectory towards ESCC. This finding aligns with experimental evidence from breast cancer cell lines, where NFE2L2 targets are upregulated or repressed through mutant p53, affecting cancer cell survival [62]. This concordance between our inferred epistatic directionality and experimentally identified molecular mechanisms supports the validity of our data-driven

evolutionary modeling. It also highlights its ability to capture and characterize functionally relevant gene-gene interactions in tumor evolution, and transforms a previously descriptive observation into a fundamental evolutionary insight: *TP53* and *NFE2L2* are not merely co-mutated due to common underlying mutagenic processes, but rather work together biologically to drive tumorigenesis in esophageal squamous epithelium.

The selective epistasis between *TP53* and *NFE2L2*, like many of the epistatic interactions we identified, is strikingly asymmetric—meaning that the selective advantage of one mutation depends on the presence of another, but not vice versa. For example, a *TP53* mutation enhances selection for *NFE2L2* mutations, whereas an *NFE2L2* mutation suppresses selection for *TP53* mutations. Such asymmetry may seem counterintuitive from a purely genetic perspective. However, this kind of asymmetry of epistasis has been both experimentally demonstrated [63] and conceptually grounded in evolutionary theory [64–67]. The presence of a driver substitution, such as in *NOTCH1* or *TP53*, will alter the cellular state and may also alter the local tissue microenvironment [68–74]. These alterations impose functional consequences, creating new positive and negative selective pressures that affect the fixation probability of subsequent mutations. These findings reinforce a central theme of this work: somatic evolution is not a linear combination of effects, but a context-dependent process shaped by the sequence and interaction of mutations. Accurate reconstruction of tumorigenesis requires not just identifying which mutations arise, but also understanding the conditionality of selection that governs when and how they are favored.

Interpreting somatic selection also requires consideration of the ecological context in which mutations arise. Both the normal tissues and the tumor tissues examined in this study were clonal

rather than heterogeneous. In biological settings where multiple selected clones coexist and compete, the probability of any clone achieving fixation is reduced due to clonal interference [75,76]. This competition lowers the probability of fixation for any competing clone, consequently reducing estimates of the strength of selection [77,78]. As a result, our own estimates of selection are likely to be conservative. However, this conservatism does not impact the relative strengths of selection we report across genes, as tumors are stochastically but equally subject to competitive tumor ecological dynamics. Exceptions would require unusual phenomena such as somatic variant clone cross-feeding or asymmetrical interclonal toxicity. Accordingly, clonal competition can be expected to reduce the frequencies of substitutions and draw down inferred estimates of the strength of selection, but should not compromise our ability to detect and quantify relative selective impact of mutations or to detect and quantify epistatic interactions between genes.

In addition to ecological context, our inferences are shaped by the resolution at which we model the action of selection. Specifically, we have assumed that putatively selected sites within each gene are subject to shared selective pressures and have shared epistatic relationships with selected sites in other genes. Our gene-specific selection estimates represent averages over a spectrum of selective effects among variants within the same gene [1,4,79]. If there were sufficient inferential power to perform it, site-specific quantification would highlight specific sites within genes with modest and outsized roles in driving tumorigenesis. However, our assumption of a single cancer effect across genes is common in cancer evolution studies [3,8,80], and is especially reasonable for genes such as tumor suppressors where inactivation of function is achieved by a range of mutations with similar phenotypic consequences. For key genes

analyzed here, including *TP53* and *NOTCH1*, incorporating variation in the strength of selection on selected sites is unlikely to alter the conclusions obtained here based on their joint estimates. In time, additional tissue sequencing will provide larger sample sizes that will empower studies to conduct measurement of independent and pairwise epistatic selective effects of individual variants, enabling fine-grained estimation of the architecture of mutational fitness landscapes in cancer evolution.

Based on gene-level modeling, we reported suggestive evidence of antagonistic epistasis between *NOTCH1* mutants and driver-site mutations of *PIK3CA*, *NFE2L2,* and *FBXW7* that was consistent with a broader pattern of *NOTCH1* mutations significantly attenuating selection on other, more frequent drivers. However, these interactions did not reach statistical significance in the current dataset ($P > 0.05$). This lack of statistical significance may derive from limited sample sizes, especially for early, histologically normal tissues that feature high levels of *NOTCH1* mutation. Sequencing data for normal esophageal epithelium and other histologically normal tissue samples are scarce, currently limiting higher-dimensional quantitative inference relating to early steps of oncogenesis. Larger datasets—especially those performing deep sequencing of normal epithelium—would provide crucial power for inference of additional epistatic interactions [cf. 81,82]. Indeed, larger datasets will empower not only inference of pairwise epistasis, but also estimation of three-way, four-way, and higher-dimensional epistatic interactions [46], revealing the fully complex, nonlinear and multidimensional fitness landscapes navigated by evolving somatic cells.

Our analyses were conducted on the selective impacts of somatic single-nucleotide mutations. However, a critical next step is the integration of copy-number alterations, which also

contribute substantially to the genetic evolution of cancer, and specifically ESCC [83], where copy-number alterations to cancer evolution have been assessed by tallying prevalence across a cohort. Ultimately, an increasingly comprehensive understanding of cancer evolution would arise from quantifying underlying rates of neutral copy-number alteration, and by integrating copy-number alterations into step-specific and epistatic selective models, yielding a more comprehensive understanding of how cancer genomes evolve under complex, dynamic selective pressures.

Our analysis has revealed the shifting selective landscape of somatic evolution from normal tissue to esophageal cancer, demonstrating the shaping of selection across cancer development subject to gene-specific epistatic interactions. These insights into *NOTCH1*, *TP53*, *NOTCH2*, *FAT1*, and *NFE2L2*, coupled with their associated epistatic interactions, underscore the importance of modeling early cancer progression as a distinct evolutionary phase, governed by selective pressures that differ fundamentally from those in advanced tumors. Crucially, the mutations most prevalent in ESCC may not represent the most effective targets for early detection or prevention. By quantifying step-specific selection and somatic genetic context-dependent epistasis, our methodology enables a precision oncology approach that is attuned to the timing and trajectory of tumor evolution. Improved early diagnosis and treatment have enormous potential to enhance survival, preserve quality of life, and reduce treatment burden. Understanding when mutations drive progression—and why they sometimes fail to—offers a precision-guided path toward clinical strategies that intervene with optimal timing and meaningful impact.

Conclusions

Our study provides new insight into the evolutionary dynamics of somatic mutations in esophageal tumorigenesis. We introduce a novel step-specific modeling approach that explicitly accounts for variation in underlying mutation rates across the progression from esophageal organogenesis through CHNE and subsequently to ESCC. This approach enables precise inference of selective pressures at discrete steps of tumor development. Applying this model to sequencing data from normal and tumor tissues of the human esophagus, we demonstrate that the underlying mutational processes and selective pressures on mutations vary substantially along the evolutionary trajectory from esophageal organogenesis to CHNE and further toward ESCC. We reveal for the first time that new mutations of *NOTCH1, NOTCH2,* and *FAT1* are strongly selected in CHNE tissue, but not during further progression to cancer, and that this shift is associated with antagonistic epistasis of *NOTCH1* and *NOTCH2* with *TP53* and *RB1*. These findings illuminate how somatic alterations can constrain rather than promote malignancy, offering crucial insight for precision oncology which must take location along the oncogenetic trajectory and somatic genetic context into account when investigating, designing, and testing translational therapeutics

List of abbreviations

CHNE: Clonal histologically normal esophageal

ESCC: Esophageal squamous-cell carcinoma

SNV: Single-nucleotide variant

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

No new datasets were generated for this study. The datasets supporting the conclusions of this article are available through cBioPortal (https://www.cbioportal.org/study/; ID escc_ucla_2014), Synapse (https://www.synapse.org/; ID syn33401857), and the supplementary materials of Martincorena et al. [8], Yokoyama et al. [7], Yuan et al. [30], and Liu et al. [32].

All analyses were conducted using a fully reproducible computational pipeline that is freely available at https://github.com/Cannataro-Lab/ESCC_step_epistasis and on Zenodo (https://doi.org/10.5281/zenodo.15659607), which utilized cancereffectsizeR v.2.10.2 [44], dependent on R (version ≥ 3.5.0).

Competing interests

The authors declare no competing interests.

Funding

This research was supported by the Elihu Endowment at Yale to JPT and by NIH NLM 5T15LM007056-32 and NIH NCI 1F31CA257288-01 to JNF. KAG was also supported by the Emmanuel College School of Science and Health.

Authors' contributions

Conceptualization: VLC and JPT; Data curation: KAG and VLC; Formal Analysis: KAG, VLC, JPT; Funding acquisition: VLC, JPT; Investigation: KAG, MJ, NF, VLC; Methodology: KAG, JDM, VLC, JPT; Project Administration: VLC, JPT; Resources: VLC, JPT; Software: KAG, JDM, VLC; Supervision: VLC, JPT; Validation: KAG, VLC; Visualization: KAG, JDM, VLC, JPT; Writing—original draft: KAG, MJ, VLC, JPT; Writing—review and editing: KAG, JDM, NF, VLC, JPT.

Acknowledgements

We thank Christopher Cross for helping to organize early efforts to start this project, Stephen Gaffney for arranging early data acquisition, Jose Costa for discussion of the topic, and Wes Lewis and Elizabeth Perry for helpful comments on an early draft.

References

1. Cannataro VL, Townsend JP. Neutral Theory and the Somatic Evolution of Cancer. Mol Biol Evol. 2018;35:1308–15.

2. Ostrow SL, Barshir R, DeGregori J, Yeger-Lotem E, Hershberg R. Cancer evolution is associated with pervasive positive selection on globally expressed genes. PLoS Genet. 2014;10:e1004239.

3. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal Patterns of Selection in Cancer and Somatic Tissues [Internet]. Cell. 2017. p. 1029–41.e21. Available from: http://dx.doi.org/10.1016/j.cell.2017.09.042

4. Cannataro VL, Gaffney SG, Townsend JP. Effect Sizes of Somatic Mutations in Cancer. J Natl Cancer Inst. 2018;110:1171–7.

5. Li L, Yue P, Song Q, Yen T-T, Asaka S, Wang T-L, et al. Genome-wide mutation analysis in precancerous lesions of endometrial carcinoma. J Pathol. 2021;253:119–28.

6. Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science. 2015;348:880–6.

7. Yokoyama A, Kakiuchi N, Yoshizato T, Nannya Y, Suzuki H, Takeuchi Y, et al. Age-related remodelling of oesophageal epithelia by mutated cancer drivers. Nature. 2019;565:312–7.

8. Martincorena I, Fowler JC, Wabik A, Lawson ARJ, Abascal F, Hall MWJ, et al. Somatic mutant clones colonize the human esophagus with age [Internet]. Science. 2018. p. 911–7. Available from: http://dx.doi.org/10.1126/science.aau3879

9. Moore L, Leongamornlert D, Coorens THH, Sanders MA, Ellis P, Dentro SC, et al. The mutational landscape of normal human endometrial epithelium. Nature. 2020;580:640–6.

10. Suda K, Nakaoka H, Yoshihara K, Ishiguro T, Tamura R, Mori Y, et al. Clonal Expansion and Diversification of Cancer-Associated Mutations in Endometriosis and Normal Endometrium. Cell Rep. 2018;24:1777–89.

11. Lee-Six H, Olafsson S, Ellis P, Osborne RJ, Sanders MA, Moore L, et al. The landscape of somatic mutation in normal colorectal epithelial cells. Nature. 2019;574:532–7.

12. Harris KL, Walia V, Gong B, McKim KL, Myers MB, Xu J, et al. Quantification of cancer driver mutations in human breast and lung DNA using targeted, error-corrected CarcSeq. Environ Mol Mutagen. 2020;61:872–89.

13. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature.

2013;499:214–8.

14. Martincorena I. Somatic mutation and clonal expansions in human tissues. Genome Med. 2019;11:35.

15. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med. 2014;371:2488–98.

16. Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med. 2014;371:2477–87.

17. Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science. 2015;348:880–6.

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

19. Majewski IJ. On fitness: how do mutations shape the biology of cancer? Biochem Soc Trans. 2019;47:559–69.

20. Rahal Z, Sinjab A, Wistuba II, Kadara H. Game of clones: Battles in the field of carcinogenesis. Pharmacol Ther. 2022;237:108251.

21. Dujon AM, Aktipis A, Alix-Panabières C, Amend SR, Boddy AM, Brown JS, et al. Identifying key questions in the ecology and evolution of cancer. Evol Appl. 2021;14:877–92.

22. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009;458:719–24.

23. Kato S, Lippman SM, Flaherty KT, Kurzrock R. The Conundrum of Genetic “Drivers” in Benign Conditions. J Natl Cancer Inst [Internet]. 2016;108. Available from: http://dx.doi.org/10.1093/jnci/djw036

24. Adashek JJ, Kato S, Lippman SM, Kurzrock R. The paradox of cancer genes in non-malignant conditions: implications for precision medicine. Genome Med. 2020;12:16.

25. Higa KC, DeGregori J. Decoy fitness peaks, tumor suppression, and aging. Aging Cell. 2019;18:e12938.

26. Hall MWJ, Shorthouse D, Alcraft R, Jones PH, Hall BA. Mutations observed in somatic evolution reveal underlying gene mechanisms. Commun Biol. 2023;6:753.

27. Colom B, Alcolea MP, Piedrafita G, Hall MWJ, Wabik A, Dentro SC, et al. Spatial competition shapes the dynamic mutational landscape of normal esophageal epithelium. Nat Genet. 2020;52:604–14.

28. Liao G, Tang J, Bai J. Early development of esophageal squamous cell cancer: Stem cells, cellular origins and early clone evolution. Cancer Lett. 2023;555:216047.

29. Abby E, Dentro SC, Hall MWJ, Fowler JC, Ong SH, Sood R, et al. Notch1 mutations drive clonal expansion in normal esophageal epithelium but impair tumor growth. Nat Genet. 2023;55:232–45.

30. Yuan W, Liu Z, Wang Y, Liu M, Pan Y, Lei W, et al. Clonal evolution of esophageal squamous cell carcinoma from normal mucosa to primary tumor and metastases. Carcinogenesis. 2019;40:1445–51.

31. Li M, Zhang Z, Wang Q, Yi Y, Li B. Integrated cohort of esophageal squamous cell cancer reveals genomic features underlying clinical characteristics. Nat Commun. 2022;13:1–15.

32. Liu X, Zhang M, Ying S, Zhang C, Lin R, Zheng J, et al. Genetic Alterations in Esophageal Tissues From Squamous Dysplasia to Carcinoma. Gastroenterology. 2017;153:166–77.

33. Lin D-C, Hao J-J, Nagata Y, Xu L, Shang L, Meng X, et al. Genomic and molecular characterization of esophageal squamous cell carcinoma. Nat Genet. 2014;46:467–73.

34. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2:401–4.

35. 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.

36. de Bruijn I, Kundra R, Mastrogiacomo B, Tran TN, Sikina L, Mazor T, et al. Analysis and Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE Biopharma Collaborative in cBioPortal. Cancer Res. 2023;83:3861–7.

37. Gao Y-B, Chen Z-L, Li J-G, Hu X-D, Shi X-J, Sun Z-M, et al. Genetic landscape of esophageal squamous cell carcinoma. Nat Genet. 2014;46:1097–102.

38. Sawada G, Niida A, Hirata H, Komatsu H, Uchi R, Shimamura T, et al. An Integrative Analysis to Identify Driver Genes in Esophageal Squamous Cell Carcinoma. PLoS One. 2015;10:e0139808.

39. Zhang Y, Zhang L, Wang R, Wang B, Hua P, Li J. LncRNA Erbb4-IR promotes esophageal squamous cell carcinoma (ESCC) by downregulating miR-145. J Cell Biochem. 2019;120:17566–72.

40. Mangalaparthi KK, Patel K, Khan AA, Manoharan M, Karunakaran C, Murugan S, et al. Mutational Landscape of Esophageal Squamous Cell Carcinoma in an Indian Cohort. Front Oncol. 2020;10:1457.

41. Wang L, Jia Y-M, Zuo J, Wang Y-D, Fan Z-S, Feng L, et al. Gene mutations of esophageal

squamous cell carcinoma based on next-generation sequencing. Chin Med J . 2021;134:708–15.

42. Zou B, Guo D, Kong P, Wang Y, Cheng X, Cui Y. Integrative Genomic Analyses of 1,145 Patient Samples Reveal New Biomarkers in Esophageal Squamous Cell Carcinoma. Front Mol Biosci. 2021;8:792779.

43. Cancer Genome Atlas Research Network, Analysis Working Group: Asan University, BC Cancer Agency, Brigham and Women's Hospital, Broad Institute, Brown University, et al. Integrated genomic characterization of oesophageal carcinoma. Nature. 2017;541:169–75.

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

45. Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, et al. OncoKB: A Precision Oncology Knowledge Base. JCO Precis Oncol [Internet]. 2017;2017. Available from: http://dx.doi.org/10.1200/PO.17.00011

46. Alfaro-Murillo JA, Townsend JP. Pairwise and higher-order epistatic effects among somatic cancer mutations across oncogenesis [Internet]. bioRxiv. 2022 [cited 2022 Jul 6]. p. 2022.01.20.477132. Available from: https://www.biorxiv.org/content/10.1101/2022.01.20.477132v1.abstract

47. Cannataro VL, Mandell JD, Townsend JP. Attribution of Cancer Origins to Endogenous, Exogenous, and Preventable Mutational Processes. Mol Biol Evol [Internet]. 2022;39. Available from: http://dx.doi.org/10.1093/molbev/msac084

48. Zhong F, Zhu T, Pan X, Zhang Y, Yang H, Wang X, et al. Comprehensive genomic profiling of high-grade serous ovarian carcinoma from Chinese patients identifies co-occurring mutations in the Ras/Raf pathway with TP53. Cancer Med. 2019;8:3928–35.

49. Remy E, Rebouissou S, Chaouiya C, Zinovyev A, Radvanyi F, Calzone L. A Modeling Approach to Explain Mutually Exclusive and Co-Occurring Genetic Alterations in Bladder Tumorigenesis. Cancer Res. 2015;75:4042–52.

50. Chang J, Zhao X, Wang Y, Liu T, Zhong C, Lao Y, et al. Genomic alterations driving precancerous to cancerous lesions in esophageal cancer development. Cancer Cell. 2023;41:2038–50.e5.

51. Ko K-P, Huang Y, Zhang S, Zou G, Kim B, Zhang J, et al. Key Genetic Determinants Driving Esophageal Squamous Cell Carcinoma Initiation and Immune Evasion. Gastroenterology. 2023;165:613–28.e20.

52. Song K, Ma C, Gu B, Wang B, Ma H, Deng X, et al. Molecular mechanism underlying epithelial-mesenchymal transformation and cisplatin resistance in esophageal squamous cell carcinoma. Thorac Cancer. 2023;14:3069–79.

53. Testa U, Castelli G, Pelosi E. The Molecular Characterization of Genetic Abnormalities in Esophageal Squamous Cell Carcinoma May Foster the Development of Targeted Therapies. Curr Oncol. 2023;30:610–40.

54. Crosby D, Bhatia S, Brindle KM, Coussens LM, Dive C, Emberton M, et al. Early detection of cancer. Science. 2022;375:eaay9040.

55. Bozic I. Quantification of the Selective Advantage of Driver Mutations Is Dependent on the Underlying Model and Stage of Tumor Evolution. Cancer Res. 2022;82:21–4.

56. Clarke MA, Woodhouse S, Piterman N, Hall BA, Fisher J. Using state space exploration to determine how gene regulatory networks constrain mutation order in cancer evolution. Computational Biology. Cham: Springer International Publishing; 2019. p. 133–53.

57. Wang J, Du J, Luo X, Guo L, Liu Y, Zhou J, et al. A platform of functional studies of ESCC-associated gene mutations identifies the roles of TGFBR2 in ESCC progression and metastasis. Cell Rep. 2024;43:114952.

58. Dotto GP. Crosstalk of Notch with p53 and p63 in cancer growth control. Nat Rev Cancer. 2009;9:587–95.

59. Kim SB, Chae GW, Lee J, Park J, Tak H, Chung JH, et al. Activated Notch1 interacts with p53 to inhibit its phosphorylation and transactivation. Cell Death Differ. 2007;14:982–91.

60. Jiang X, Zhou X, Yu X, Chen X, Hu X, Lu J, et al. High expression of nuclear NRF2 combined with NFE2L2 alterations predicts poor prognosis in esophageal squamous cell carcinoma patients. Mod Pathol. 2022;35:929–37.

61. Liu J, Liu S, Li D, Li H, Zhang F. Prevalence and Associations of Co-occurrence of Mutations and Chromosome 3q26 Amplification in Lung Cancer. Glob Med Genet. 2024;11:150–8.

62. Lisek K, Campaner E, Ciani Y, Walerych D, Del Sal G. Mutant p53 tunes the NRF2-dependent antioxidant response to support survival of cancer cells. Oncotarget. 2018;9:20508–23.

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

64. Weinreich DM, Watson RA, Chao L. Perspective: Sign epistasis and genetic costraint on evolutionary trajectories. Evolution. 2005;59:1165–74.

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

66. Zhang J, Chen F, Li X. Mechanistic causes of sign epistasis and its applications. Front Genet.

2024;15:1366917.

67. Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ. Empirical fitness landscapes reveal accessible evolutionary paths. Nature. 2007;445:383–6.

68. Shi Y, Ren X, Cao S, Chen X, Yuan B, Brasil da Costa FH, et al. gain-of-function mutation modulates the immunosuppressive microenvironment in non-HPV-associated oral squamous cell carcinoma. J Immunother Cancer [Internet]. 2023;11. Available from: http://dx.doi.org/10.1136/jitc-2023-006666

69. Chen X, Zhang T, Su W, Dou Z, Zhao D, Jin X, et al. Mutant p53 in cancer: from molecular mechanism to therapeutic modulation. Cell Death Dis. 2022;13:974.

70. Asl ER, Rostamzadeh D, Duijf PHG, Mafi S, Mansoori B, Barati S, et al. Mutant P53 in the formation and progression of the tumor microenvironment: Friend or foe. Life Sci. 2023;315:121361.

71. D'Assoro AB, Leon-Ferre R, Braune E-B, Lendahl U. Roles of Notch Signaling in the Tumor Microenvironment. Int J Mol Sci [Internet]. 2022;23. Available from: http://dx.doi.org/10.3390/ijms23116241

72. Li W, Ye L, Huang Y, Zhou F, Wu C, Wu F, et al. Characteristics of Notch signaling pathway and its correlation with immune microenvironment in SCLC. Lung Cancer. 2022;167:25–33.

73. Wang F, Long J, Li L, Zhao Z-B, Wei F, Yao Y, et al. Mutations in the notch signalling pathway are associated with enhanced anti-tumour immunity in colorectal cancer. J Cell Mol Med. 2020;24:12176–87.

74. Levine AJ. p53: 800 million years of evolution and 40 years of discovery. Nat Rev Cancer. 2020;20:471–80.

75. Greaves M, Maley CC. Clonal evolution in cancer. Nature. 2012;481:306–13.

76. de Visser JAGM, Rozen DE. Clonal interference and the periodic selection of new beneficial mutations in Escherichia coli. Genetics. 2006;172:2093–100.

77. Gerrish PJ, Lenski RE. The fate of competing beneficial mutations in an asexual population. Genetica. 1998;102-103:127–44.

78. Mon Père N, Terenzi F, Werner B. The dynamic fitness landscape of ageing haematopoiesis through clonal competition [Internet]. bioRxiv. 2024. Available from: http://biorxiv.org/lookup/doi/10.1101/2024.04.16.589764

79. Kennedy MC, Lowe SW. Mutant p53: it's not all one and the same. Cell Death Differ. 2022;29:983–7.

80. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of

somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016;534:47–54.

81. Armenia J, Wankowicz SAM, Liu D, Gao J, Kundra R, Reznik E, et al. The long tail of oncogenic drivers in prostate cancer. Nat Genet [Internet]. 2018; Available from: http://dx.doi.org/10.1038/s41588-018-0078-z

82. Cannataro VL, Townsend JP. Wagging the long tail of drivers of prostate cancer. PLoS Genet. 2019;15:e1007820.

83. Liang J, Wang Y, Cai L, Liu J, Yan J, Chen X, et al. Comparative Genomic Analysis Reveals Genetic Variations in Multiple Primary Esophageal Squamous Cell Carcinoma of Chinese Population. Front Oncol. 2022;12:868301.

Additional Files

Supplementary Table 1. Source and classifications of all normal and tumor samples included in this analysis.

[a] Study indicates the publication or dataset from which the sample was obtained for this analysis. This may differ from the original publication in which the sample was first described.

[b] Normal_or_Tumor indicates whether the sample was derived from clonal histologically normal esophageal epithelium (CHNE) or tumor tissue (ESCC).

[c] Coverage refers to the type of sequencing performed (i.e., targeted panel, whole-exome, or whole-genome).

[d] Notes provide the original source study for samples drawn from the integrated cohort assembled by Li et al. (2022).

Supplementary Table 2. Summary of sequencing targets and variant calling pipelines used across studies.

[a] Study, publication year, and PMID refer to the original source for sequenced samples.

[b] Samples indicate the number of normal and tumor samples processed in the cited study, as reported by the authors.

[c] Average sequencing depth refers to the reported median or mean sequencing coverage per sample, when available.

[d] Additional comments on methodology summarize study-specific quality filters, exclusion criteria, or other processing steps not captured in the preceding columns.