

# Impact of deleterious passenger mutations on cancer progression

Christopher D. McFarland<sup>a</sup>, Kirill S. Korolev<sup>b</sup>, Gregory V. Kryukov<sup>c,d</sup>, Shamil R. Sunyaev<sup>a,c,d</sup>, and Leonid A. Mirny<sup>a,b,c,e,1</sup>

<sup>a</sup>Graduate Program in Biophysics, Harvard University, Boston, MA 02115; <sup>b</sup>Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02129; <sup>c</sup>Broad Institute of MIT and Harvard, Cambridge, MA 02139; <sup>d</sup>Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115; and <sup>e</sup>Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited\* by Robert H. Austin, Princeton University, Princeton, NJ, and approved January 4, 2013 (received for review August 23, 2012)

**Cancer progression is driven by the accumulation of a small number of genetic alterations. However, these few *driver* alterations reside in a cancer genome alongside tens of thousands of additional mutations termed *passengers*. Passengers are widely believed to have no role in cancer, yet many passengers fall within protein-coding genes and other functional elements that can have potentially deleterious effects on cancer cells. Here we investigate the potential of moderately deleterious passengers to accumulate and alter the course of neoplastic progression. Our approach combines evolutionary simulations of cancer progression with an analysis of cancer sequencing data. From simulations, we find that passengers accumulate and largely evade natural selection during progression. Although individually weak, the collective burden of passengers alters the course of progression, leading to several oncological phenomena that are hard to explain with a traditional driver-centric view. We then tested the predictions of our model using cancer genomics data and confirmed that many passengers are likely damaging and have largely evaded negative selection. Finally, we use our model to explore cancer treatments that exploit the load of passengers by either (i) increasing the mutation rate or (ii) exacerbating their deleterious effects. Though both approaches lead to cancer regression, the latter is a more effective therapy. Our results suggest a unique framework for understanding cancer progression as a balance of driver and passenger mutations.**

chaperones | stochastic simulations | population genetics | unfolding response pathway

Recent advances in sequencing and genotyping of cancer tissues at a genome level have found that individual cancers contain tens of thousands of somatic alterations (1–4). These encompass many genetic alterations, such as single-nucleotide substitutions, insertions, deletions, rearrangements, Loss Of Heterozygosity (LOH) events, copy-number alterations, and whole-chromosome duplications/deletions (1); epigenetic alterations (5); and inheritable changes in cell state. It is generally believed that only a few (2–15) of these alterations cause the cancer phenotype, called driver alterations or simply drivers, whereas the overwhelming majority of events in cancer are believed to have nonsignificant phenotypes and are called passenger alterations or simply passengers. Drivers confer advantageous phenotypes to neoplastic cells (i.e., phenotypes that allow cells in the population to proliferate further). This property is inferred by their effect on cancer-related pathways; frequent occurrence at the same genes, loci, or pathways in different patients (3, 4, 6); and by the structure of cancer incidence rates (7). Because driver events are so critical to cancer progression, their discovery has been the primary goal of genome-wide cancer sequencing (8).

Conversely, little attention has been paid to passengers, which constitute the vast majority of observed somatic alterations in cancer (Table 1) (4). These alterations are assumed to be phenotypically neutral in cancer cells because they are nonrecurrent and are dispersed across a cancer genome (8, 9); however, their phenotype has never been systematically tested. If passengers arise as random alterations, then many can be deleterious to cancer cells (10–12), potentially via proteotoxic stress (13, 14),

loss of function (15), provoking an immune response (16), or other mechanisms. Though highly deleterious passengers are weeded out by negative selection, moderately deleterious passengers can evade negative selection and accumulate by mutation-selection balance, ratcheting, or similar mechanisms studied in population genetics (17). Because cancer genomes contain hundreds to thousands of accumulated protein-coding passengers, they may individually exert small effect, yet collectively be significant enough to alter the course of cancer progression.

Here we investigate the possible role of deleterious passenger alterations in cancer progression and examine their potential as an unexploited therapeutic target. First, using an evolutionary model, where passengers can arise alongside drivers in cancer cells, we find that moderately deleterious mutations evade purifying selection and accumulate. The accumulation of passengers alters the dynamics of cancer progression and may explain several clinical phenomena, such as slow progression, long periods of dormancy, the prevalence of small subclinical cancers, spontaneous regression, and heterogeneity in growth rates. These phenomenon cannot be easily explained without considering deleterious passengers. Unlike the current driver-centric paradigm of cancer progression, our analyses demonstrate that progression depends on drivers overcoming passengers. Second, we test the model's predictions by analyzing somatic mutations sequenced in cancers. This analysis shows that, in agreement with the model, individual passengers are likely to be damaging to cells and have largely evaded negative selection. Third, we use our model to explore two possible therapeutic approaches that target passengers and find that increasing either the mutation rate or the deleterious effect of passengers leads to cancer meltdown. The latter therapy may be possible by targeting pathways that buffer the effects of mutations, e.g., unfolded protein response (UPR) pathways. Finally, we present and discuss clinical and biological evidence that supports an important role of passenger alterations in cancer.

## Results and Discussion

### Evolutionary Model of Cancer Progression Incorporating Passengers.

Existing evolutionary models of cancer progression have several limitations. Many models have considered a population of a constant or externally controlled size (18, 19), which does not depend on the absolute fitness of cells. Other models study exponentially growing cancer populations (7, 19, 20), whereas logistic-like behavior has been observed in cancer (21). Most importantly, the vast majority of cancer models (with the exception of ref. 22; see below) neglect the effects of passenger alterations.

Author contributions: C.D.M., K.S.K., G.V.K., S.R.S., and L.A.M. designed research; C.D.M., K.S.K., and L.A.M. performed research; C.D.M., K.S.K., and G.V.K. analyzed data; and C.D.M., K.S.K., and L.A.M. wrote the paper.

The authors declare no conflict of interest.

\*This Direct Submission article had a prearranged editor.

<sup>1</sup>To whom correspondence should be addressed. E-mail: leonid@mit.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1213968110/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1213968110/-DCSupplemental).

**Table 1. Passenger mutations in whole-genome sequences**

Cancer(s)	Protein-coding mutations	Driver mutations*	Ref(s).
11 breast	115.4	5.1	(4)
10 colon	75	4	(4)
4 astrocytomas	206	5.5	(52)
Acute myeloid leukemia	10	2	(53)
26 melanomas	366	4	(1, 32)
Small-cell lung	100	4	(2)

In most tumors, hundreds of protein-coding mutations accrue, yet only a few are putative drivers. These values are consistent with our model's results. Deleterious passengers may be most exploitable in carcinomas, because leukemia and many blood cancers are generally more sensitive to DNA damage and have earlier incidence rates.

\*Classified as drivers by COSMIC (8).

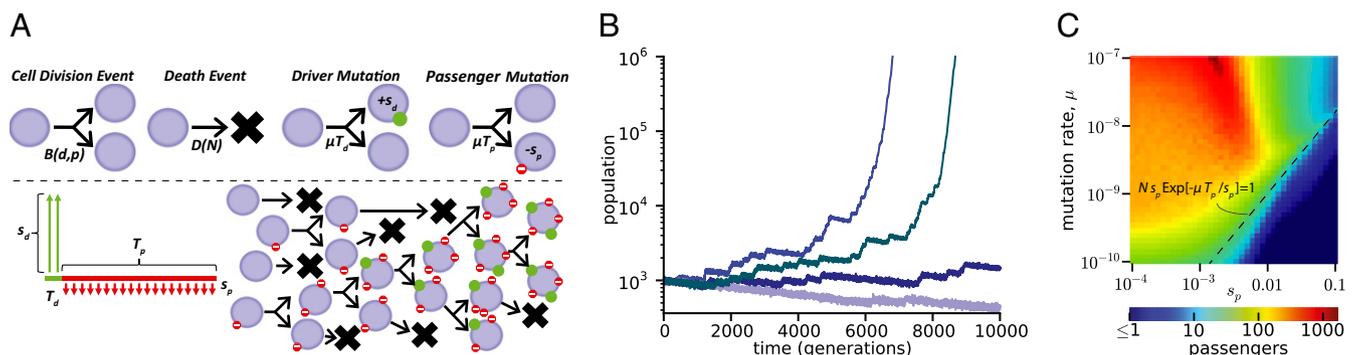
In our stochastic model, individual cells can divide, potentially acquiring driver or passenger alterations, and die. Population size changes with the birth and death of individual cells (Fig. 1A). Generally, the birth and death rates of a cell depend on the effect of accumulated drivers and passengers, and the environment. Assuming that all drivers/passengers possess equal fitness advantage/disadvantage, the birth and death rates  $B(d,p,N)$  and  $D(d,p,N)$  of each cell depend on the number of drivers  $d$ , the number of passengers  $p$ , and the total hyperplasia or population size  $N$ . Driver mutations increase population size by either increasing the birth rate (e.g., an activating mutation in *KRAS*) or by decreasing the death rate [e.g., a *TP53* knockout that diminishes contact inhibition (23) and apoptosis]. Though specific drivers and passengers will have differing effects on the birth and death rates, we find that aggregating the effects of mutations into the birth rate, and placing the effects of population size into the death rate, does not alter population dynamics from models where mutational effects are split between the two (SI Appendix, Fig. S1). Thus, we use

$$B(d,p) = \frac{(1+s_d)^d}{(1+s_p)^p} \quad D(N) = \frac{N}{K} \quad [1]$$

where  $s_d$  is the fitness advantage (selection coefficient) of a driver,  $s_p$  is the fitness disadvantage conferred by a passenger, and  $K$  is the initial equilibrium population size—reflecting the effects of the tumor microenvironment. This model assumes multiplicative epistasis and is equivalent or similar to other possible forms (SI Appendix, SI Text), which all exhibit qualitatively similar behavior

(SI Appendix, Fig. S1). We also let  $D(N) = \log(1 + N/K)$ , for large cancers (grown to  $10^6$  cells). For small  $N/K$  this reduces to the linear model above [similar to previous neoplastic (24) and ecological (25) models], but for large  $N/K$  this recapitulates Gompertzian dynamics observed experimentally for large tumors (26). The death rate's dependence on population size is a coarse approximation of many size-dependent factors that tumors must overcome as they expand via additional drivers: contact inhibition, competition between cells for space and resources (e.g., due to a limited crypt size), homeostatic pressure, hypoxia, angiogenesis, limited paracrine signaling, and immune/inflammatory responses to larger tumors (16).

We model cancer progression as a stochastic system of birth (with or without mutations) and death events with defined reaction rates using a standard Gillespie algorithm (27). The system is fully defined by five parameters:  $s_p$ ,  $s_d$ ,  $\mu T_p$ ,  $\mu T_d$ , and  $K$ , where  $\mu$  is the mutation rate and  $T_{d/p}$  are the mutation target sizes for drivers/passengers. Though driver and passenger alterations take many forms, we parameterized our model using single-nucleotide substitution data, as these mutations have been more thoroughly quantified. Because of extensive cancer heterogeneity and limited quantitative knowledge, we varied all parameters by 2–3 orders of magnitude. The ranges we explored centered on values obtained from the literature (SI Appendix, Table S1). The mutation rate ( $\mu \sim 10^{-8}$  nt $^{-1} \times$  division $^{-1}$ ; range  $10^{-10}$ – $10^{-6}$ ) approximates cells with a mutator phenotype (28). Our initial equilibrium population size ( $K \sim 10^3$  cells; range  $10^2$ – $10^4$ ) was estimated from hyperplasias within a mouse colonic crypt observed 2 wk after an initiating *APC* deletion (29). The target size for drivers ( $T_d \sim 700$  nt; range 70–7,000) approximately 10 potential hotspot mutations per gene (oncogene or tumor suppressor) times 70 driver genes (4). This value was used in previous simulations (19) and is close to the 571 loci with recurrent mutations in colon cancer (30). The target size for functional (nonsynonymous) passengers ( $T_p \sim 5 \times 10^6$  nt; range  $5 \times 10^5$ – $5 \times 10^7$ ) was estimated as  $10^3$  nonsynonymous loci per gene times 5,000 well-expressed, non-cancer-related genes in cancer (9). This value is comparable, but less than, a previous estimate of 10 million deleterious loci in cancer (31); does not attempt to capture the  $10^4$ – $10^5$  noncoding passenger mutations per cancer genome (2, 32); and yet is thousands of times greater than  $T_d$ . The chosen driver strength [ $s_d \sim 0.1$  (i.e., 10% growth increase per driver); range 0.01–1] was shown to be congruent with cancer onset (19). Passenger deleteriousness ( $s_p \sim 10^{-3}$ ; range  $10^{-1}$ – $10^{-4}$ ) was estimated from the effects of near-neutral germ-line mutations in humans (33) and randomly introduced mutations in yeast (14). Simulations where drivers or passengers conferred a distribution of  $s_p$  and  $s_d$  did not significantly differ from our fixed-effect



**Fig. 1. Dynamics of cancer progression.** (A) Our evolutionary model: individual cancer cells stochastically divide (potentially acquiring new drivers/passengers) and die. A new driver increases the birth rate by  $s_d$ , whereas a passenger decreases it by  $s_p$  (Eq. 1). Drivers arise rarely, but have large effects, while passengers are common, but have small individual effects. (B) Simulated cancer progression using a Gompertz death rate; despite identical parameters, trajectories exhibit markedly different behavior, sometimes regressing to extinction or having long periods of dormancy. (C) The number of accumulated passengers increases with mutation rate and depends, nonmonotonically, on passenger strength.

model (*SI Appendix*, Fig. S1 and discussed below), suggesting that this fixed-effect model adequately captures cancer dynamics. We find that deleterious passengers accumulate under a broad range of conditions (*SI Appendix*, Fig. S2).

We consider death to be any process that prevents a cell from replicating indefinitely, i.e., necrosis, apoptosis, senescence, or differentiation. Thus,  $N$  represents only cells capable of infinite division and of carrying the (epi)genetic information in cancer. For this reason, our model lacks asymmetric cell divisions, as this yields differentiated cells. Because we explored the initial population size across two orders of magnitude, our model applies equally well to tumor subtypes dominated by only a small cohort of cancer stem cells and subtypes where cancer may arise from progenitor cells (34). Our model ignores the spatial structure of cancer. Previous studies of asexual populations suggest that ignoring spatial structure will (i) underestimate the time for beneficial drivers to sweep through the population and hence the degree of clonal interference, and (ii) overestimate the effective strength of selection, which only acts at the geographic boundary between clones (35, 36). Hence, models considering spatial structure should find that more passengers fixate relative to those that do not, strengthening the conclusions of our model.

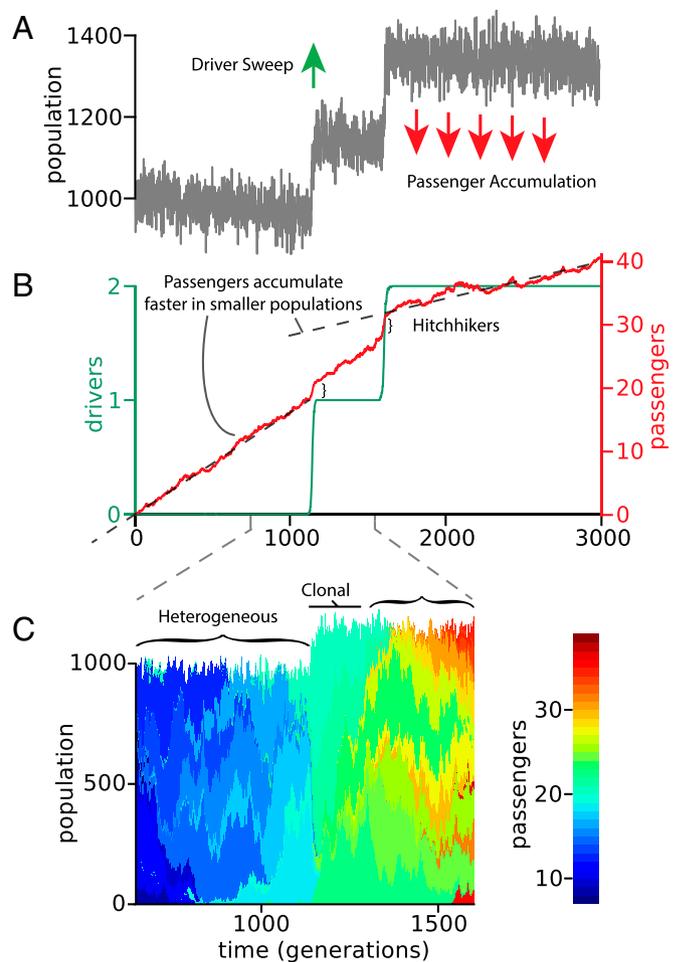
#### Moderately Deleterious Passengers Fixate and Alter Cancer Progression.

Fig. 1*B* presents typical population trajectories of cancer beginning at the first driver mutation. All trajectories consist of intervals of rapid growth and gradual decline. A new driver leads to a clonal expansion of the subpopulation carrying this driver, causing short periods of rapid growth. Growth stops when the effect of this driver is balanced by the death rate, which increases with population size. While the population waits for the next driver to arise, passengers steadily accumulate, causing a gradual decline of population size. Together, these processes cause trajectories to grow in a sawtooth pattern.

Simulated tumors exhibit either unconstrained growth or regression, often after a period of dormancy (Fig. 1*B*). We find that the probability of either outcome depends on the tumor size: tumors larger than a critical size ( $N_{critical}$ ) are likely to progress, whereas smaller tumors are likely to regress (*SI Appendix*, Fig. S3). Indeed, larger populations acquire drivers more frequently, as they have more cells in which drivers can arise. Moreover, natural selection weeds out deleterious mutations more efficiently in larger populations (Fig. 2*B*). The phenomena of dormancy and spontaneous regression, observed both in our model and clinically (37), do not occur in models lacking deleterious passengers. In *SI Appendix*, *SI Text*, we estimate  $N_{critical}$  for cancer and provide a framework for understanding where deleterious passengers are most relevant (*SI Appendix*, Fig. S4).

Importantly, simulations show that hyperplasias that progress to clinical size (i.e.,  $10^6$  cells, 15–20 drivers) accumulate many deleterious passengers. Evasion of purifying selection and fixation of deleterious passengers is an unexpected result not programmed into the model. Although the exact number of accumulated passengers depends on  $\mu$  and  $s_p$  (Fig. 1*C*),  $10^2$ – $10^3$  deleterious passengers are obtained for a broad range of parameters, consistent with the numbers of nonsynonymous substitutions observed in cancer genomics studies (Table 1), suggesting that observed passengers in sequencing data can be moderately deleterious.

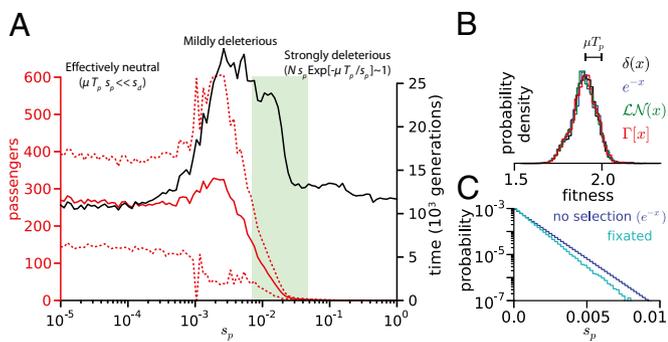
We then studied how deleterious mutations can accumulate despite negative selection. Previous studies have calculated the rate of accumulation of deleterious mutations in the absence of clonal expansions (38–40). We identified two previously known processes that allow passengers to evade negative selection in cancer: hitchhiking alongside a driver and Muller's ratchet (25) (Fig. 2). Deleterious passengers hitchhike when the cell they reside in acquires a new driver, which then leads to a clonal expansion and fixation of all the mutations in that cell. Muller's ratchet, in turn, is a process of gradual accumulation of deleterious mutations and population decline in the absence of drivers. In Muller's ratchet, a mutation-selection balance arises after driver sweeps, which creates a steady-state Poisson distribution of the



**Fig. 2.** Mechanisms of passenger accumulation. (A) Spurts of population growth, caused by the acquisition of a new driver, are interspersed with a gradual decline due to passenger accumulation. (B) Passengers accumulate both steadily between the arrival of drivers and by hitchhiking during clonal expansions. (C) Each subclone, containing a unique number of passengers (shown by color), grows and declines stochastically, eventually to extinction. In between drivers, the population becomes heterogeneous. A new driver will promote only one clone, creating a clonal population. Afterward, new mutations on top of the previous hitchhikers restore heterogeneity.

number of passengers per cell with mean and variance  $\mu T_p/s_p$  (first described in ref. 41; *SI Appendix*, *SI Text* and Fig. S5). The fittest subpopulation—cells with the fewest passengers:  $\sim N \text{Exp}[-\mu T_p/s_p]$  cells—is much smaller than the whole population, so it can spontaneously shrink to extinction (Fig. 2*C*). When back mutations are rare, such an extinction leads to the irreversible loss of this least-mutated fraction of cell and corresponds to a “click” of Muller's ratchet (25). This process is especially rapid during clonal expansions when the size of the expanding clone is small. Both of the above processes, well known in population genetics, are augmented in cancer because of the presence of strong drivers.

Simulations show that moderately deleterious, rather than highly deleterious or neutral, passengers have a major effect on cancer progression (Fig. 3*A*). Indeed, almost-neutral passengers have very little effect on cancer cells, and passengers of large effect do not accumulate (31). By slowing progression to cancer, moderately deleterious passengers accumulate in even greater numbers than neutral mutations despite their slower accumulation rate (Fig. 3*A*). Importantly, we find that moderately deleterious passengers affect progression for  $s_p$  from  $10^{-3}$  to  $2 \times 10^{-2}$ , which subsumes the best literature estimates of the strength of



**Fig. 3.** Moderately deleterious passengers alter cancer progression and mostly evade selection. (A) Passengers of intermediate fitness effect  $s_p$  prolong the time to cancer and accumulate in large, highly variable quantities (red solid, mean; dotted,  $\pm 1$  SD). Moderately deleterious passengers affect cancer only if they are strong or frequent enough to be comparable to the effects of drivers, yet weak enough to avoid selection (SI Appendix, SI Text). Experimentally observed fitness effects of random point mutations in YFP in yeast ranged from 0.007 to 0.028 (green shading) (14). (B) Population dynamics did not change noticeably when passengers were drawn from various distributions of fitness distributions (SI Appendix, SI Text). (C) Passenger fixation probability declined only moderately with increasing deleteriousness.

deleterious mutations (14, 33). Such small selection coefficients for individual passengers are typically undetectable in cell cultures, yet critical for long-term cancer dynamics.

We then relaxed our assumption that  $s_p$  is constant for all passengers, by simulating cancer progression with passengers drawn from distributions of deleteriousness (SI Appendix, Fig. S6). The strength of driver and passenger mutations affects their fixation probability (Fig. 3B and C). For passengers, the variation in fitness within a population is mostly invariant to the type of distribution of passenger effects (Fig. 3B). Negative selection against passenger fixation appears to be largely inefficient, except for highly deleterious passengers (Fig. 3C).

The significant variance in cell fitness within the population, caused by deleterious passengers (Fig. 3B), also affects the probability of driver fixation. Because a driver will generally occur in a cell of average fitness, it is unlikely to fixate unless its new fitness is greater than the fittest cells. The difference between the fittest cells and average cells in the population is approximately  $\mu T_p$  and independent of  $s_p$  (Fig. 3B) (17); therefore, a driver must confer a benefit greater than  $\mu T_p$  to fixate (SI Appendix, Fig. S1). This argues that weak drivers are unlikely to fixate in cancer or be observed in genomic sequencing.

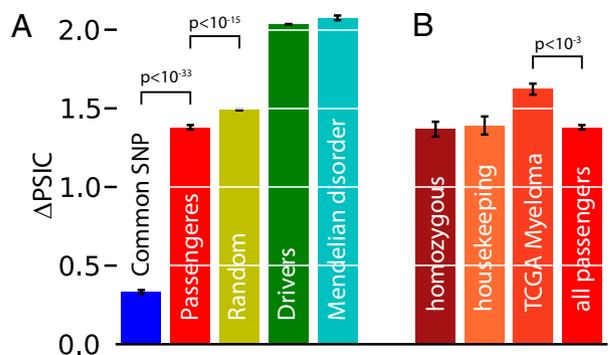
In summary, our simulations demonstrate that despite the moderately deleterious effect of individual passengers, they accumulate in large numbers during neoplastic progression, reducing the fitness of cancer cells and altering the course of neoplastic progression. We find several reasons why deleterious passengers accumulate more than might be expected a priori: (i) mutator phenotypes [a hallmark of cancer (28)] accelerate accumulation rates; (ii) small population sizes in the early stages of cancer progression enhance accumulation rates; (iii) driver-induced bottlenecks and hitchhiking contribute additional passengers; (iv) passengers prolong progression—offering more time for accumulation; and (v) passengers arising as part of a distribution of deleteriousness fixate more often than equivalent passengers considered in isolation. These first three phenomenon, though undocumented in cancer theory, have been previously observed in population genetics (12).

**Passenger Mutations Observed in Cancer Can Be Damaging.** Our model makes several testable predictions: (i) accumulated passengers in cancer populations can be deleterious to cancer cells; (ii) the deleterious effect of an individual passenger has little bearing on its likelihood of accumulation; and (iii) fixed

drivers should have larger effects on phenotype than passengers. Cancer genomics data provide an opportunity to test these predictions. First, we test whether nonsynonymous passengers found in cancer are damaging or neutral to protein function using comparative genomics. Second, we test whether selection acting against passengers is effective at preventing fixation or largely ineffective, as suggested by our simulations.

We analyzed 116,977 cancer mutations curated by the Catalogue of Somatic Mutations in Cancer (COSMIC) and The Cancer Genome Atlas (TCGA). We classified them as driver and passenger mutation groups and then characterized their effects using PolyPhen, a tool widely used in population and medical genetics to predict the damaging effect of missense mutations (15). Passengers were identified as missense mutations that show no recurrence and affect genes not listed in a census of possible cancer-causing genes (SI Appendix, SI Text). The  $\Delta PSIC$  metric of PolyPhen measures the degree of evolutionary conservation of a mutated residue (42) by calculating the negative log-likelihood of observing a specific mutation, given the evolutionary history of the protein. Specifically, a mutation with a  $\Delta PSIC$  of 1 is  $e$  ( $= 2.71 \dots$ ) times less likely to be observed than the wild-type allele, as computed from a multiple alignment. Thus, a mutation with high  $\Delta PSIC$  is more likely to be damaging to molecular function (43) because this implies the mutation disrupts a well-conserved residue. PolyPhen has been extensively tested and benchmarked (15).

Fig. 4 presents this analysis for passengers, drivers, and three reference datasets: (i) common human missense SNPs; (ii) simulated de novo mutations (randomly generated using a cancer-specific three-parameter model; SI Appendix, SI Text); and (iii) damaging, pathogenic missense mutations causing human Mendelian diseases (from the Human Gene Mutation Database). As expected, common SNPs are benign and exhibit small  $\Delta PSIC$  values, whereas disease-causing mutations, with known damaging effect, exhibit large  $\Delta PSIC$  values (Fig. 4A). Driver mutations exhibit similarly high values of  $\Delta PSIC$ , significantly greater than randomly generated mutations, indicating that drivers tend to occur at well-conserved loci. From a biochemical perspective, this result shows that, to activate an oncogene or to disable a tumor suppressor, the driver mutation must change a critical and well-conserved residue, e.g., the GTP binding site of Ras or DNA binding domain of p53. From an evolutionary perspective, the conservation of residues that promote tumorigenesis when mutated suggests strong natural selection against the early development of cancer. The ability of  $\Delta PSIC$  score to identify drivers as having highly nonneutral phenotypes (i.e., damaging or altering



**Fig. 4.** Characterization of missense mutations in cancer sequencing data. (A) Mutations were assayed using the  $\Delta PSIC$  score of PolyPhen, which estimates the damaging effect of a new mutation, given known homologs; mutations with high  $\Delta PSIC$  scores are most likely damaging (43). Passengers have large  $\Delta PSIC$ , close to random mutations, suggesting that they are deleterious. (B) Deleterious passenger phenotypes were observed in all subsets of passengers studied, arguing that these results cannot be explained by recessive phenotypes, or lack of expression, or database biases.

molecular function) validates its use for characterizing somatic cancer mutations. The exceptionally high  $\Delta PSIC$  scores for these mutations are consistent with our third prediction that drivers must be of strong effect.

Most importantly, passenger mutations exhibit  $\Delta PSIC$  values that are on average much greater than neutral mutations (Fig. 4A;  $P < 10^{-33}$ ); therefore, many passengers affect conserved residues and are likely damaging to protein function. This result clearly demonstrates that passenger mutations are nonneutral. To ensure that our set of putative passenger mutations was not contaminated by drivers, we increased our stringency of passenger classification, but found no statistically significant change ( $P = 0.69$ ) in mean  $\Delta PSIC$  (SI Appendix, SI Text); additional safeguards are explored below.

Passengers exhibit  $\Delta PSIC$  values much lower than drivers (Fig. 3A), supporting the assumption of our evolutionary model that deleterious passengers are generally much weaker than drivers ( $s_p \ll s_d$ ). The  $\Delta PSIC$  values of passengers are close to, but lower than, values of randomly generated mutations (Fig. 3A;  $P < 10^{-15}$ ), suggesting that many passenger mutations evade purifying selection. Still, a statistically significant difference between these two sets demonstrates slight negative selection against the most deleterious passengers. This comparison of passengers and random mutations fully supports our model's prediction that selection against moderately deleterious passengers is largely ineffective in neoplastic progression (Fig. 3C).

To rule out possible caveats where passengers have damaging effects on protein function but no effect on the fitness of cancer cells, we performed additional tests. For example, passengers with deleterious scores could affect only genes that are functionally unimportant or not expressed in cancer cells. Thus, we considered only passengers in essential and ubiquitously expressed housekeeping genes, but still observe equally high  $\Delta PSIC$  scores (Fig. 4B). This eliminates the possibility that damaging passengers are not expressed or present in unimportant genes. Alternatively, perhaps only recessive heterozygous passengers exhibit high  $\Delta PSIC$  scores; if so, cell fitness would remain unchanged because the other allele provides sufficient functionality. We observe equally high  $\Delta PSIC$  scores for homozygous passengers (which can arise via LOH events or chromosomal losses), refuting this possibility (Fig. 4B). Collectively, our analyses show that signatures of damaging mutations are ubiquitous in known passengers and likely affect the fitness of cancerous cells.

As an alternative powerful test, we assayed for signatures of selection in driver and passenger genes by comparing the observed ratio of nonsynonymous to synonymous mutations ( $\omega$ ) to the predicted ratio using a random model of mutations (SI Appendix, Fig. S7). This distribution reaffirmed COSMIC's driver and passenger classifications. Genes with  $\omega < 1$  likely experience purifying selection and these genes were generally classified as passengers by COSMIC. Conversely, genes with  $\omega > 1$  likely experience positive selection and were nearly all classified as drivers by COSMIC. Most importantly, the shape of this distribution corroborates the narrative of a few strong drivers overlaid with copious passengers experiencing nearly undetectable negative selection that we observe in both our modeling and  $\Delta PSIC$  analysis: A total of 94% of genes had an observed  $\omega < 1$ , and their occurrence was only very moderately enriched relative to our neutral model—on the fringe of statistical significance:  $P = 0.012$ —and not nearly as pronounced as the signal for drivers. The rare driver genes, with  $\omega > 1$ , often exhibited extreme nonsynonymous substitution rates vastly greater than expected from a neutral model of evolution:  $\omega$  of *KRAS*, *TP53*, *BRAF*, and *PTEN* were all greater than 40.

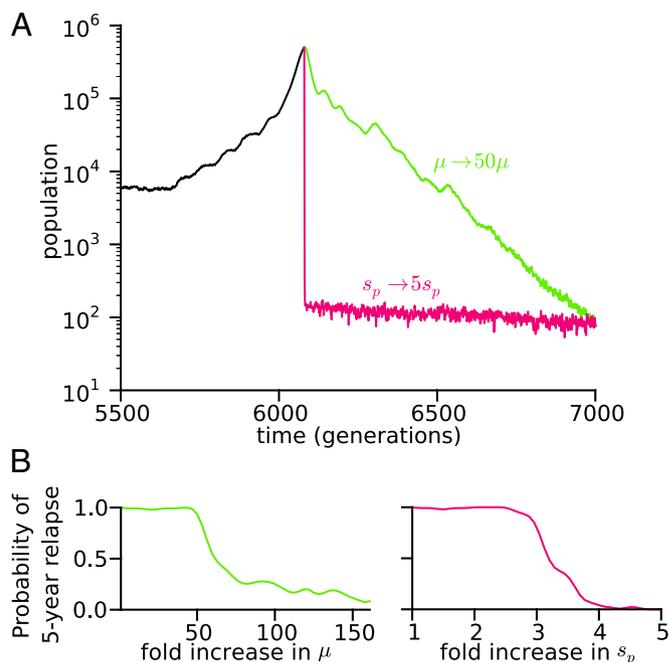
Though our genomic analysis of passenger mutations focused on missense substitutions, our model is generalizable to all inheritable (epi)genetic alterations, including those that are present at low frequency in the cancer population. Indeed, the length distribution of somatic copy number alterations (SCNAs) in cancer suggests these alterations are under purifying selection as well (44). Hence, the total load of accumulated deleterious

passengers in cancer may be greater than that estimated from single nucleotide mutations detected in genome sequencing.

**Accumulated Passenger Mutations Can Be Exploited for Cancer Treatment.** Using our evolutionary model, we probed how cancers that accumulated passenger alterations respond to passenger-centric treatments. We tested two strategies: (i) increasing the overall mutation rate ( $\mu$ ), thus increasing the rate of passenger accumulation, and (ii) magnifying the deleterious effect of passengers ( $s_p$ ), as described below. Both strategies reduce cancer size (Fig. 5); however, mutagenic strategies require more severe increases ( $\sim 50$ -fold) in the mutation rate to succeed (Fig. 5A), whereas fivefold magnifications of deleterious effect suffice. Even with large mutation rate increases, the probability of 5-y relapse following treatment initiation is considerable (Fig. 5B). This behavior resembles patient responses to existing chemotherapeutic agents that elevate mutation rates.

In practice, increasing the deleterious effect of passengers, both mutations and chromosomal alterations (45), could be achieved by inhibiting cellular mechanisms that buffer against the effects of mutations or incorrect protein dosage (14). Hence, deleterious effect could be increased by targeting chaperones, proteasomes, or other components of UPR pathways (46); or by elevating ER stress (47); or by stimulating protein misfolding using hyperthermia (48). These passenger-mediated therapies should specifically affect cancer cells because somatic mutations are generally rarer in normal tissues (9). For example, a recent study of clonal mosaicism in human brains found only 1.5 SCNAs per adult sample (49), whereas a recent pan-cancer survey found 42 SCNAs per cancer (3).

Several experiments support this strategy of exacerbating passengers' effect. First, chaperones are widely expressed in cancer, indicative of poor prognosis (50), and their inhibition (or proteasome inhibition) exhibits antitumor activity (46). Though other specific roles of chaperones and proteasomes in cancer were proposed, our framework suggests that cancers buffer against the effects of passenger alterations using UPR machinery. In our paradigm, inhibiting the UPR unleashes the effects of accumulated



**Fig. 5.** Deleterious passengers can be exploited for treatment. (A) Cancers grown to  $10^6$  cells are treated by increasing the mutation rate (green) or deleterious effect of passengers (magenta). Both strategies lead to reduction in cancer size. (B) Much smaller increase of the deleterious effect of passengers is sufficient to prevent 5-y relapse.

passengers. Recent discoveries that aneuploidy and chromosomal imbalance lead to proteotoxic stress (45) and a dependence on the UPR for survival (47), and that very high levels of DNA damage correlate with better clinical outcomes (43, 51) (a paradoxical result in the classical paradigm of cancer), are consistent with our framework.

One of the major limitations of driver-targeted therapies is that they can be defeated by a cancer's ability to rapidly evolve resistance by acquiring new mutations. Our approach, of increasing the deleterious effects of passengers, is different as it targets not only existing cancer cells, but also cancer's ability to accumulate new mutations and thus its evolvability.

Cancer research has focused primarily on driver alterations with little attention to the overwhelming majority of potentially harmful passenger alterations that arise along the way. To the best of our knowledge, this "dark matter" of cancer genomes has not yet been explored. We developed an evolutionary model of cancer progression that clearly demonstrates that deleterious passengers can accumulate in cancer, while our genomic analysis confirms that passengers presented in sequenced cancers have

damaging phenotypes. Importantly, considering cancers as a balance between drivers and deleterious passengers reproduces many observed phenomena in cancer, including (*i*) slow initial and rapid late growth; (*ii*) a critical cancer size for dormancy or spontaneous regression; and (*iii*) short-term response to mutagenic therapies (*SI Appendix, Table S2*). These phenomena were not preprogrammed into the model, suggesting that the deleterious effect of passengers explains many properties of cancers.

## Materials and Methods

Simulations were executed using the Next Reaction (27), a Gillespie algorithm. All cancer mutations were collected from COSMIC version 42 (30). See *SI Appendix, SI Text* for details.

**ACKNOWLEDGMENTS.** The authors thank M. Imakaev and G. Fudenberg for many productive discussions and comments on the manuscript. This work was supported by the National Institutes of Health/National Cancer Institute Physical Sciences Oncology Center at Massachusetts Institutes of Technology Grant U54CA143874.

- Pleasant ED, et al. (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 463(7278):191–196.
- Pleasant ED, et al. (2010) A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 463(7278):184–190.
- Beroukhi R, et al. (2010) The landscape of somatic copy-number alteration across human cancers. *Nature* 463(7283):899–905.
- Sjöblom T, et al. (2006) The consensus coding sequences of human breast and colorectal cancers. *Science* 314(5797):268–274.
- Sharma S, Kelly TK, Jones PA (2010) Epigenetics in cancer. *Carcinogenesis* 31(1):27–36.
- Maley CC, et al. (2004) Selectively advantageous mutations and hitchhikers in neoplasms: p16 lesions are selected in Barrett's esophagus. *Cancer Res* 64(10):3414–3427.
- Haeno H, Iwasa Y, Michor F (2007) The evolution of two mutations during clonal expansion. *Genetics* 177(4):2209–2221.
- Futreal PA, et al. (2004) A census of human cancer genes. *Nat Rev Cancer* 4(3):177–183.
- Chapman MA, et al. (2011) Initial genome sequencing and analysis of multiple myeloma. *Nature* 471(7339):467–472.
- Moran NA (1996) Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci USA* 93(7):2873–2878.
- Walczak AM, Nicolaisen LE, Plotkin JB, Desai MM (2012) The structure of genealogies in the presence of purifying selection: A fitness-class coalescent. *Genetics* 190(2):753–779.
- Lynch M, Bürger R, Butcher D, Gabriel W (1993) The mutational meltdown in asexual populations. *J Hered* 84(5):339–344.
- Neznanov N, Komarov AP, Neznanova L, Stanhope-Baker P, Gudkov AV (2011) Proteotoxic stress targeted therapy (PSTT): Induction of protein misfolding enhances the antitumor effect of the proteasome inhibitor bortezomib. *Oncotarget* 2(3):209–221.
- Geiler-Samerotte KA, et al. (2011) Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. *Proc Natl Acad Sci USA* 108(2):680–685.
- Boyko AR, et al. (2008) Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet* 4(5):e1000083.
- Andersen MH, Schrama D, Thor Straten P, Becker JC (2006) Cytotoxic T cells. *J Invest Dermatol* 126(1):32–41.
- Gillespie JH (2004) *Population Genetics: A Concise Guide* (Johns Hopkins Univ Press, Baltimore).
- Spencer SL, Gerety RA, Pienta KJ, Forrest S (2006) Modeling somatic evolution in tumorigenesis. *PLoS Comput Biol* 2(8):e108.
- Beerenwinkel N, et al. (2007) Genetic progression and the waiting time to cancer. *PLoS Comput Biol* 3(11):e225.
- Iwasa Y, Nowak MA, Michor F (2006) Evolution of resistance during clonal expansion. *Genetics* 172(4):2557–2566.
- Laird AK (1964) Dynamics of tumor growth. *Br J Cancer* 13:490–502.
- Bozic I, et al. (2010) Accumulation of driver and passenger mutations during tumor progression. *Proc Natl Acad Sci USA* 107(43):18545–18550.
- Kim S, Chin K, Gray JW, Bishop JM (2004) A screen for genes that suppress loss of contact inhibition: Identification of ING4 as a candidate tumor suppressor gene in human cancer. *Proc Natl Acad Sci USA* 101(46):16251–16256.
- Johnston MD, Edwards CM, Bodmer WF, Maini PK, Chapman SJ (2007) Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer. *Proc Natl Acad Sci USA* 104(10):4008–4013.
- Felsenstein J (1974) The evolutionary advantage of recombination. *Genetics* 78(2):737–756.
- Guthrie J (1966) Effects of a synthetic antigonadotrophin (2-amino,5-nitrothiazole) on growth of experimental testicular teratomas. *Br J Cancer* 20(3):582–587.
- Gibson MA, Bruck J (2000) Efficient exact stochastic simulation of chemical systems with many species and many channels. *J Phys Chem A* 104(9):1876–1889.
- Loeb LA, Bielas JH, Beckman RA (2008) Cancers exhibit a mutator phenotype: Clinical implications. *Cancer Res* 68(10):3551–3557, discussion 3557.
- Cole AM, et al. (2010) Cyclin D2-cyclin-dependent kinase 4/6 is required for efficient proliferation and tumorigenesis following Apc loss. *Cancer Res* 70(20):8149–8158.
- Forbes SA, et al. (2008) The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum Genet* Chapter 10:Unit 10.11.
- Beckman RA, Loeb LA (2005) Negative clonal selection in tumor evolution. *Genetics* 171(4):2123–2131.
- Berger MF, et al. (2012) Melanoma genome sequencing reveals frequent PREX2 mutations. *Nature* 485(7399):502–506.
- Fischer A, Greenman C, Mustonen V (2011) Germline fitness-based scoring of cancer mutations. *Genetics* 188(2):383–393.
- Gupta PB, et al. (2011) Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* 146(4):633–644.
- Korolev KS, et al. (2012) Selective sweeps in growing microbial colonies. *Phys Biol* 9(2):026008.
- Korolev KS, Avlund M, Hallatschek O, Nelson DR (2010) Genetic demixing and evolution in linear stepping stone models. *Rev Mod Phys* 82(2):1691–1718.
- Aguirre-Ghiso JA (2006) The problem of cancer dormancy: Understanding the basic mechanisms and identifying therapeutic opportunities. *Cell Cycle* 5(16):1740–1743.
- Brunet E, Rouzine IM, Wilke CO (2008) The stochastic edge in adaptive evolution. *Genetics* 179(1):603–620.
- Gordo I, Charlesworth B (2000) On the speed of Muller's ratchet. *Genetics* 156(4):2137–2140.
- Neher RA, Shraiman BI (2012) Fluctuations of fitness distributions and the rate of Muller's ratchet. *Genetics* 191(4):1283–1293.
- Haigh J (1978) The accumulation of deleterious genes in a population—Muller's ratchet. *Theor Popul Biol* 14(2):251–267.
- Sunyaev SR, et al. (1999) PSIC: Profile extraction from sequence alignments with position-specific counts of independent observations. *Protein Eng* 12(5):387–394.
- Birkbak NJ, et al. (2011) Paradoxical relationship between chromosomal instability and survival outcome in cancer. *Cancer Res* 71(10):3447–3452.
- Fudenberg G, Getz G, Meyerson M, Mirny LA (2011) High order chromatin architecture shapes the landscape of chromosomal alterations in cancer. *Nat Biotechnol* 29(12):1109–1113.
- Sheltzer JM, et al. (2011) Aneuploidy drives genomic instability in yeast. *Science* 333(6045):1026–1030.
- Dai C, Whitesell L, Rogers AB, Lindquist S (2007) Heat shock factor 1 is a powerful multifaceted modifier of carcinogenesis. *Cell* 130(6):1005–1018.
- De Raedt T, et al. (2011) Exploiting cancer cell vulnerabilities to develop a combination therapy for ras-driven tumors. *Cancer Cell* 20(3):400–413.
- Wust P, et al. (2002) Hyperthermia in combined treatment of cancer. *Lancet Oncol* 3(8):487–497.
- Soong NW, Hinton DR, Cortopassi G, Arnheim N (1992) Mosaicism for a specific somatic mitochondrial DNA mutation in adult human brain. *Nat Genet* 2(4):318–323.
- Santagata S, et al. (2011) High levels of nuclear heat-shock factor 1 (HSF1) are associated with poor prognosis in breast cancer. *Proc Natl Acad Sci USA* 108(45):18378–18383.
- Silva JM, et al. (2000) DNA damage after chemotherapy correlates with tumor response and survival in small cell lung cancer patients. *Mutat Res* 456(1–2):65–71.
- Parsons DW, et al. (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science* 321:1807–1812.
- Ley TJ, et al. (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 456:66–72.

# The impact of deleterious passenger mutations on cancer progression

Christopher D McFarland, Kirill S Korolev, Gregory V Kryukov, Shamil Sunyaev, and Leonid A Mirny

## Supplemental Text and Figures

**Design of our model for neoplastic growth.** Cells in our populations were fully described by their number of drivers  $d$  and passengers  $p$ . Birth and death events were modeled using an implementation of the Next Reaction (1), a Gillespie Algorithm that orders events using a Heap Queue. Generation time in our model was defined as the inverse of the mean birth rate of the population:  $1 / \langle B(d, p) \rangle$ . All mutation events occurred during cell division. If mutations occur per unit of time, rather than *per generation*, we would expect that rapidly growing tumors would acquire drivers slower because generation times shrink as the growth rate increases. Likewise, we would expect that rapidly declining populations would acquire passengers faster. Because  $\mu T_p$  exceeds 1 for large mutation rates, each daughter cell acquires a Poisson-distributed, pseudo-random number of new passenger mutations from its parent, with mean  $\mu T_p$ .

**Estimating and exploring the parameters and functional form of the model.** To account for the extensive heterogeneity of cancer between and within sub-types, and to account for our limited quantitative knowledge of cancer, we varied the values of all parameters by 2-3 orders of magnitude. The ranges we explored centered about values derived from the literature (**Table S1**). We found that deleterious passenger mutations accumulate under this broad range of conditions (**Fig. S2**).

The effect of each driver was assumed to be very significant ( $s_d \approx 0.1$ , i.e. individual drivers increase the growth rate by 10%) because previous studies found this rate to be congruent with the time to cancer onset (2). Simulations with fixed  $s_d$  were varied between 0.001 and 1. Simulations, where drivers conferred a Gaussian or exponential distribution of fitness advantages did not differ qualitatively from our fixed-effect model (**Fig. S1**).

Recent estimates of the effects of near-neutral germ line mutations in humans (3), as well as randomly introduced mutations in yeast (4), suggest the disadvantage conferred by a passenger is very weak ( $s_p \approx 0.001$ , range  $10^{-4} - 10^{-1}$ ). Like drivers, we simulated trajectories with passenger mutations drawn from several potential fitness distributions (5). The effect of each passenger  $x$ , was drawn from either an exponential distribution, a Log-Normal distribution, or a Gamma distribution with the following density functions.

$$\begin{aligned} \text{Exponential}(x | s_p) &= \text{Exp}[-x / s_p] s_p^{-1} \\ \mathcal{L} \mathcal{N}(x | s_p, \sigma = 1) &= \frac{1}{x \sqrt{2\pi}} \text{Exp}[-(\ln(x / s_p) + 1/2)^2 / 2] \\ \Gamma(x | k = 2, \theta) &= s_p^{-2} x \text{Exp}[-2x / s_p] \end{aligned}$$

All distributions have mean value  $s_p = 0.001$ . The probability of fixation of passengers declined only for very deleterious passengers (**Fig. 3B, Fig. S1**).

The target size for driver mutations  $T_d$ , represents all relevant sites in all genes associated with cancer development (i.e. mutational hot spots in oncogenes and tumor suppressors). Our estimate,  $T_d \approx 700$  positions per genome represents 70 genes [estimated in (6)] times approximately 10 activating mutations per gene (range 70-7,000). This value has been used previously in simulations (2), is projected from current sequencing data, and is approximately equal in number to the 571 loci with observed recurrent mutations in colon cancer (7). While the number of potentially inactivating mutations for any tumor suppressors is certainly larger than 10 per gene, and mutational effects should remain silent until a LOH event, for parsimony we assume that these nuances collectively result in an effective target size for a tumor suppressor that is still approximately 10. Because the number of observed mutations in major oncogenes is of the same order as the observed number of mutations in major tumor suppressors (7), we assume this approximation is reasonable.

Our estimate for the target size of potentially deleterious passengers represents  $\sim 5,000$  non-cancer related genes expressed in cancer multiplied by  $\sim 1,000$  non-synonymous and non-neutral loci per gene. This quantity is significantly smaller than the entire protein-coding genome and even previous estimates of deleterious loci in cancer of 10,000,000 (8) because our model is intently focused on the subset of loci in the genome that have intermediate deleterious effect (**Fig. S6**). Hence, we are ignoring loci with either neutral or lethal effect. Nevertheless,  $T_p$  is almost ten thousand times the target size for drivers. The mutation rate ( $\mu \approx 10^{-8} \text{ nucleotide}^{-1} \cdot \text{division}^{-1}$ , range  $10^{-10} - 10^{-6}$ ) approximates cells that have acquired a mutator phenotype (9). The initial equilibrium population size ( $K \approx 1,000$ , range 100–10,000) is on the order of the population size of the niche in which carcinomas originate (i.e. the colonic crypt size, lung alveoli size, etc). We estimated this quantity from the observed size of hyperplasias within a mouse colonic crypt 2 weeks following an initiating *APC* deletion (10).

Collectively, our choice of form and parameters for our model appears appropriate to cancer. Our model exhibits many of the key features of tumor progression without being pre-programed to do so (**Table S2**).

**Estimating the critical population size.** The occurrence of a critical population size can be understood using classical population genetics theory and several simplifying approximations. If we assume  $B(d, p) = D(N)$  [verified in (**Fig. S1**)] and consider the accumulation of advantageous drivers and deleterious passengers as independent processes, then the change in population size can be written as:

$$\frac{dN}{dt} = v_d - v_p$$

Where  $v_d$  is the increase in population per unit time (velocity) due to fixation of drivers and  $v_p$  is the decrease in population per unit time due passenger fixation. New drivers and passengers arise in the population with rates  $\mu T_d N$  and  $\mu T_p N$  respectively. After occurring, their probability of fixation in the

absence of interfering mutations is  $\pi_d = \frac{s_d}{1 + s_d} \approx s_d$  (11) for an advantageous driver, and  $\pi_p \approx 1/N$

for a mildly deleterious passenger that fixates at an effectively neutral rate (which is expected for high mutation rates or small effect size, as we will show below). Once fixated, each mutation alters the population size by  $\Delta N_d = N_{d+1} - N_d$  and  $\Delta N_p = N_{p+1} - N_p$  respectively. Evoking the earlier assumption:  $B(d, p, N_d) = D(d, p, N_d)$ , and  $B(d + 1, p, N_{d+1}) = D(d + 1, p, N_{d+1})$ , yields  $\Delta N_d \approx N s_d$  and  $\Delta N_p \approx N s_p$ .

Upon multiplying the probability of occurrence of a mutation, the probability that it fixates, and the resulting change in population size, we obtain:

$$v_d = \mu T_d N \cdot s_d \cdot N s_d = \mu T_d s_d^2 N^2$$

$$v_p = \mu T_p N \cdot \frac{1}{N} \cdot N s_p = \mu T_p s_p N$$

$$\frac{dN}{dt} = \mu T_d s_d^2 N^2 - \mu T_p s_p N = \mu T_p s_p N \left( \frac{N}{N_{critical}} - 1 \right), \text{ where}$$

$$N_{critical} = \frac{T_p s_p}{T_d s_d^2}$$

Thus, the population, on average, will decline for small  $N$  and increase for large  $N$ .  $N_{critical}$  represents an unstable fixed point, which roughly corresponds to the observed critical population size in our simulations (**Fig. S3**).

**Accumulation of passengers of fixed-effect.** The rate of accumulation of deleterious mutations, in the absence of clonal expansions, has been well studied previously (11–14). Here we use existing theories to describe how passengers accumulate. In the absence of drivers, the critical parameters are:  $N$ ,  $s_p$ , and  $\mu_p$ , where  $\mu_p = \mu T_p$  is the genome-wide passenger mutation rate. The evolution dynamics of passenger accumulation strongly depends upon  $\lambda = \mu_p/s_p$ , which represents the strength of mutation relative to natural selection.

Unless  $\lambda$  is very large, the population is approximately in mutation selection balance between passenger fixation events. This means that the increase in passengers load via mutations is balance by natural selection:

$$\mu_p(n_{p-1} - n_p) - s_p(P - \bar{P})n_p = 0$$

Here, the number of cells with  $P$  passengers is  $n_p$  and  $\bar{P}$  corresponds to the mean number of passengers in the population. This mutation selection balance leads to a stationary distribution of  $n_p$  of the following form:

$$n_{p, stationary} = N \frac{\lambda^p}{p!} e^{-\lambda}$$

Note that this solution takes the shape of a Poisson distribution with mean and variance  $\lambda$  (**Fig. S5**).

When  $\lambda \ll 1$ , mutations are so rare and/or passengers are so deleterious that essentially only two alleles segregate in the population. In these circumstances, the exact probability that a mutation fixates ( $\Pi$ ) is described in (11) and is:

$$\Pi(s_p, N) = \frac{s_p}{1 - (1 + s_p)^{-N}}$$

Note that this expression assumes a particular dynamics known as a Moran process. The probability of fixation declines very steeply as  $s_p$  or  $N$  increase.

For larger  $\lambda$ , the population with the least number of passengers, also known as the fittest class, is only a fraction of the total population size,  $n_0 = Ne^{-\lambda}$ . Therefore, this fittest class can stochastically go extinct due to fluctuations in population size. This extinction is irreversible, so the population must shift to a lower fitness value in a process known as Muller's Ratchet. For moderately large  $\lambda$ , the time between ratcheting events  $T_{click}$  was calculated in (12, 15, 16) as:

$$T_{click} = \frac{e-1}{s_p} \text{Exp}\left[\frac{s_p N(e-1)}{2} e^{-\lambda}\right]$$

For even larger  $\lambda$  (when  $s_p Ne^{-\lambda} \approx 1$ ), the time between the clicks of the ratchet becomes so short that there is insufficient time for the population to reach mutation selection balance. In this regime the system can be approximated by a traveling wave with the accumulation rate  $v$  (13), given by :

$$1 \approx \frac{v}{2\mu_p} \left( \log^2 \left[ \frac{e\mu_p}{v} \right] + 1 \right) - \lambda^{-1} \log \left[ N s_p \lambda^{-\frac{1}{2}} \frac{e\mu_p}{v} \right] / \left( 1 - \frac{v}{\mu_p} \log \left[ \frac{e\mu_p}{v} \right] + \frac{5}{6\lambda} \right)$$

For very large  $\lambda$ , mutations are effectively neutral and the accumulation rate becomes  $\mu_p$ .

Muller's Ratchet, the traveling wave solution, and the neutral approximation collectively explain the rate of passenger accumulation in the absence of driver mutations across time scales relevant to cancer progression (**Fig. S4**).

**Accumulation of passengers of varying effect.** When deleterious mutations are drawn from a distribution of effect sizes, estimating their accumulation rate becomes more complicated and rigorous analytical treatment of this situation has not been published to our knowledge. However, we will offer an approximate treatment in the limit of small  $\lambda$  here. **Figure 3B** demonstrates that the population fitness distributions remain constant between the fixed-effect and variable effect models, with one caveat: the mean and variance of fitness relative to the fittest class are now  $\bar{\lambda} = \mu_p / \bar{s}_p$ . Here,  $\bar{s}_p$  represents the mean deleteriousness of a passenger. While  $\bar{\lambda}$  is relatively large in **Figure 3B** (approximately 50), this congruence should presumably hold so long as mutation effect sizes are generally smaller than the width of fitness distribution. Obviously, if the variance of the distribution of fitness were very large or undefined (e.g. the variance of a power-law distribution) this approximation may not hold. So long as this is not the case, a new mutation, with its particular fitness effect  $\hat{s}_p$ , has, in the limit of small  $\bar{\lambda}$ , a fixation probability of

$$\Pi = \hat{s}_p / (1 - (1 + \hat{s}_p)^{\text{NExp}[-\bar{\lambda}]})$$

Here we used the fixation probability in the weak mutation limit, but with population size described by the size of the fittest class. For small  $\lambda$ , it is clear that mutations in the deleterious tail, where  $\hat{s}_p > \bar{s}_p$ , will fixate in this model more often than in a model where all passengers exhibit effect an effect size of  $\hat{s}_p$ . Indeed, in the fixed-effect model, as mutations become more deleterious they 1) are less likely to fixate in a fittest class of equivalent size, and 2) increase the size of this fittest class—further reducing their probability of fixation. However in a distribution of effect sizes, the size of the fittest class is defined by  $\bar{s}_p$  not  $\hat{s}_p$ . Hence, these rare, very deleterious, mutations only reduce their probability of fixation in the fittest class, but do not affect the size of this fittest class. Their fixation probability then declines less rapidly than one might expect from the fixed-effect model (**Fig. 3B**).

For large  $\bar{\lambda}$ , accumulation rates are high and well approximated by a neutral model (**Fig. S4**), so we expect at most a modest effect of  $\hat{s}_p$  relative to  $\bar{s}_p$  on the rate of accumulation, although we know of

no existing theory that describes this scenario. The fixation rate should still exhibit some decline, if the fitness distribution has a long deleterious tail—as illustrated in **Figure 3B, S1E and S1F**.

When considering the effects of treatment strategies on passengers of varying effect, it is important to keep in mind that nearly all of the passengers residing in a population that progress to cancer have fixated. Therefore, nearly all of the effects of increased selection against passengers will be on mutations that cannot change in frequency in the population. For this reason, cancers grown under previously estimated parameters with passengers drawn from an exponential distribution exhibited similar relapse rate as cancers grown under the fixed effect model.

**Analysis of somatic mutations in cancer.** All cancer mutations were collected from the ongoing COSMIC database at <http://www.sanger.ac.uk/genetics/CGP/cosmic/> (7). COSMIC and TCGA, along with other cancer genomics consortia, have focused on identifying driver mutations (i.e. distinguishing drivers from passengers) by their recurrence in multiple patients or samples (7). Using COSMIC, we identified 4,195 missense passenger mutations (non-synonymous, amino acid changes) from a total of 116,977 mutations. We defined a mutation as a passenger if it arose in a gene not listed in the census of possible cancer-causing genes (17). These 4,195 ‘passenger’ mutations show no recurrence and are dispersed across 3,172 genes, further supporting their classification as passengers. We then contrasted these mutations with driver mutations and three reference datasets: 1) benign, common human non-synonymous SNPs; 2) simulated *de novo* mutations (randomly generated using a cancer-specific 3-parameter model described in detail below); and 3) damaging, pathogenic missense mutations causing Mendelian human diseases (from the Human Gene Mutation Database, HGMD). Common SNPs and disease causing mutations were obtained previously for validation of *POLYPHEN2* (18). In our more stringent classification of passenger mutations, we discarded: 1) all passengers in genes that harbored more than one passenger, 2) passengers in any genes where  $\omega > 1$  (**Fig. S7**), and 3) passengers that were not ‘confirmed’ somatic mutations in the COSMIC dataset (only 29.4% of mutations in the database were ‘confirmed’ by follow-up Sanger sequencing). Mean  $\Delta PSIC$  for this stringent set of passengers did not differ significantly ( $p < 0.69$ ) from our original set, so it was not used for further controls as it greatly reduced sample size.

To stratify passengers into various subsets, we used several resources. 372 passenger mutations were classified as ‘Homozygous’ by COSMIC, presumably due to some kind of Loss of Heterozygosity event. ‘Housekeeping’ genes, were 195 genes with passenger mutations and with one-to-one orthologs in *S. cerevisiae*, identified using InParanoid (19). These genes are well expressed in humans, so we believe it is highly likely that they are expressed in cancer. We could not directly normalize mutations in our dataset by their expression levels because mutations in the COSMIC database derive from varied literature sources (which often lack direct expression data). 881 non-COSMIC, non-synonymous passenger mutations were obtained from The Cancer Genome Atlas’ analysis of 38 Multiple Myeloma genomes (20). This subset was used as a control to ensure that any biases, which COSMIC might introduce via literature curation, did not account for our observed scores.

To parameterize our random model of pan-cancer mutations, we collected all 1,128 synonymous mutations present in COSMIC at the time of this study. Given our sample size, we parameterized our model to account for 3 types of point mutations: transversions, CpG to TpG transitions, and all other transitions, as these 3 processes seemed to explain observed mutation patterns best. Because some genes in COSMIC, like *KRAS* or *TP53*, are sequenced more often than others, we weighted both our estimated parameters and simulated mutations by the frequency with which each gene was sequenced;

COSMIC notes in their manifest that they take care to record how often a gene is sequenced since it is critical to determining whether-or-not a gene is abnormally mutated. We chose to weight genes this way because the frequency that genes were sequenced was highly variable and highly skewed, especially towards driver genes. Hence, the frequencies of mutations  $f_i$  for all three mutational processes (transversions, CpG transitions, other transitions) were estimated as follows:

$$f_i = \sum_j^{all\ genes} w_j \frac{O_{ij}}{P_{ij}}$$

Here,  $O_{ij}$  is the number of observed synonymous mutations belonging to mutational class  $i$ , for a particular gene  $j$ ,  $P_{ij}$  is the number of possible unique synonymous mutations for the class  $i$  of gene  $j$ , and  $w_j$  is the fractional of cancer sequences reported in COSMIC that belong to gene  $j$ . This model explained the observed patterns of non-synonymous mutations with greater log-likelihood than two-parameter models, as well as more sophisticated 10-parameter models or models developed for human germ-line mutations (21). Using this model, random mutations were drawn with probability  $f_i w_j$  from the set of all possible genome-wide, non-synonymous mutations. These randomly-generated mutations were not only used as a null model for evolutionary conservation, but also as a neutral null-model to test for signatures of positive and negative selection in cancer genomes (**Fig. S7**).

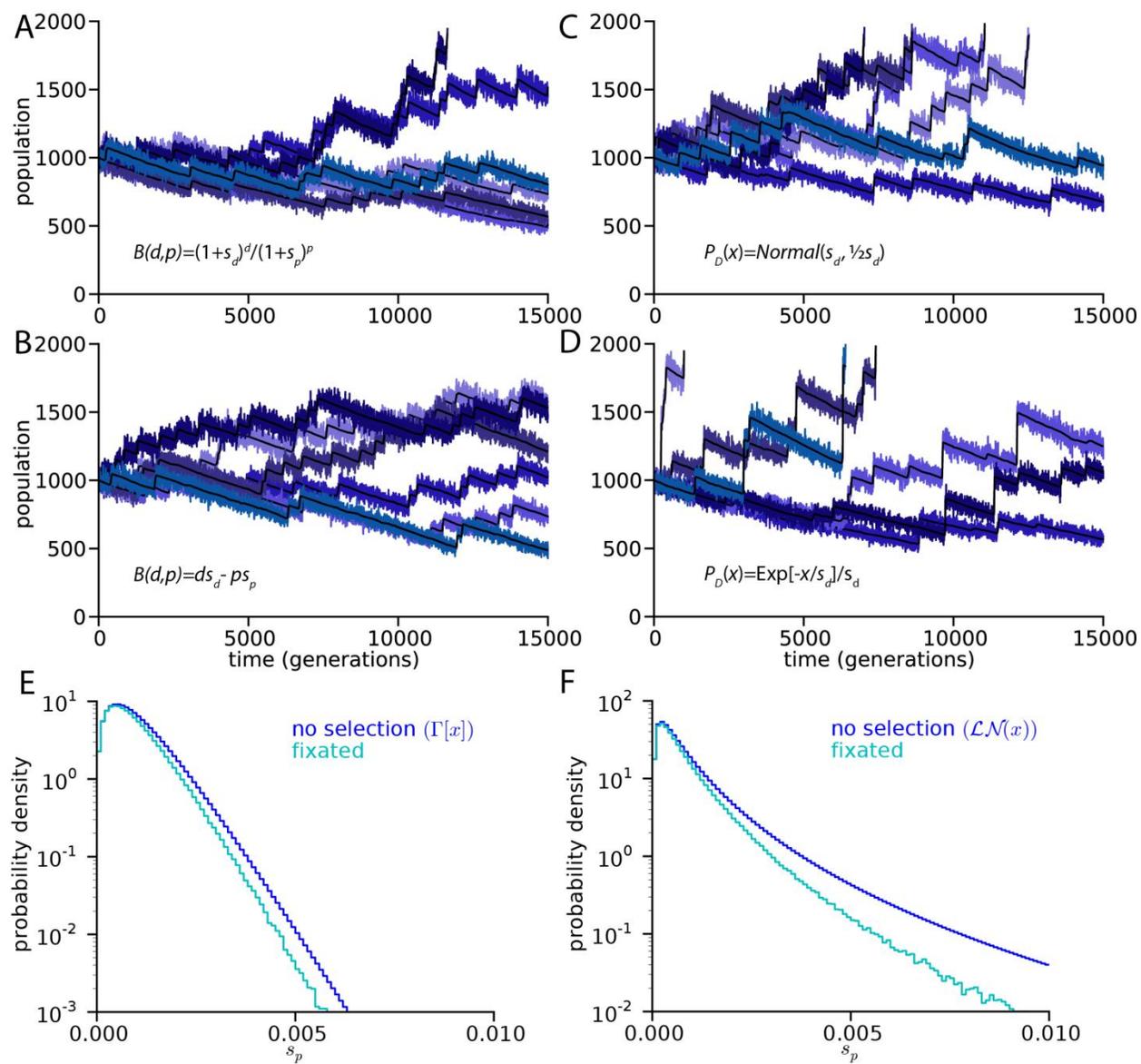
## REFERENCES

1. Gibson M a., Bruck J (2000) Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *The Journal of Physical Chemistry A* 104:1876–1889.
2. Beerenwinkel N et al. (2007) Genetic progression and the waiting time to cancer. *PLoS computational biology* 3:e225.
3. Fischer A, Greenman C, Mustonen V (2011) Germline fitness-based scoring of cancer mutations. *Genetics* 188:383–93.
4. Geiler-Samerotte KA et al. (2011) Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. *Proceedings of the National Academy of Sciences of the United States of America* 108:680–5.
5. Nielsen R, Yang Z (2003) Estimating the distribution of selection coefficients from phylogenetic data with applications to mitochondrial and viral DNA. *Molecular biology and evolution* 20:1231–9.
6. Sjöblom T et al. (2006) The consensus coding sequences of human breast and colorectal cancers. *Science (New York, N.Y.)* 314:268–74.
7. Forbes S a et al. (2008) The Catalogue of Somatic Mutations in Cancer (COSMIC). *Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.]* Chapter 10:Unit 10.11.

8. Beckman R a, Loeb L a (2005) Negative clonal selection in tumor evolution. *Genetics* 171:2123–31.
9. Loeb L a, Bielas JH, Beckman R a (2008) Cancers exhibit a mutator phenotype: clinical implications. *Cancer research* 68:3551–7; discussion 3557.
10. Cole AM et al. (2010) Cyclin D2-cyclin-dependent kinase 4/6 is required for efficient proliferation and tumorigenesis following Apc loss. *Cancer research* 70:8149–58.
11. Nowak MA (2006) *Evolutionary dynamics: exploring the equations of life* (Belknap Press).
12. Gordo I, Charlesworth B (2000) On the Speed of Muller ' s Ratchet. *Genetics* 2140:2137–2140.
13. Brunet E, Rouzine IM, Wilke CO (2008) The stochastic edge in adaptive evolution. *Genetics* 179:603–20.
14. Neher R a, Shraiman BI (2012) Fluctuations of fitness distributions and the rate of Muller's ratchet. *Genetics* 191:1283–93.
15. Neher R a., Shraiman BI (2009) Competition between recombination and epistasis can cause a transition from allele to genotype selection. *Proceedings of the National Academy of Sciences* 106:6866–6871.
16. Gordo I, Charlesworth B (2000) The Degeneration of Asexual Haploid Populations and the Speed of Muller ' s Ratchet.
17. Futreal PA et al. (2004) A census of human cancer genes. *Nature reviews. Cancer* 4:177–83.
18. Adzhubei I a et al. (2010) A method and server for predicting damaging missense mutations. *Nature methods* 7:248–9.
19. O'Brien KP, Remm M, Sonnhammer ELL (2005) Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic acids research* 33:D476–80.
20. Chapman M a et al. (2011) Initial genome sequencing and analysis of multiple myeloma. *Nature* 471:467–72.
21. Stamatoyannopoulos J a et al. (2009) Human mutation rate associated with DNA replication timing. *Nature genetics* 41:393–5.
22. Jordan DM, Ramensky VE, Sunyaev SR (2010) Human allelic variation: perspective from protein function, structure, and evolution. *Current opinion in structural biology* 20:342–50.
23. Tong a H et al. (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science (New York, N.Y.)* 294:2364–8.
24. Jackson a L, Loeb L a (1998) The mutation rate and cancer. *Genetics* 148:1483–90.

25. Boyko AR et al. (2008) Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS genetics* 4:e1000083.
26. Beroukhim R et al. (2010) The landscape of somatic copy-number alteration across human cancers. *Nature* 463:899–905.
27. Aguirre-Ghiso JA (2006) The problem of cancer dormancy: understanding the basic mechanisms and identifying therapeutic opportunities. *Cell cycle (Georgetown, Tex.)* 5:1740–3.
28. Hanahan D, Weinberg R a (2011) Hallmarks of cancer: the next generation. *Cell* 144:646–74.
29. Pleasance ED et al. (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 463:191–6.
30. Pleasance ED et al. (2010) A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 463:184–90.
31. Ley TJ et al. (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 456:66–72.
32. Yachida S et al. (2010) Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* 467:1114–7.
33. Campbell PJ et al. (2008) Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 105:13081–6.
34. Silva JM et al. (2000) DNA damage after chemotherapy correlates with tumor response and survival in small cell lung cancer patients. *Mutation research* 456:65–71.

## FIGURES



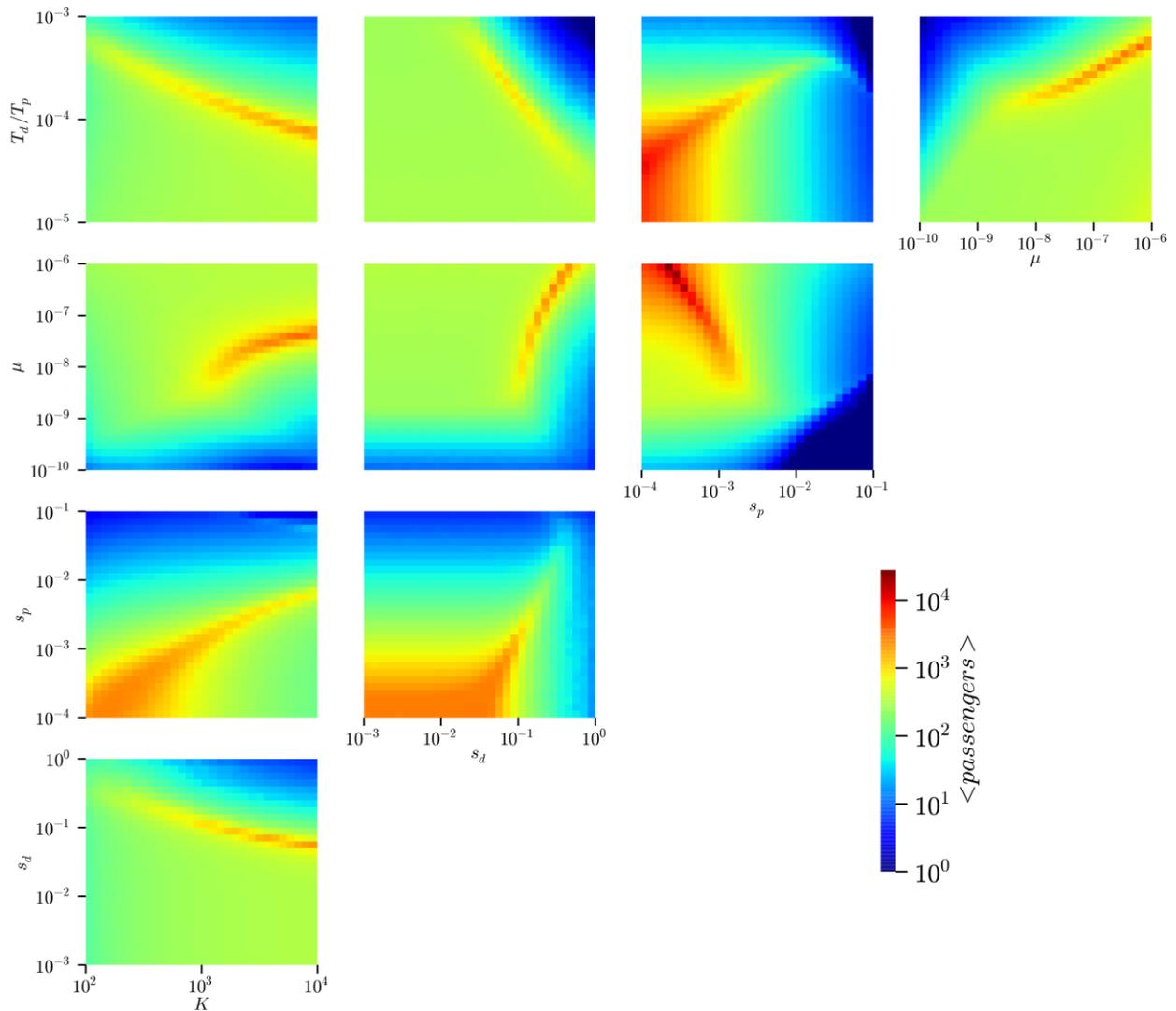
**Fig. S1. Qualitative behavior of model is invariant to shape of birth and death functions. Six**

trajectories (blue shades) with rates<sup>1</sup>  $B(d, p) = \frac{(1 + s_d)^d}{(1 + s_p)^p}$  and  $D(N) = N/K$  (**A**) appear qualitatively

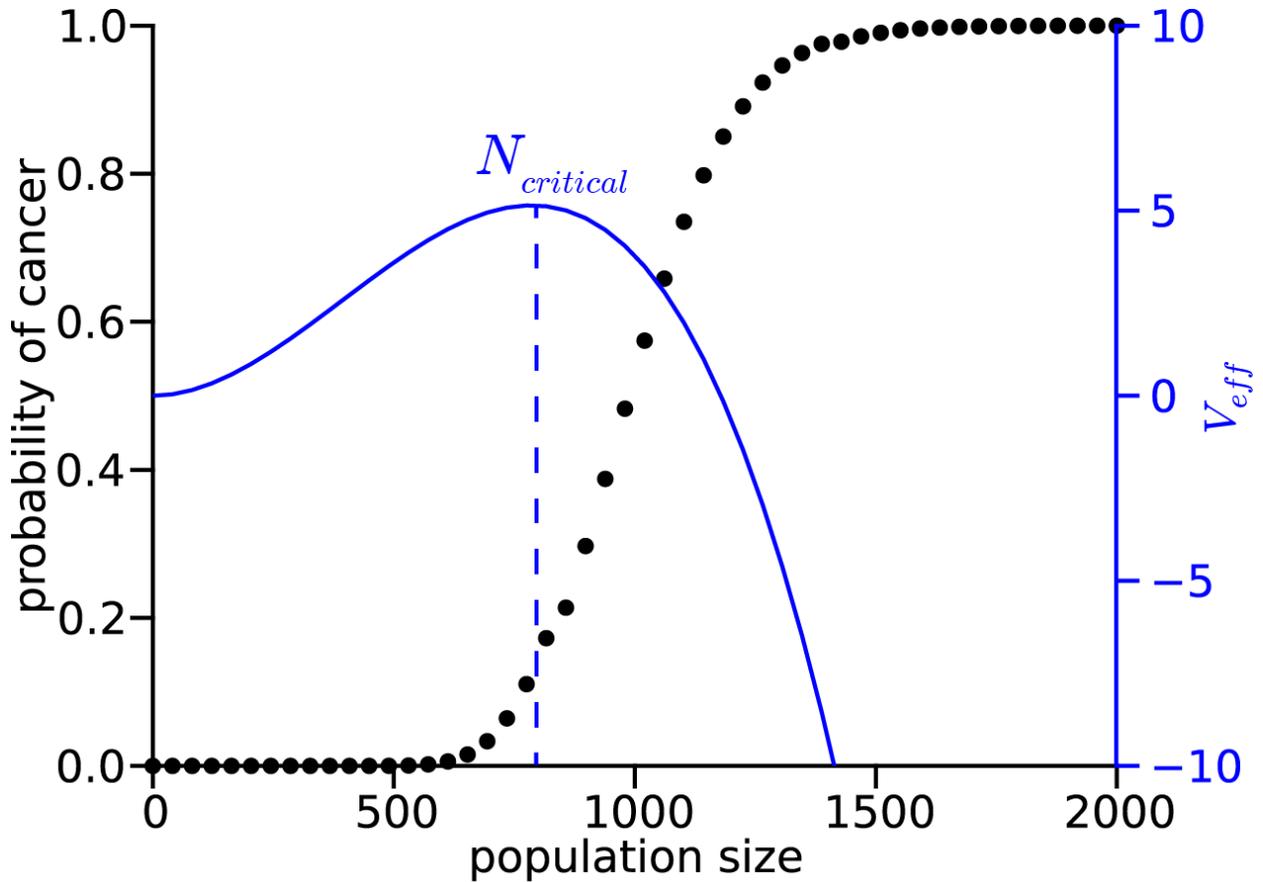
similar to simulations with rates with rates  $B(d, p) = 1 + s_d d - s_p p$  and  $D(N) = N/K$ . (**B**). They are also mathematically equivalent to first order expansion of  $(d, p)$ . Black lines represent  $N$  which satisfies  $D(N) = B(d, p)$ . The strong overlap of this black line with the observed population size indicates that birth and death are balanced throughout progression. For this reason, a model where mutations alter death rates, rather than birth rates, would be equivalent to our model provided that time is measured in units of generations. Trajectories where driver loci are assigned fitness advantage by sampling from a Gaussian distribution (**C**) or Exponential distribution (**D**) appear qualitatively similar to trajectories with fixed fitness effects.  $P_D$  represents the probability density function from which drivers were drawn. For the Normal distribution  $P_D(x) = \sqrt{\frac{2}{s_d \pi}} \text{Exp} \left[ -\frac{2(x-s_d)}{s_d^2} \right]$ , while  $P_D = \text{Exp}[-x/s_d] s_d^{-1}$  for the exponential distribution. All trajectories have a mean  $s_d$  of 0.1. Likewise, passengers drawn from a Gamma distribution (**E**) or Log-normal distribution (**F**) accumulate at a similar rate to passengers with fixed effects.

---

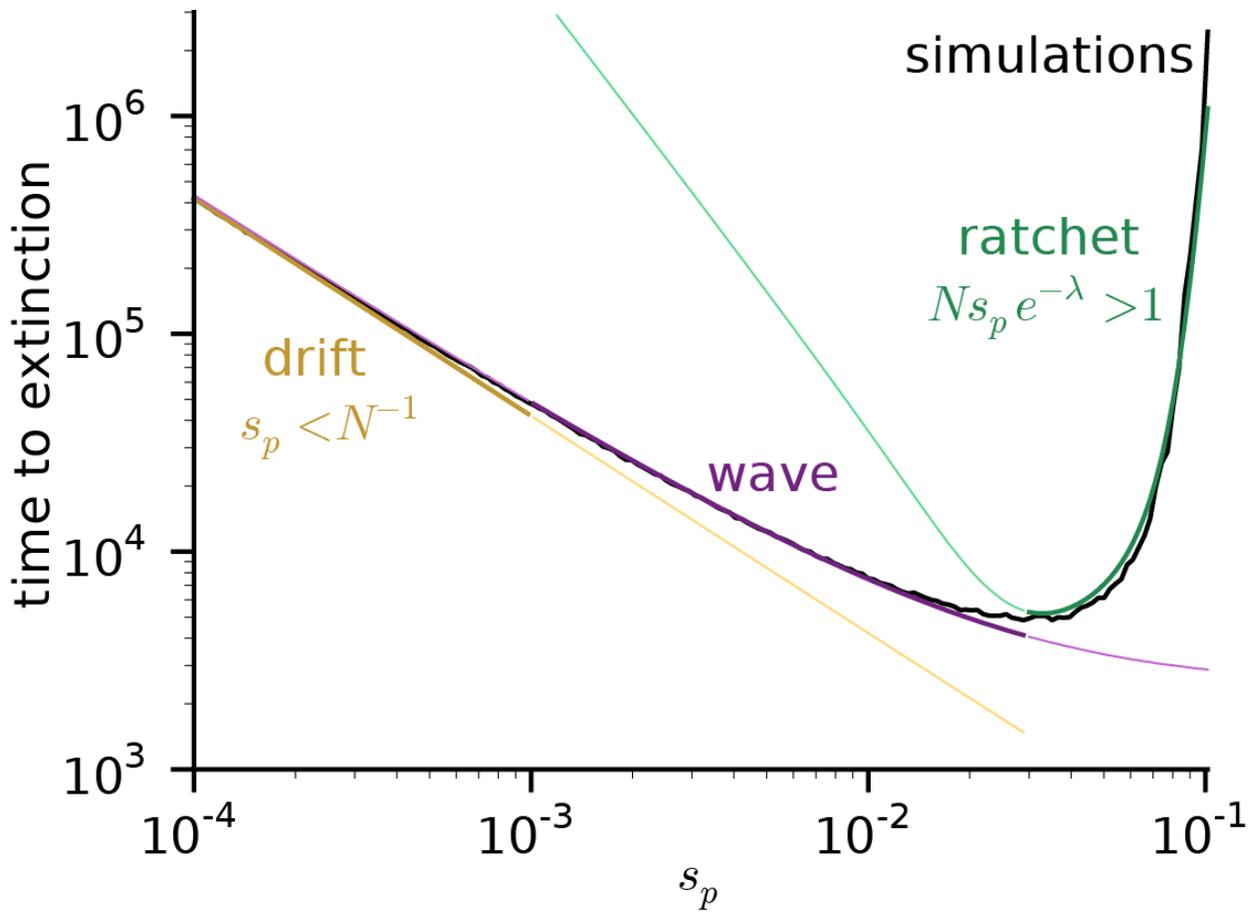
<sup>1</sup> This functional form of  $B(d, p)$  is exactly equivalent to  $(1 + s_d)^d (1 - s_p)^p$  by rescaling  $s_p$ .



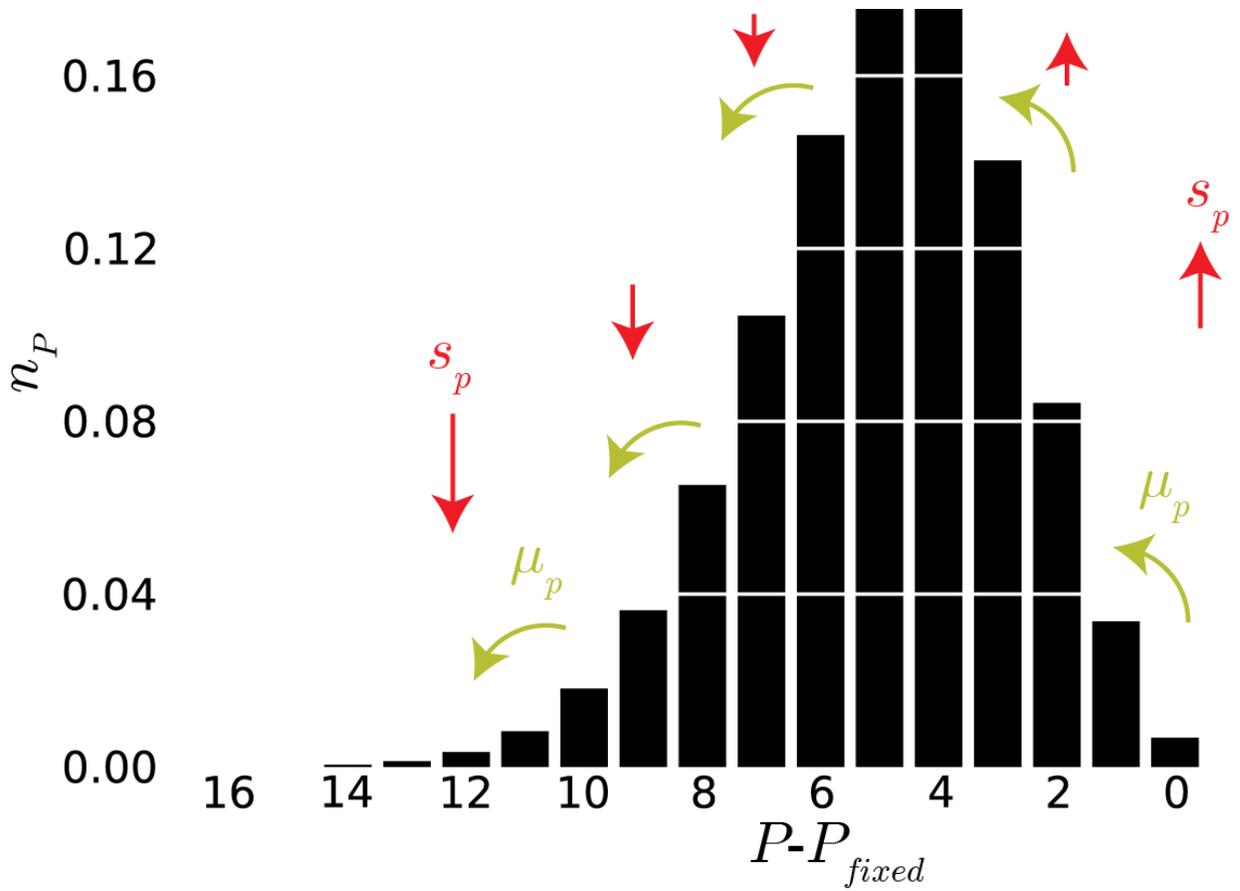
**Fig. S2. Passengers accumulate non-monotonically across the entire parameter space.** We simulated tumor progression across all parameters. For each heat map, the three parameters not shown on the x or y axes were set to estimated values (**Table S1**). For each element in the heat map, 200 trajectories were simulated until they progressed to cancer, progress to extinction, or until 15,000 generations (~50 years), whichever was sooner. We define progression as doubling of cancer size because, after doubling, populations progress quickly to cancer and acquire few additional passengers. The mean number of accrued passengers is shown with color. Because variation in the mutation rate and  $s_p$  alter the time to cancer and the probability of progression to cancer, the number of passengers depends non-monotonically on both parameters. We plan to explore and explain our model's rich behavior across its phase space in a future article.



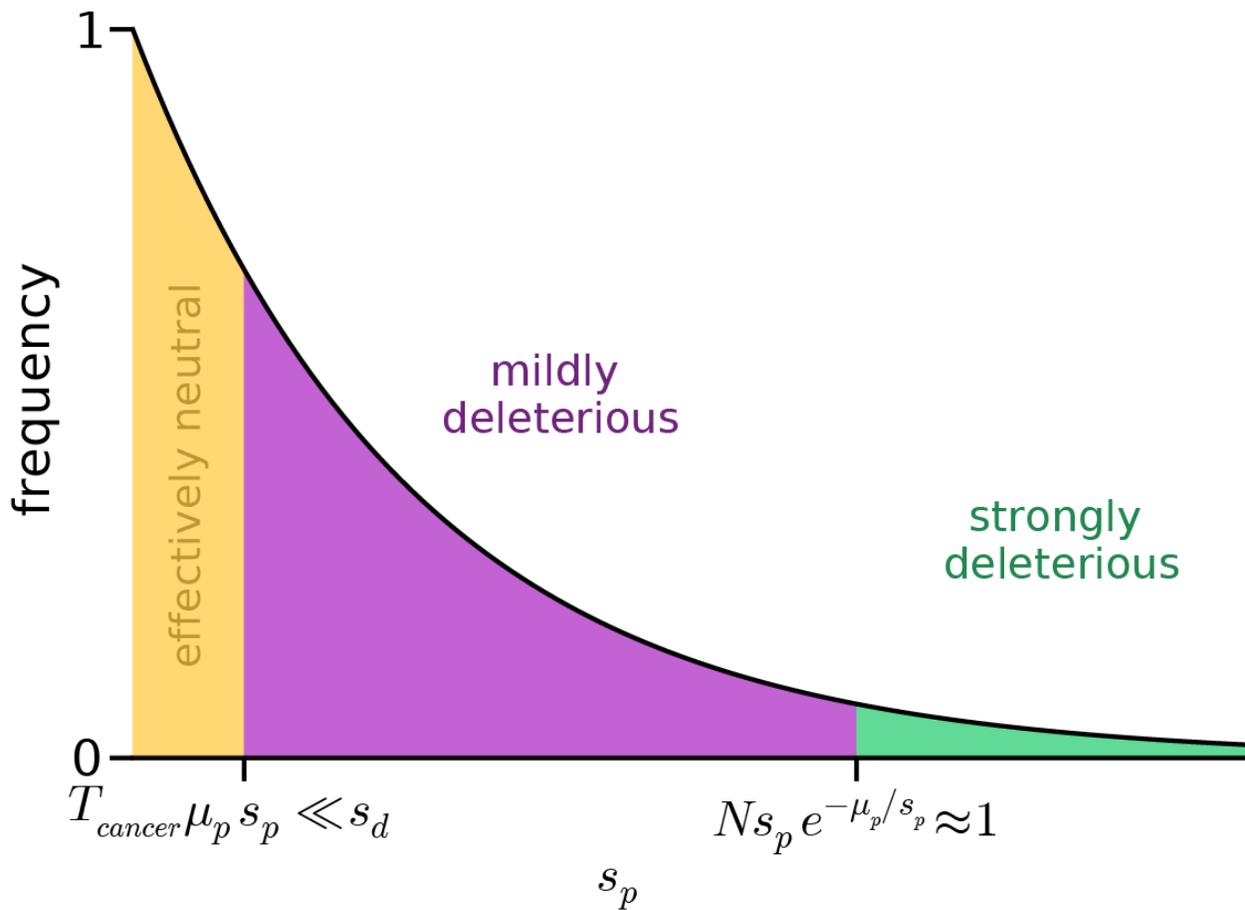
**Fig. S3. Existence of critical population size for progression to cancer.** The probability of progressing to cancer (black dots; 500 simulations using parameters in **Table S1**) exhibited a sigmoidal dependence upon the initial population size ( $K$ ). Below this critical size, populations most often regressed from passenger accumulation and above this critical size they most often progressed to cancer. To the first approximation, overcoming this critical size is a stochastic process of barrier crossing, akin to many processes in chemical kinetics, where  $N$  is a reaction coordinate. In this analogy, the energy of the reaction is given by  $V_{eff} = -\int \frac{dN}{dt} dN$ . Hence,  $N_{critical}$  is the transition state or barrier maxima of  $V_{eff}$ . If the population stochastically exceeds  $N_{critical}$ , then the pull of drivers becomes larger than the pull of passengers and the population expands rapidly. Populations smaller than  $N_{critical}$  are overpowered by passengers, so they most often go extinct or regress. If all passengers are neutral ( $s_p=0$ ), the barrier is absent and tumors grow faster than exponentially.  $N_{critical}$  depends upon the ratio of the target sizes  $T_p/T_d$  and selection coefficients of drivers and passengers  $s_p/s_d^2$ . A large difference in the target sizes for passengers and drivers  $T_p/T_d \approx 10^3 - 10^5$  (see main text) and a smaller difference in selection coefficients  $s_p/s_d^2 \approx 10^{-3} - 10^0$  renders a larger  $N_{critical}$ .



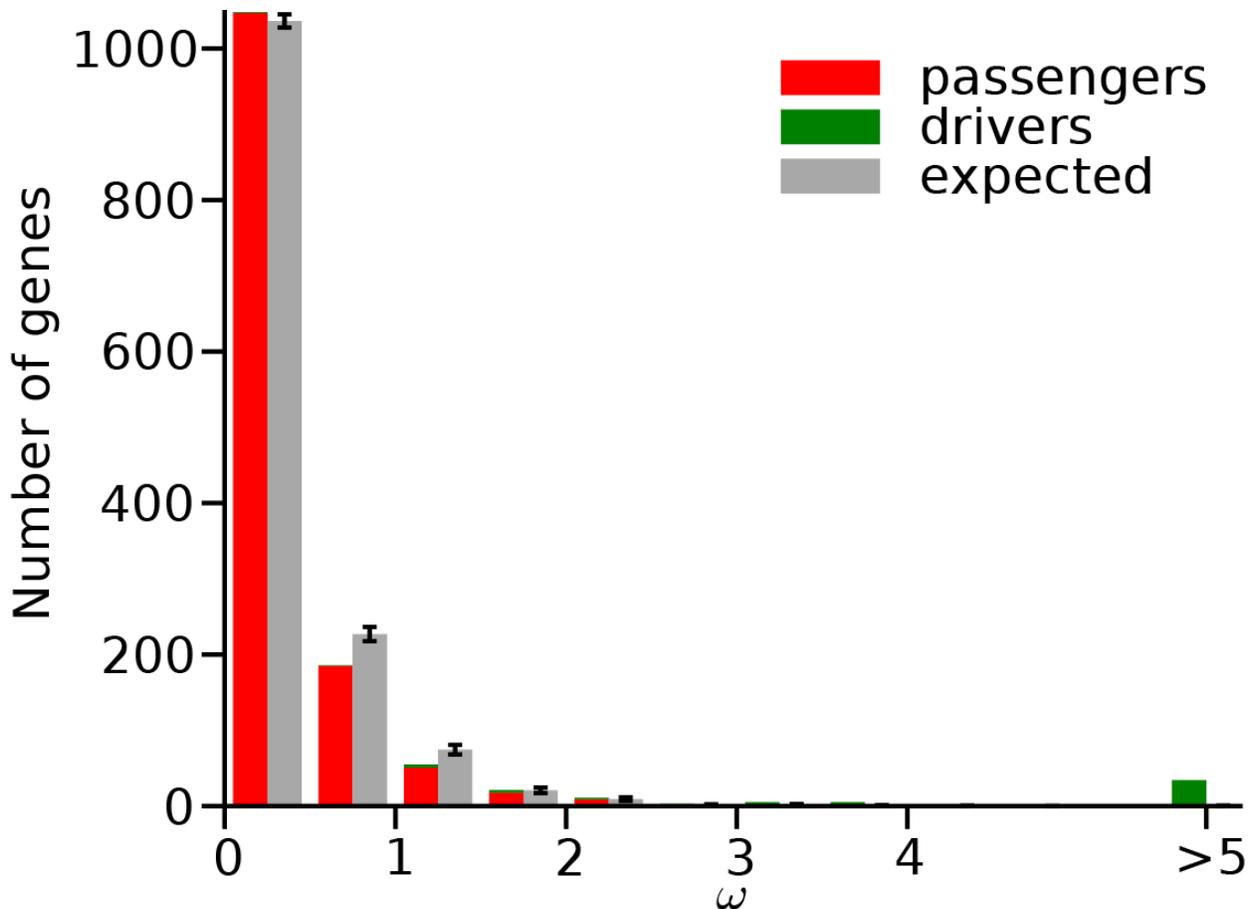
**Fig. S4. Analytical estimates of passenger accumulation.** To test the accuracy of our three analytical approximations of passenger accumulation: Muller’s Ratchet, a traveling wave, and a neutral model (drift); we simulated cancer populations with estimated parameters (**Table S1**), except driver mutations were impossible. All simulations progressed to extinction via passenger accumulation, but did so with varying time. By assuming  $N(p) = K(1 + s_p)^{-p}$  and that a population goes extinct once  $N = 15$  (i.e. stochastic fluctuations kill very small populations), we could analytically estimate the time to extinction. Our analytical approximations agree with the simulations for their regimes of validity (heavy lines). For completeness, the predictions of the analytical results outside of their applicability regions are also show, but with light lines.



**Fig. S5. Balance between mutation and selection in the absence of drivers.** The distribution of passengers per cell approaches a steady-state Poisson distribution with mean and variance  $\lambda = \mu_p/s_p$  provided  $Ne^{-\lambda s_p} \gg 1$ . A cell transitions to a fitness class with one additional passenger every time a mutation arises. This pushes the distribution to the left (lower fitness), while natural selection favors fitter haplotypes and pulls the distribution to the right.



**Fig. S6. Full fitness distribution of deleterious mutations in cancer.** While our model focused specifically on deleterious mutations of intermediate effect, passenger mutations may have phenotypes with a wide range of magnitudes (5). Nevertheless, analysis of human variation suggests that most possible missense mutations in human populations are deleterious, as defined by HGMD (22). Although it may be difficult to distinguish between mildly and strongly deleterious mutations, only approximately 20% of gene knockouts in yeast, a single cell eukaryote like cancer, are lethal (23). Our estimate of  $T_p = 5,000,000$  attempts to accommodate these considerations. The boundaries, between what we consider effectively neutral, moderately deleterious, and strongly deleterious, are defined by the point at which the combined effects of accumulated passengers is negligible compared to the effects of a driver ( $T_{cancer} \mu_p s_p \ll s_d$ ) and the point at which selection against passengers prevents them from efficiently fixing in the population ( $N s_p e^{-\mu_p/s_p} \approx 1$ ). Note  $T_{cancer}$  represents the time to cancer.



**Fig. S7. Cancer mutations show evidence of positive and negative selection.** A histogram of the number of genes, under positive selection ( $\omega > 1$ ) and negative selection ( $\omega < 1$ ). Genes classified by COSMIC as passengers (red) generally experienced neutral or negative selection, while genes classified by COSMIC as putative drivers (green) generally experienced positive selection. Because of the very small number of observed mutations within each gene, most genes are expected to have values of  $\omega$  outside of 1 even under neutral evolution simply by chance (grey); nevertheless, the observed distribution has an inordinate number of genes with  $\omega < 1$  as well as an inordinate number above 1. This suggests that there exist both passenger genes under very mild negative selection as well as driver gene under strong positive selection. Most of the genes used in *PSIC* analysis contained only 1 nonsynonymous mutation and could not be included in this histogram. Because some publications do not report synonymous mutations and because the above distribution is most likely a convolution of some genes under negative selection and others under positive selection, we suspect that the true extent of purifying selection in cancer may be greater than suggested by this analysis of the COSMIC database.

$$\omega = \frac{\omega_{observed}}{\omega_{expected}} = \frac{O_{nonsynonymous}/O_{synonymous}}{E_{nonsynonymous}/E_{synonymous}}$$

Here  $O$  represents observed mutations in a gene (synonymous or nonsynonymous) and  $E$  represents the expected mutations in the gene using our 3-parameter random model of mutagenesis, if the gene were to experience no evolution. We generated expected histograms, by binomially sampling

nonsynonymous and synonymous mutations from each gene using  $\frac{E_{nonsynonymous}}{E_{nonsynonymous} + E_{synonymous}}$  as the probability of a nonsynonymous mutations. The number of trials in each gene sampling was constrained to equal the number of observed mutations and we discarded trials with no synonymous mutations—just as we did for observed data.

Parameter	Symbol	Estimate	Range	Citation
Mutation rate	$\mu$	$10^{-8}$	$10^{-10}$ - $10^{-7}$	(24)
Driver Loci	$T_d$	700	70-7,000	(2, 6, 17)
Passenger Loci	$T_p$	$5 \times 10^6$	$5 \times 10^5$ - $5 \times 10^7$	(8, 25)
Driver strength	$s_d$	0.1	0.001-1	(2, 26)
Passenger strength	$s_p$	0.001	$10^{-4}$ - $10^{-1}$	(4, 25)
Initial Population Size	$K$	1000*	100-10,000	(10)

**Table S1. Parameters of model and estimated range.** Our model contains 5 independent variables that were estimated from the literature and explored across a range of values ( $\mu$ ,  $T_d$  and  $T_p$  can be abstracted to a genome-wide driver and passenger mutation rate). \*Estimated from labeled populations in mice colonic crypts 2 weeks after an induced initiating *APC* deletion.

Phenomenon observed in our model	Experimental Observation
Clonal expansion, delayed growth, and extinction	(27, 28)
More mutations accumulate with high mutation rate	(9)
Approximately 50-300 deleterious mutations acquired under realistic parameters	(29, 30)
Tumors cells have a large degree of heterogeneity in growth rate, yet driver mutations fixate clonally	(31–33)
Mutagenic therapies often relapse after a period of remission	(34)

**Table S2. The deleterious passenger model reproduces many properties of cancer.** Many of the above phenomena would not be observed in our model without the inclusion of deleterious passengers. None of the above phenomena were pre-programmed into the model of neoplastic growth (i.e. population size was not fixed, nor was the number of mutations).