

# Genomic clustering of fitness-affecting mutations favors the evolution of chromosomal instability

Yevgeniy Raynes  | Daniel M. Weinreich

Department of Ecology and Evolutionary Biology, Center for Computational Molecular Biology, Brown University, Providence, Rhode Island

## Correspondence

Yevgeniy Raynes, Department of Ecology and Evolutionary Biology, Center for Computational Molecular Biology, Brown University, Providence, RI  
Email: yevgeniy\_raynes@brown.edu

## Funding information

Center for Computational Molecular Biology at Brown University

## Abstract

Most solid cancers are characterized by chromosomal instability (CIN)—an elevated rate of large-scale chromosomal aberrations and ploidy changes. Chromosomal instability may arise through mutations in a range of genomic integrity loci and is commonly associated with fast disease progression, poor prognosis, and multidrug resistance. However, the evolutionary forces promoting CIN-inducing alleles (hereafter, CIN mutators) during carcinogenesis remain poorly understood. Here, we develop a stochastic, individual-based model of indirect selection experienced by CIN mutators via genomic associations with fitness-affecting mutations. Because mutations associated with CIN affect large swaths of the genome and have the potential to simultaneously comprise many individual loci, we show that indirect selection on CIN mutators is critically influenced by genome organization. In particular, we find strong support for a key role played by the spatial clustering of loci with either beneficial or deleterious mutational effects. Genomic clustering of selected loci allows CIN mutators to generate favorable chromosomal changes that facilitate their rapid expansion within a neoplasm and, in turn, accelerate carcinogenesis. We then examine the distribution of oncogenic and tumor-suppressing loci in the human genome and find both to be potentially more clustered along the chromosome than expected, leading us to speculate that human genome may be susceptible to CIN hitchhiking. More quantitative data on fitness effects of individual mutations will be necessary, though, to assess the true levels of clustering in the human genome and the effectiveness of indirect selection for CIN. Finally, we use our model to examine how therapeutic strategies that increase the deleterious burden of genetically unstable cells by raising either the rate of CIN or the cost of deleterious mutations affect CIN evolution. We find that both can inhibit CIN hitchhiking and delay carcinogenesis in some circumstances, yet, in line with earlier work, we find the latter to be considerably more effective.

## KEYWORDS

carcinogenesis, chromosomal instability, indirect selection, mutator

## 1 | INTRODUCTION

Genomic instability is a hallmark of carcinogenesis and may account for much of the extensive genetic heterogeneity of solid tumors (Bedard, Hansen, Ratain, & Siu, 2013; Burrell, McGranahan, Bartek, & Swanton, 2013; Marusyk, Almendro, & Polyak, 2012; Salk, Fox, & Loeb, 2010; Vogelstein et al., 2013). Genomic instability can be categorized as nucleotide instability (NIN), microsatellite instability (MSI), or chromosomal instability (CIN). Nucleotide instability manifests as an elevated rate of single nucleotide alterations and arises from defects in the nucleotide and base excision repair pathways (Pikor, Thu, Vucic, & Lam, 2013). Microsatellite instability is associated with an elevated rate of nucleotide mismatches and short deletions and insertions (Boland & Goel, 2010; Thibodeau, Bren, & Schaid, 1993) and usually results from defects in the mismatch repair system (Vilar & Gruber, 2010). Most commonly, solid tumors exhibit chromosomal instability—an elevated rate of large chromosomal aberrations, such as somatic copy number alterations (SCNAs) and aneuploidies (Bakhom & Compton, 2012; Hanahan & Weinberg, 2011; Heng et al., 2013). Unlike NIN or MSI, molecular mechanisms of CIN—especially in sporadic cancers—remain poorly defined. Mutations in a number of different pathways have been implicated in CIN, including, among others, mitotic checkpoints (Cahill et al., 1998), chromatid cohesion (Solomon et al., 2011), and double-strand break repair (Lord & Ashworth, 2012). Chromosomal instability has been observed in most high mortality cancers including colon (Fearon, 2011; Lengauer, Kinzler, & Vogelstein, 1998), breast (Kwei, Kung, Salari, Holcomb, & Pollack, 2010), and lung (Masuda & Takahashi, 2002) and has been linked with poor prognosis (Carter, Eklund, Kohane, Harris, & Szallasi, 2006; McGranahan, Burrell, Endesfelder, Novelli, & Swanton, 2012; Walther, Houlston, & Tomlinson, 2008) and multidrug resistance (Lee et al., 2011). Despite its clinical importance, the evolutionary forces favoring the emergence of CIN and its role in cancer progression have long been a subject of debate in the literature and remain poorly understood (Cahill, Kinzler, Vogelstein, & Lengauer, 1999; Datta, Gutteridge, Swanton, Maley, & Graham, 2013; Loeb, 2011; Michor, 2005; Negrini, Gorgoulis, & Halazonetis, 2010; Nowak et al., 2002).

Current theories for the evolution of CIN, and, in fact, the evolution of genomic instability in general, usually invoke two distinct, yet not mutually exclusive, mechanisms of selection on instability-inducing mutator alleles. Chromosomal instability mutators may intrinsically raise a neoplastic cell's chances of survival and reproduction, that is, its Darwinian fitness. As a result, CIN mutators may be directly favored by natural selection. For example, it has been suggested that CIN may originate from directly beneficial oncogenic mutations that simultaneously induce genomic instability [the oncogene-induced DNA replication stress hypothesis] (Halazonetis, Gorgoulis, & Bartek, 2008; Negrini et al., 2010).

Alternatively, mutators may be favored not for their own intrinsic effects on a cell's fitness but through genetic association with intrinsically beneficial mutations elsewhere in the genome. In other words, CIN mutators may evolve via so-called indirect selection by hitchhiking (Smith & Haigh, 1974) with selectively favored oncogenic

and tumor-suppressing mutations they generate (Loeb, 2001; Sprouffske, Merlo, Gerrish, Maley, & Sniegowski, 2012). Indirect selection on alleles that increase the genomic mutation rate (e.g., CIN-inducing mutators) is expected to be particularly effective in asexual populations, such as cancers, in which the genetic association between mutators and beneficial mutations can never be disrupted by recombination. In fact, numerous theoretical and experimental studies in microbes have already shown that mutators may spread through non-recombining populations by hitchhiking if beneficial mutations are readily available (reviewed in: Raynes & Sniegowski, 2014; Sniegowski, Gerrish, Johnson, & Shaver, 2000).

Importantly, while genomic instability may increase the rate of beneficial mutations, it necessarily also increases the rate of deleterious mutations, which are generally more common (Cahill et al., 1999). However, alleles that elevate the point mutation rate (i.e., NIN- or MSI-inducing mutators) generate DNA changes confined to only a few nucleotides. As a result, beneficial and deleterious mutations are likely introduced independently from each other by separate mutational events. Thus, while such mutators may be frequently lost to selection against the increased load of deleterious mutations, they may also occasionally expand within a neoplasm by hitchhiking with a rare beneficial mutation. In contrast, CIN mutators generate large-scale SCNAs that are not confined to single loci. Instead, SCNAs may disrupt many neighboring loci, thereby simultaneously introducing both beneficial and deleterious changes. For example, a single SCNA may delete a tumor suppressor and a neighboring housekeeping gene. Genetic linkage between the relatively rare beneficial loci and the more common deleterious ones may drastically limit the availability of SCNAs with net beneficial effects (needed to facilitate CIN hitchhiking) and, thus, inhibit CIN evolution.

We have hypothesized that indirect selection could, nevertheless, favor CIN given a spatial organization of the genome that minimizes co-occurrence of beneficial and deleterious loci in SCNAs. Specifically, we wanted to test the hypothesis that CIN would be favored by indirect selection in genomes in which either beneficial or deleterious loci were spatially clustered along the chromosome. To do so, we developed an individual-based stochastic population model of clonal evolution in spatially organized genomes. In simulation, we investigated the effect of the spatial distribution of beneficial and deleterious loci on CIN evolution and cancer progression. We also tested the effectiveness of therapeutic strategies that aim to raise the deleterious costs of CIN in order to inhibit CIN evolution. Finally, we examined the spatial distributions of candidate human oncogene and tumor suppressor loci (identified in Davoli et al., 2013) for evidence of spatial clustering that could facilitate indirect selection for CIN in real cancers.

## 2 | METHODS

### 2.1 | Stochastic simulations

To model the evolutionary progression of a neoplastic cell population to cancer, we developed and simulated an individual-based

computational model of clonal evolution based on the earlier work of Beerenwinkel et al. (2007) and Datta et al. (2013). Like these earlier studies, we employed a Wright–Fisher model (Ewens, 2004), in which a population of neoplastic cells evolves in discrete, non-overlapping generations with the probability of each cell's reproduction being proportional to its relative fitness. As in earlier studies, cells in the model could acquire beneficial and deleterious fitness-affecting mutations upon reproduction. Unlike these earlier studies, however, we also allowed for the evolution of CIN by adding SCNA-type mutations (described below) and introducing CIN mutator alleles.

As in the work of Beerenwinkel et al. (2007) and Datta et al. (2013), simulations here start with a tumor population of initial size  $N_0 = 10^6$  cells and end when the tumor develops into a cancer. Following Datta et al. (2013), we defined cancer as a tumor in which 10% of all cells have acquired at least 20 beneficial mutations.

The total size of the neoplastic population is constrained to grow exponentially at a rate proportional to the average fitness of the population. The size of the population at generation  $t + 1$  is defined as  $N_{t+1} = N_t \cdot (1 + \bar{w}\beta)$ , where  $\bar{w}$  is the average fitness of the tumor population (see below) and  $\beta$  is a constant that governs the rate of population growth; as in Datta et al. (2013), we set  $\beta = 0.0016$ .

Neoplastic cells can acquire single-locus mutations that either change fitness or induce CIN. Chromosomal instability-inducing mutations result in a mutator phenotype which allows cells to generate SCNA mutations (described below). In order to explicitly explore the role of indirect selection and genetic hitchhiking in the evolution of CIN, CIN-inducing mutations are assumed to have no intrinsic direct effect on a cell's fitness. Furthermore, a single CIN mutator mutation is assumed to be sufficient for the mutator phenotype; additional CIN mutations have no effect on the mutation rate and so cannot affect the dynamics of the mutator lineage.

Among the fitness-affecting mutations, beneficial mutations increase a cell's fitness by  $s_{\text{ben}}$  while deleterious mutations decrease a cell's fitness by  $s_{\text{del}}$ . Since  $s_{\text{ben}}$  and  $s_{\text{del}}$  are held constant, fitness of a cell with  $x$  beneficial mutations and  $y$  deleterious mutations can be computed as  $w_{xy} = 1 + x \cdot s_{\text{ben}} - y \cdot s_{\text{del}}$ . Correspondingly, if  $f_{xy}$  is the fraction of cells with  $x$  and  $y$  beneficial and deleterious mutations, respectively, in a tumor, then the average fitness of the tumor population is  $\bar{w} = \sum_x \sum_y f_{xy} w_{xy}$ . Unlike in the earlier work (Beerenwinkel et al., 2007; Datta et al., 2013), the effect of multiple mutations in our model is additive rather than multiplicative. Note that there is little difference between additive and multiplicative fitness for genotypes containing a small number of mutations. (Algebraically,  $(1 + s_{\text{ben}})^x (1 - s_{\text{del}})^y \approx 1 + x \cdot s_{\text{ben}} - y \cdot s_{\text{del}}$  for small  $x$  and  $y$ .) However, for genotypes containing many mutations (like some CIN genotypes in our model), multiplicative fitness, which increases exponentially with the number of mutations, results in unrealistically high values.

The tumor population is composed of genetic lineages of neoplastic cells defined by the counts of fitness-affecting and CIN mutator mutations they carry. As per the Wright–Fisher model, the size of a lineage with  $x$  beneficial mutations and  $y$  deleterious mutations at generation  $t + 1$  is drawn from a multinomial distribution with

expectation given by  $N_{t+1} f_{xy} w_{xy} / \bar{w}$ , where  $f_{xy}$  is the frequency of the lineage in generation  $t$ , and  $w_{xy} / \bar{w}$  is its relative fitness.

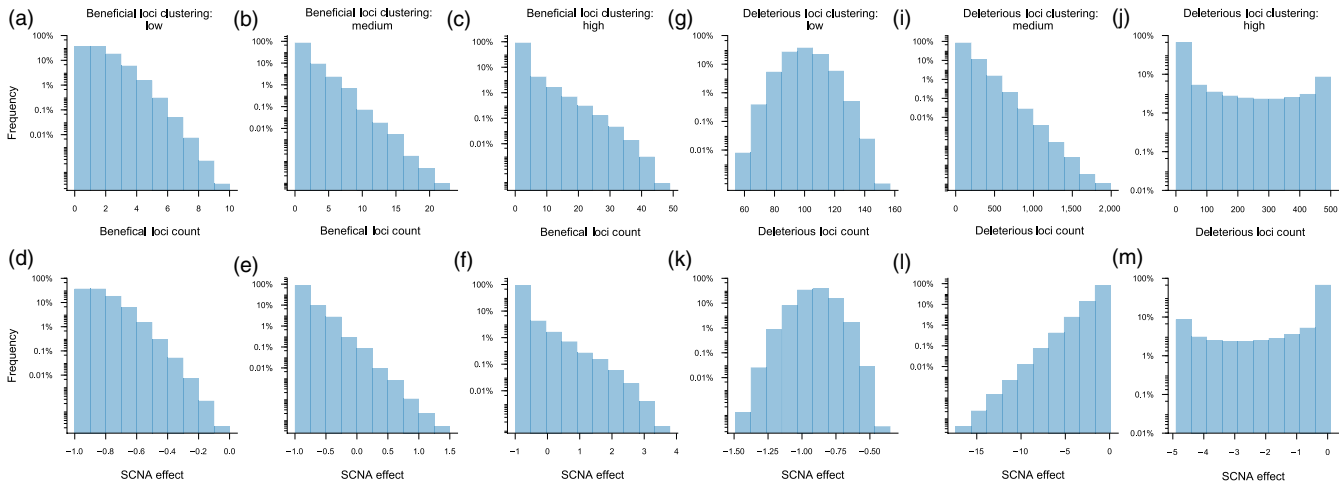
Upon reproduction, every surviving lineage acquires a random number  $M_{\text{ben}}$  of beneficial mutations,  $M_{\text{del}}$  of deleterious mutations, and  $M_{\text{CIN}}$  of mutator mutations drawn from a multinomial distribution with expectations given by  $N_i U_{\text{ben}}$ ,  $N_i U_{\text{del}}$ , and  $N_i U_{\text{CIN}}$ , respectively. Here,  $N_i$  is the size of the lineage, and  $U_{\text{ben}}$ ,  $U_{\text{del}}$ , and  $U_{\text{CIN}}$  are the per cell rates of beneficial, deleterious, and CIN mutator mutations. Thus, as a result of mutation, a lineage with  $x$  beneficial and  $y$  deleterious mutations gives rise to a new lineage with  $M_{\text{ben}}$  individuals carrying  $x + 1$  beneficial mutations, a new lineage with  $M_{\text{del}}$  individuals carrying  $y + 1$  deleterious mutations, and a third lineage with  $M_{\text{CIN}}$  individuals carrying the CIN mutator allele.

Genetic lineages carrying CIN mutator mutations also acquire a Poisson-distributed number of SCNAs,  $M_{\text{SCNA}}$  with mean  $N_i U_{\text{SCNA}}$ , where  $U_{\text{SCNA}}$  is the per cell rate of SCNA production in CIN mutators. Each new SCNA mutation contains a randomly generated number of beneficial,  $B_{\text{SCNA}}$ , and a randomly generated number of deleterious mutations,  $D_{\text{SCNA}}$ . Each new SCNA mutation in a lineage with  $x$  beneficial and  $y$  deleterious mutations, thus, gives rise to a new lineage with a single individual carrying  $x + B_{\text{SCNA}}$  beneficial and  $y + D_{\text{SCNA}}$  deleterious mutations.

To model spatial clustering of fitness-affecting loci, we use discrete distributions with a constant mean number of mutations ( $\mu$ ) but increasing variance ( $\sigma^2$ ) to draw the values of  $B_{\text{SCNA}}$  and  $D_{\text{SCNA}}$  for each new SCNA. Thus, we assume the infinite sites model for new SCNAs. Note that while the available supply of large chromosomal aberrations, such as aneuploidies, in real cancers is relatively small, shorter SCNAs will be plentiful. In fact, the supply of SCNAs containing about 100 loci (~0.5% of the human genome) like the ones in our model should be extremely high, justifying our assumption of the infinite sites model. Furthermore, consider that in our model a tumor samples and substitutes only an infinitesimal fraction of the available SCNA distribution, similar to real cancers that have been shown to contain a median number of only about 20 SCNAs (Beroukhi et al., 2010).

Because we assume the infinite sites model for new SCNAs, spatially organized genomes are not explicitly modeled, which allows increased computational efficiency. Instead, clustering of loci in the model is controlled by the variance in the distributions of  $B_{\text{SCNA}}$  and  $D_{\text{SCNA}}$ . As can be seen from Figure 1, distributions of  $B_{\text{SCNA}}$  and  $D_{\text{SCNA}}$  characterized by lower variance result in less clustering of fitness-affecting loci across potential SCNAs with most SCNAs containing relatively similar numbers of loci. On the other hand, higher variance in either  $B_{\text{SCNA}}$  or  $D_{\text{SCNA}}$  distributions produces more clustering in fitness-affecting loci across potential SCNAs with relatively few SCNAs containing large clusters of loci and many SCNAs without any.

The model was parameterized as follows. Based on earlier theoretical studies of Beerenwinkel et al. (2007), Bozic et al. (2010), Datta et al. (2013), and McFarland, Korolev, Kryukov, Sunyaev, and Mirny (2013), we set  $U_{\text{ben}} = 10^{-5}$ . Because estimates of the deleterious mutation rate and effects are limited, we set  $s_{\text{del}} = 0.01$  and we assumed



**FIGURE 1** Spatial clustering of fitness-affecting loci in the model. Beneficial loci of effect  $s_{\text{ben}} = 0.1$  were distributed across SCNAs using either Poisson (a; low clustering), geometric (b; intermediate clustering), or beta-binomial (c; high clustering) distributions with  $\mu = 1$ . Deleterious loci of effect  $s_{\text{del}} = 0.01$  were distributed using the Dirac delta function with  $\mu = 100$ . These distributions of beneficial loci produced the distributions of SCNA fitness effects shown in panels d, e, and f, respectively, calculated as  $w_{\text{SCNA}} = B \cdot s_{\text{ben}} - D \cdot s_{\text{del}}$  (where  $B$  and  $D$  are the counts of beneficial and deleterious mutations, respectively). Likewise, deleterious mutations were distributed across SCNAs using Poisson (g; low clustering), geometric (i; intermediate clustering), or beta-binomial (j; high clustering) distributions with  $\mu = 100$ . Beneficial loci were distributed using the Dirac delta function with  $\mu = 1$ . These distributions of deleterious loci produced the distributions of SCNA fitness effects shown in panels k, l, and m, respectively, calculated as above. See Methods for distribution details

that deleterious mutations outnumber the beneficial ones one hundred-fold and set  $U_{\text{del}} = 10^{-3}$ . Note that higher  $U_{\text{del}}$  or  $s_{\text{del}}$  would raise the deleterious load of SCNAs, reinforcing the importance of genomic clustering in promoting CIN; on the other hand, lower  $U_{\text{del}}$  or  $s_{\text{del}}$  would lessen the role of genomic clustering in facilitating CIN evolution.  $U_{\text{CIN}}$  has no effect on the role of genomic clustering and was set at  $U_{\text{CIN}} = 10^{-5}$ . We set  $U_{\text{SCNA}} = 0.01$  SCNA mutations per cell per generation after Lengauer, Kinzler, and Vogelstein (1997) in lineages carrying CIN mutators. Finally, we set  $s_{\text{ben}} = 0.1$ , based on the estimates of beneficial effects of  $\sim 0.004$  to  $\sim 0.6$  obtained from earlier theoretical studies (Beerenwinkel et al., 2007; Bozic et al., 2010; McFarland, Mirny, & Korolev, 2014).

The mean number of mutations ( $\mu$ ) of all  $B_{\text{SCNA}}$  distributions is set to 1 (i.e., on average a single beneficial mutation per SCNA), while  $\mu$  of all  $D_{\text{SCNA}}$  distributions is set to 100: The ratio of the two is thus the same as the ratio of  $U_{\text{ben}}$  and  $U_{\text{del}}$  above. To generate  $B_{\text{SCNA}}$  values, we use either a Dirac delta function with a single value (i.e., no clustering,  $\mu = 1$ ,  $\sigma^2 = 0$ ), a Poisson distribution (minimal clustering,  $\mu = 1$ ,  $\sigma^2 = 1$ ; Figure 1a), a geometric distribution (intermediate clustering,  $\mu = 1$ ,  $\sigma^2 = 2$ ; Figure 1b), or a beta-binomial distribution (high clustering, parameters of the beta distribution:  $n = 50$ ,  $\alpha = 0.1$ ,  $\beta = 4.9$ ;  $\mu = 1$ ,  $\sigma^2 \approx 8.98$ ; Figure 1c). Similarly,  $D_{\text{SCNA}}$  is randomly drawn from either a Dirac delta function ( $\mu = 100$ ,  $\sigma^2 = 0$ ), a Poisson distribution ( $\mu = 100$ ,  $\sigma^2 = 100$ ; Figure 1g), a geometric distribution ( $\mu = 100$ ,  $\sigma^2 = 10,100$ ; Figure 1i), or a beta-binomial distribution (parameters of the beta distribution:  $n = 500$ ,  $\alpha = 0.1$ ,  $\beta = 0.4$ ;  $\mu = 100$ ,  $\sigma^2 \approx 26,693$ ; Figure 1j). Figure 1 also illustrates how changing the variance in the physical distribution of beneficial and deleterious loci affects the distribution of available SCNA fitness effects (calculated for an SCNA with  $B$  beneficial and

$D$  deleterious mutations as  $w_{\text{SCNA}} = B \cdot s_{\text{ben}} - D \cdot s_{\text{del}}$ ). It is important to note that while the expected number of loci in an SCNA is constant for all distributions (100 deleterious plus 1 beneficial), the realized number of loci in an SCNA is not constant unless both  $B_{\text{SCNA}}$  and  $D_{\text{SCNA}}$  are drawn from the Dirac delta function. Consequently, at higher variance in either  $B_{\text{SCNA}}$  or  $D_{\text{SCNA}}$ , some SCNAs may comprise more loci than others. Moreover, while we model SCNAs of a given mean size (101 loci), it is certainly not the case that SCNAs of such size are representative of actual SCNAs in cancer (in fact, based on Beroukhi et al. (2010), it is likely that most focal SCNAs comprise fewer loci). We explore these assumptions in the simulation in Supporting Information Figures S1 and S2, respectively, and find that the effect of clustering on CIN evolution is robust to both of them.

Since all simulations end with cancer, to assess the influence of genomic clustering and CIN evolution on carcinogenesis, we calculate the mean waiting time required for a tumor to evolve into cancer. To assess whether CIN mutators are favored by indirect selection or not, we calculate the mean waiting time to CIN establishment. We assume that a mutator mutation becomes established if it reaches the frequency of 10% of the population. We then compare the probability of mutator establishment ( $P_{10\%}^{\text{CIN}}$ ) to the probability of a neutral mutation reaching the frequency of 10% ( $P_{10\%}^{\text{neutral}}$ ).  $P_{10\%}^{\text{neutral}}$  is calculated as the frequency (over  $10^6$  runs of simulation) at which a neutral mutation appearing at the same rate as the mutator ( $U_{\text{CIN}}$ ) reaches 10% of the population in control simulations without CIN. Correspondingly, mutators are favored by indirect selection when  $P_{10\%}^{\text{CIN}} > P_{10\%}^{\text{neutral}}$  and disfavored when  $P_{10\%}^{\text{CIN}} < P_{10\%}^{\text{neutral}}$  (e.g., Raynes, Wylie, Sniegowski, & Weinreich, 2018).

## 2.2 | Genomic analysis

To examine the spatial distribution of the carcinogenic mutations in the human genome, we focused on the candidate oncogene and tumor suppressor loci identified by the TUSON algorithm of Davoli et al. (2013). The TUSON algorithm predicts the likelihood that a given gene acts as a tumor suppressor or an oncogene based on its mutational status in sequenced tumors versus normal tissue. To identify the most reliable mutational parameters for prediction of tumor suppressors and oncogenes, Davoli et al. compiled a data set of ~1,200,000 mutations from more than 8,200 tumor samples of >20 different tumor types and developed 22 different parameters based on the different classes of mutations. The algorithm was tested on three different training sets of known tumor suppressors and oncogenes from the Cancer Gene Census (Futreal et al., 2004). In the original study, TUSON predictions were used to rank every gene in the genome based on its potential as a tumor suppressor or an oncogene (Davoli et al., 2013). For our analysis, we used the top 300 tumor suppressors and 249 oncogenes identified by TUSON in the original study. One of the potential tumor suppressors (C3orf27) from the original top 250 list of Davoli et al. have since been shown to be a long intergenic non-protein-coding RNA and was excluded from the present analysis. As in Davoli et al., each potential tumor suppressor and oncogene was assigned a weight, hereafter  $w_{\text{Davoli}}$  (to differentiate from fitness notation in our model), calculated as  $w_{\text{Davoli}} = T - r$  where  $T$  is the total number of genes in the respective list (300 for suppressors and 249 for oncogenes) and  $r$  is the rank of that gene in the list.  $w_{\text{Davoli}}$  can thus be thought as proxy for the potential fitness effect of a gene (most potent tumor suppressors and oncogenes have the highest  $w_{\text{Davoli}}$  scores).

Then, also as in Davoli et al., we quantified the tumor-suppressive and oncogenic potentials of human SCNAs based on the density and potency ( $w_{\text{Davoli}}$ ) of tumor suppressors and oncogenes they encompass. For SCNAs the size of a chromosome or a chromosome arm we used Chrom and Charm scores provided in the study of Davoli et al. These scores were calculated as the sums of weights ( $w_{\text{Davoli}}$ ) of tumor suppressors or oncogenes contained in each chromosome or a chromosome arm divided by the total number of genes contained in that chromosome or arm. In other words, Chrom and Charm scores represent the potential beneficial effect that SCNAs comprising a given chromosome or chromosome arm may have on a neoplastic cell's fitness.

For focal SCNAs shorter than a chromosome arm, we employed a sliding window approach. For computational efficiency, we advanced the window from the beginning of each chromosome arm, 10,000 nucleotides at a time, up until the centromere or the end of the chromosome was reached. The score of each window was calculated as the sum of weights of oncogenes or tumors suppressors contained in the window divided by the total number of genes in the window. To quantify spatial clustering of oncogenes and tumor suppressors across chromosomes, chromosome arms, and focal SCNAs, we calculated the variance in the distributions of Chrom, Charm, and focal SCNA scores of each size.

To assess whether human oncogenes and tumor suppressors are clustered more than random, we used a permutation approach. To create each random permutation of tumor suppressors and oncogenes, weight ( $w_{\text{Davoli}}$ ) values from 1 to 300 (for the tumor suppressor distribution) or from 1 to 249 (for the oncogene distribution) were assigned to 300 (or 249) human loci, randomly sampled from the genome. For each such random permutation, we calculated the Chrom, Charm, and focal SCNA scores as above. We then calculated the variance in the distributions of Chrom, Charm, and focal SCNA scores for each permutation. A total of 10,000 permutations of weights were performed resulting in distributions of 10,000 variance values for each size of potential SCNAs. Observed variance in the distribution of Chrom, Charm, and focal scores of all sizes for tumor suppressors and oncogenes was then compared to the null distributions obtained by permutation. Observed variances were considered to be significantly higher than expected if they were greater than 95% of variances sampled by permutation.

Ensemble database, release 84 (Yates et al., 2016), was used to obtain the genomic locations of all tumor-suppressing and oncogenic genes identified by TUSON as well as the list of 20,837 human genes using the following two gene types to filter the results: protein-coding genes and TEC (potential protein-coding genes that require experimental confirmation). Genes on the Y chromosome were excluded from the analysis. The UCSC human genome browser (Karolchik et al., 2004) (Assembly Dec 2013, GRCh/hg38) was used to obtain the lengths of all human chromosomes and the locations of centromeres.

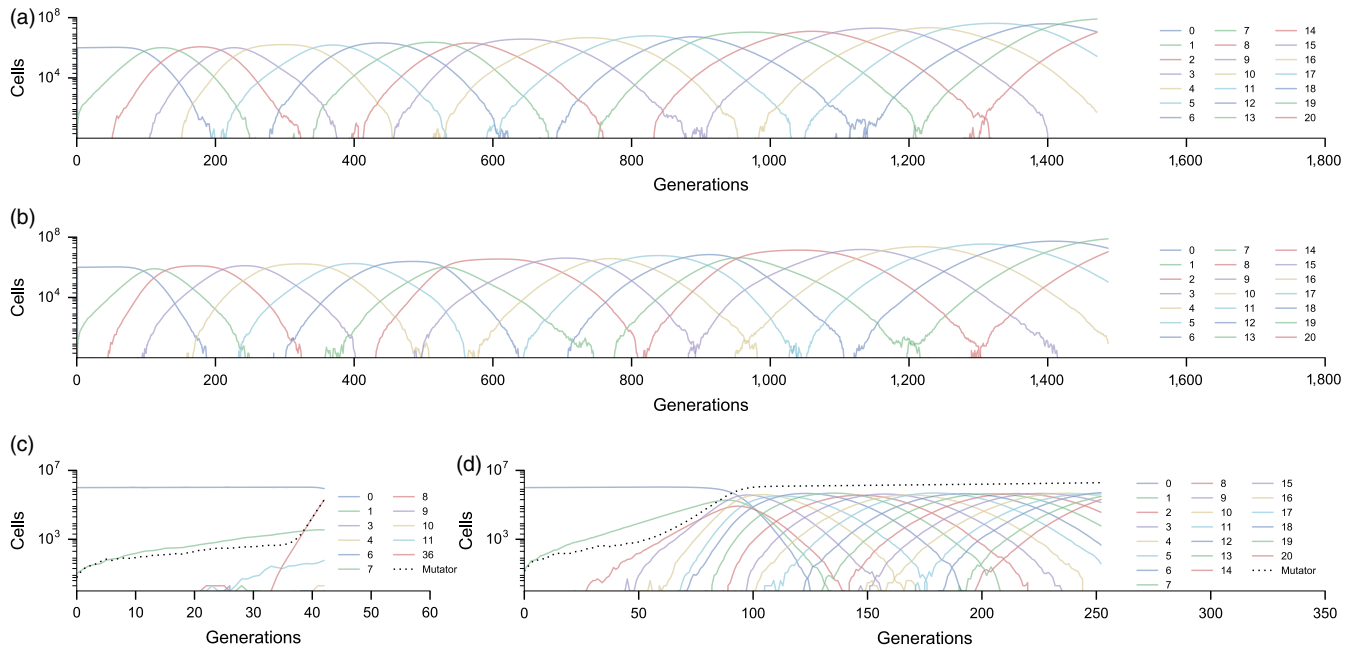
## 3 | RESULTS

### 3.1 | Spatial clustering of fitness-affecting loci across SCNAs promotes the evolution of CIN

Figure 2 illustrates the dynamics of tumor evolution in representative runs of the simulation with and without CIN mutator loci in the genome (note the difference in x-axis scale between panels). Consistent with earlier studies (Beerenwinkel et al., 2007; Datta et al., 2013), in the absence of CIN mutators, our model exhibits a "traveling wave" of clonal expansions and contractions (Figure 2a). Clones with increasing numbers of beneficial mutations successively expand and replace less adapted clones until a clone with 20 beneficial mutations appears and expands to 10% of the population (at which point the simulation halts). Importantly, each clonal expansion is driven by only a single additional beneficial mutation.

When CIN mutators are added to the genome, the dynamics of tumor evolution become strongly dependent on the spatial distribution of beneficial and deleterious loci across potential SCNAs. Recall that in our model, CIN mutators themselves do not affect fitness and only experience indirect selection via genetic association with SCNA mutations they generate. When neither beneficial nor deleterious loci are spatially clustered, that is, distributed across SCNAs with no variance, all available SCNAs are deleterious to fitness (Figure 1). As a result, CIN mutators are strongly disfavored by indirect selection





**FIGURE 2** Clonal dynamics in representative simulated populations. Solid lines represent the size of clones with numbers of beneficial mutations indicated. Dashed lines in c and d represent the size of the mutator population within the tumor. (a) Without CIN mutators, (b) with CIN mutators; beneficial and deleterious mutations distributed with no clustering (using the Dirac delta function), (c) with CIN mutators; beneficial mutations distributed with high clustering (using the beta-binomial distribution) and deleterious loci not clustered, (d) with CIN mutators; deleterious mutations distributed with high clustering (beta-binomial) and beneficial loci not clustered. See Methods for distribution details. Model parameters:  $U_{ben} = 10^{-5}$ ,  $U_{del} = 10^{-3}$ ,  $U_{CIN} = 10^{-5}$  per cell per generation,  $s_{del} = 0.01$ ,  $s_{ben} = 0.1$ ,  $U_{SCNA} = 0.01$  SCNA mutations per cell per generation.  $\mu$  for all beneficial loci distributions = 1,  $\mu$  for all deleterious loci distributions = 100

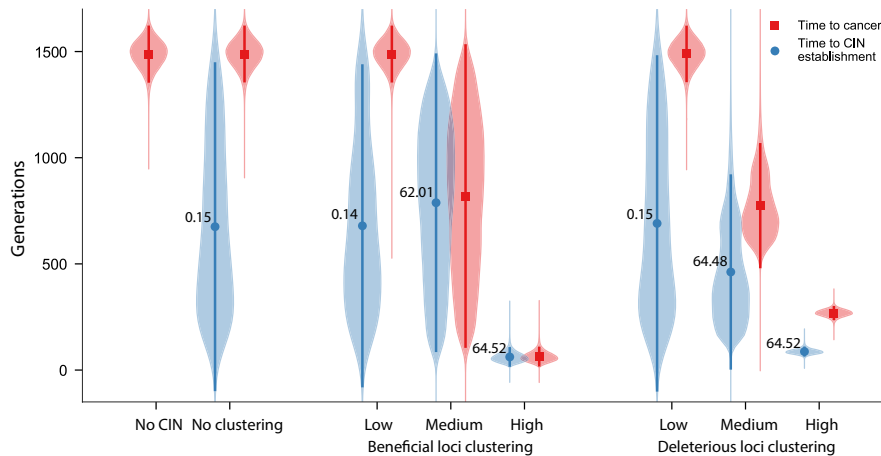
and have virtually no effect on cancer progression, which remains indistinguishable from that in genomes without CIN (Figure 2b).

However, when either beneficial or deleterious loci are highly spatially clustered, that is, distributed with the highest possible variance in the genome, cancer progression is significantly accelerated by CIN, albeit in different ways. In genomes with highly clustered beneficial loci, large numbers of beneficial mutations co-occur in a few SCNAs, resulting in some extremely beneficial SCNAs (Figure 1c,f). Chromosomal instability mutators rapidly increase in frequency by acquiring such SCNAs, which, in turn, allows the tumor to very quickly evolve into cancer by substituting only a few SCNAs. For example, in Figure 2c, a single CIN mutator lineage acquires a cluster of 36 beneficial mutations, which then quickly expands to the threshold frequency of 10%. On the other hand, in genomes with highly clustered deleterious mutations (Figure 1j,m), large numbers of deleterious loci are sequestered into few very harmful SCNAs, which reduces the deleterious load of other available SCNAs. However, spatial clustering of deleterious mutations does not affect the number of beneficial mutations in individual SCNAs. As a result, CIN mutators are still favored by indirect selection but spread by hitchhiking with SCNAs containing fewer than the mean number of deleterious mutations and only a single beneficial mutation. Correspondingly, up to 20 SCNA mutations may substitute in the CIN mutator population before the tumor develops into cancer. Populations with clustered deleterious mutations exhibit very similar traveling wave dynamics as populations without CIN, although

the progression is accelerated once CIN mutators become established (Figure 2d).

Figure 3 summarizes the influence of the spatial distribution of fitness-affecting loci across SCNAs on CIN evolution and the speed of carcinogenesis. As expected from population dynamics seen in Figure 2, cancer progression is slowest in genomes without CIN mutators. Even in genomes with CIN mutators, cancer progression remains unaffected by CIN when both beneficial and deleterious loci are distributed with no or even low clustering across SCNAs. In these genomes, CIN mutators are strongly disfavored by indirect selection ( $p_{10\%}^{CIN} < p_{10\%}^{neutral}$ ) and the waiting time to cancer is not significantly different than in genomes without CIN (Figure 3). However, as beneficial and deleterious loci become increasingly clustered, the probability of mutator establishment raises dramatically above that of a neutral mutation ( $p_{10\%}^{CIN} > p_{10\%}^{neutral}$ ). In other words, CIN mutators switch from being disfavored to being strongly favored by indirect selection. Correspondingly, as selected loci become increasingly clustered, cancer progression is accelerated, with the waiting time to cancer minimized in genomes with the most clustered loci.

Note that in genomes with clustered beneficial loci, the waiting time to CIN establishment is not significantly different from the waiting time to cancer. Here, CIN mutators become established by hitchhiking with SCNAs containing multiple beneficial mutation. As a result, an expanding mutator lineage may already carry enough beneficial mutations for the tumor to become cancerous as soon as it is



**FIGURE 3** Clustering of fitness-affecting loci promotes CIN evolution and accelerates cancer development. Waiting time to CIN establishment (blue) and cancer (red) as a function of genomic organization. For genomes with clustered beneficial mutations ( $\mu = 1$ ), deleterious loci were distributed using the Dirac delta function ( $\mu = 100$ ). For genomes with clustered deleterious mutations ( $\mu = 100$ ), beneficial loci were distributed using the Dirac delta function ( $\mu = 1$ ). The rest of model parameters are as in Figure 2. Circles are mean values calculated over 100,000 runs of simulation (error bars represent  $\pm 95\%$  CI, all times are represented with violin plots).  $P_{10\%}^{CIN} / P_{10\%}^{neutral}$  values next to each time to CIN establishment data point. Mutators are favored by selection when  $P_{10\%}^{CIN} / P_{10\%}^{neutral} > 1$ .  $P_{est}^{neutral} \approx 0.016$  (see Methods). See Methods for more distribution details

established (as seen, for example, in Figure 1c) or can quickly acquire additional mutations in large clusters thereafter. On the other hand, in genomes with clustered deleterious loci, CIN mutators spread by hitchhiking with SCNAs generally containing only a single beneficial mutation. Consequently, even after CIN is established, multiple beneficial mutation must still sweep through the population before cancer evolves. Chromosomal instability mutators do, however, significantly accelerate cancer evolution by rapidly producing the necessary beneficial mutations (albeit still via SCNAs with only a single beneficial mutation each).

### 3.2 | Oncogenic and tumor-suppressive loci in the human genome are distributed with relatively high variance

Simulations show that increased variance in either beneficial or deleterious fitness effects of SCNAs, resulting from increased spatial clustering of fitness-affecting mutations, can promote the evolution of CIN. To assess the variance in beneficial fitness effects of potential SCNAs in the human genome, we examined the spatial distribution of tumor suppressor and oncogene loci identified by the TUSON algorithm, developed by Davoli et al. (2013). Intriguingly, the work of Davoli et al. showed a significant relationship between the genomic distribution and potency of tumor suppressors and oncogenes (summarized as Chrom and Charm scores, see Methods) and patterns of chromosomal deletions and amplifications from sequenced tumors.

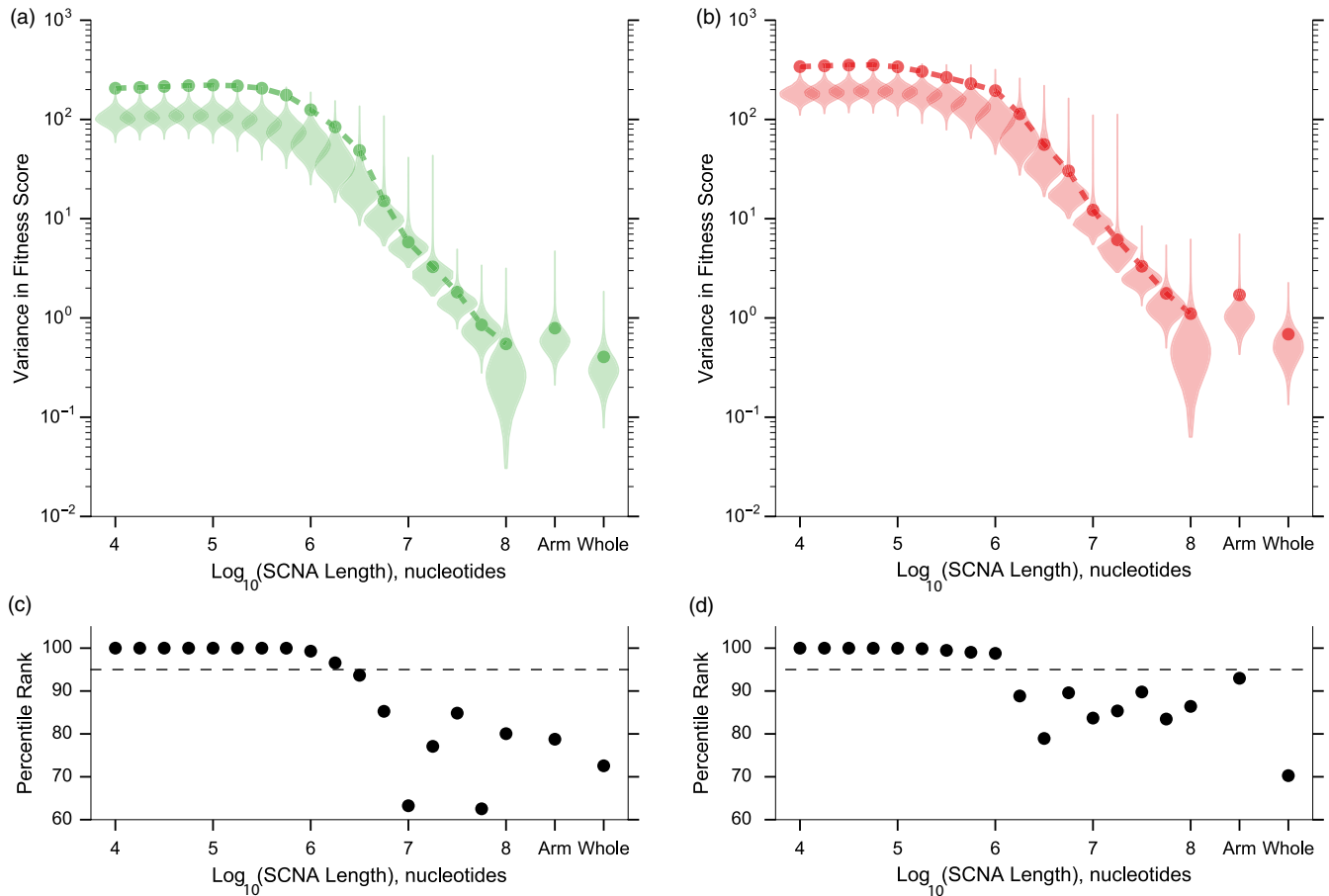
Using the methodology of their study, we evaluated variance in the distribution of oncogenic and tumor-suppressive effects of potential human SCNAs of different size, including focal SCNAs shorter than a chromosome arm, as well as SCNAs the length of a chromosome arm and a whole chromosome (Methods). We observed that variance in the distributions of oncogenic (Figure 4a) and

tumor-suppressive (Figure 4b) effects was maximized for shorter SCNAs and decreased for longer SCNAs. We then evaluated the observed variance using a permutation approach (Methods). We found that the true variance in the distributions of both oncogenes and tumor suppressors was consistently higher than the mean variance of the permuted distributions across all of SCNAs lengths examined and significantly higher (above the 95th percentile) for focal SCNAs shorter than  $\sim 10^6$  nucleotides. Thus, it appears that beneficial oncogenic and tumor-suppressive effects are, in fact, more clustered in potential SCNAs than expected by chance, suggesting that the human genome could be organized in a way that could promote CIN hitchhiking.

### 3.3 | Spatial organization of the genome affects the success of CIN-inhibiting therapies

In simulations, we showed that high variance in the spatial distribution of beneficial loci across SCNAs can promote the evolution of CIN, which in turn can significantly accelerate carcinogenesis. Furthermore, our analysis of the distribution of known oncogenes and tumor suppressors suggested that mutations beneficial to neoplastic cells may be more clustered than random in the human genome. In light of these observations, we investigated whether the evolution of CIN could be inhibited by either increasing the mutation rate of CIN mutators or by exacerbating the effects of individual deleterious loci. Both therapeutic strategies have been previously shown to successfully reduce tumor size by exploiting its deleterious mutational load (McFarland et al., 2013). Correspondingly, we wanted to test whether these strategies could also inhibit CIN evolution by increasing the deleterious load associated with CIN mutators.

Using our model, we assessed the effect of increasing both CIN rate ( $U_{SCNA}$ ) and deleterious mutation effects ( $s_{del}$ ) in genomes



**FIGURE 4** Oncogenes and tumor suppressors are distributed with relatively high variance in the human genome. Variance in the distributions of (a) oncogenic and (b) tumor-suppressive effects of the human SCNAs (circles). Violin plots in a and b: Variances of 10,000 distributions generated by permutation (see Methods). Percentile ranks of observed variances in (c) oncogenic effect distributions and (d) tumor-suppressive effect distributions among variances of permuted distributions (in panels a and b, respectively). Horizontal dashed line: 95th percentile rank

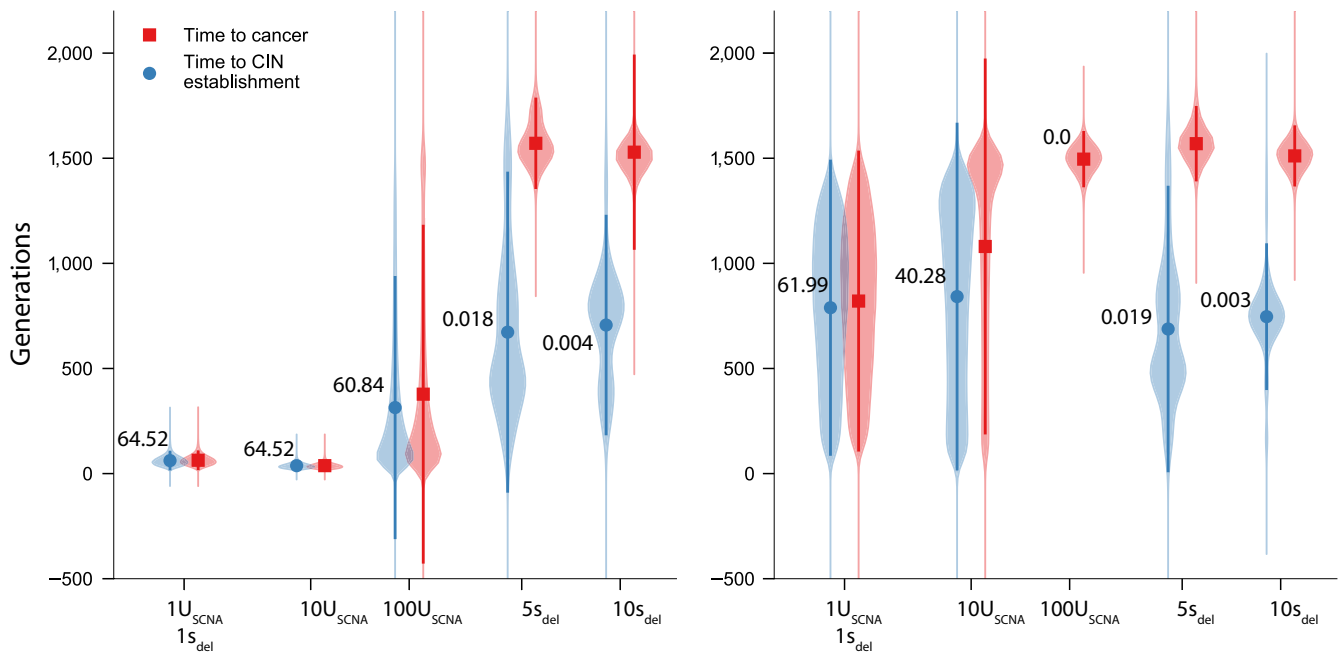
characterized by high (beta-binomially distributed, Figure 1c) and intermediate (geometrically distributed, Figure 1b) clustering of beneficial mutations. For the most clustered genomes (Figure 5a), increasing  $U_{\text{SCNA}}$  10-fold was completely ineffective at preventing either the establishment of CIN or rapid carcinogenesis. Increasing  $U_{\text{SCNA}}$  100-fold (to 1 SCNA per cell per generation) produced only a minor effect on the probability of establishment of CIN mutators, although both CIN establishment and carcinogenesis were somewhat delayed. Increasing  $U_{\text{SCNA}}$  was considerably more effective at inhibiting its establishment and minimizing its role in carcinogenesis in genomes characterized by intermediate clustering (Figure 5b). In this case, 100-fold stronger CIN mutators were completely inhibited while the probability of establishment of 10-fold stronger mutators was reduced by ~30%. On the other hand, increasing  $s_{\text{del}}$  only five-fold resulted in CIN mutators becoming strongly disfavored by selection in both genomes, while the average waiting time to cancer increased to non-CIN levels seen in Figure 3. Thus, while increasing CIN rate may successfully inhibit CIN evolution in some spatially clustered genomes, magnifying the effects of deleterious mutations appears to be a considerably

more effective strategy. We speculate on the reasons for this difference below.

## 4 | DISCUSSION

Here, we have developed a stochastic, individual-based simulation model of clonal populations to examine the evolution of chromosomal instability (CIN) via indirect selection on associated beneficial variation. The propensity of genomic mutators to spread in clonal populations via indirect selection has been extensively studied in evolutionary theory (Gerrish, Colato, Perelson, & Sniegowski, 2007; Kimura, 1967; Taddei et al., 1997) and demonstrated in experimental microbial populations (Chao & Cox, 1983; Raynes & Sniegowski, 2014). Indirect selection on mutators and their potential role in carcinogenesis have also been investigated in computational and analytic models of cancer progression (Beckman & Loeb, 2006; Datta et al., 2013). However, whether indirect selection could favor CIN mutators during carcinogenesis has remained unclear. The reason is that CIN mutators generate SCNAs large enough to simultaneously





**FIGURE 5** Exacerbating deleterious mutations is more effective at inhibiting CIN than increasing the rate of CIN. (a) In genomes with high clustering of beneficial mutations (beta-binomially distributed,  $\mu = 1$ ). (b) In genomes with intermediate clustering of beneficial mutations (geometrically distributed,  $\mu = 1$ ). In both panels: deleterious mutations distributed with no clustering (Dirac delta,  $\mu = 100$ ). Circles are mean values calculated over 100,000 runs of simulation (error bars represent  $\pm 95\%$  CI, all times are shown with violin plots). Model parameters as in Figures 2 and 3 except where noted. See Methods for distribution details.  $P_{CIN_{10\%}} / P_{neutral_{10\%}}$  values next to each data point

affect multiple loci. As deleterious mutations generally outnumber beneficial ones, SCNAs with an overall beneficial effect may be too scarce to allow for CIN hitchhiking. Thus, we hypothesized that indirect selection could only favor CIN in genomes in which fitness-affecting loci were distributed in a way that minimized the co-occurrence of beneficial and deleterious loci in potential SCNAs.

In agreement with our hypothesis, simulations showed that the genomic distribution of fitness-affecting loci can strongly influence indirect selection on CIN-inducing mutators. CIN mutators failed to establish or affect carcinogenesis when both beneficial and deleterious loci were evenly distributed among potential SCNAs. However, spatial clustering of either the beneficial or the deleterious loci promoted rapid hitchhiking of CIN mutators and accelerated carcinogenesis. In genomes characterized by the higher clustering of the beneficial loci, CIN mutators succeeded by acquiring SCNAs containing multiple beneficial mutations, rather than acquiring such mutations individually (as previously seen in models of MSI mutators, (Datta et al., 2013)). On the other hand, in genomes characterized by the higher clustering of the deleterious loci, CIN mutators succeeded by acquiring SCNAs with few beneficial mutations but a reduced load of deleterious ones. Once established, CIN mutators in such genomes were able to accelerate carcinogenesis by rapidly producing additional beneficial mutations via further SCNAs.

Importantly, in our model, we assume that all single-locus mutations have a constant effect on a cell's fitness. Therefore, the availability of beneficial SCNAs that could facilitate hitchhiking in simulation depended solely on the variance in the physical distribution of beneficial and deleterious loci. In a real tumor, however,

different mutations will likely have different effects on a cell's fitness. As a result, the distribution of beneficial effects of real SCNAs will be determined by both the physical distribution of individual loci and the fitness distribution of their mutational effects. For example, the overall beneficial effect of an SCNA could be set by a cluster of smaller effect beneficial mutations or, instead, a single mutation of large effect. Hitchhiking of CIN mutators should then depend on the variance in the distribution of beneficial effects of potential SCNAs being sufficiently high to allow for SCNAs whose beneficial effects could compensate for their deleterious load.

Unfortunately, while many potential oncogenic and tumor-suppressing loci have been discovered, little is known about their fitness effects as these have been difficult to measure empirically. Thus, as a first approximation of the variance in fitness of potential human SCNAs, we used the TUSON algorithm by Davoli et al. (2013), which ranks human loci based on the likelihood of their mutations acting as either oncogenes or tumor suppressors. Like Davoli and colleagues, we assigned each gene a fitness effect corresponding to its rank, resulting in a discrete uniform distribution of fitness effects. Given this simple scheme, both potential oncogenes and tumor suppressors appeared to be more spatially clustered than expected, with shorter SCNAs up to  $\sim 10^6$  nucleotides significantly so. Thus, the human genome may be organized in such a way that some of the available SCNAs have sufficiently large beneficial effects that overcome their deleterious load. If such SCNAs are available, our model suggests that alleles that induce CIN may be favored by indirect selection even in the absence of any direct benefit to a neoplastic cell's fitness. Furthermore, the relatively high clustering of selected

loci across the shorter SCNAs suggests that shorter SCNAs are particularly likely to be positively selected during carcinogenesis and should thus be overrepresented in genomically unstable tumors. Intriguingly, a comprehensive survey of focal SCNA length across multiple cancer types by Beroukhi et al. (2010) showed an inverse relationship between SCNA length and frequency, with a median length of  $1.8 \times 10^6$  nucleotides (Beroukhi et al., 2010). Note that this observation is only consistent with the prediction that natural selection should favor shorter SCNAs and is not evidence for the role of indirect selection in CIN evolution.

It is surprising that the human genome may be organized in a way that promotes CIN evolution. After all, natural selection might have been expected to eliminate or at least reduce clustering of oncogenes and tumor suppressors to lower cancer susceptibility. However, it is important to note that such selection would have been only one of the determinants of genome organization. It is becoming well understood that eukaryotic genes are not randomly distributed across the genome. Related genes and gene families that have arisen through gene duplication may be expected to co-localize (Demuth & Hahn, 2009). Genes with similar or coordinated expression are also frequently clustered (Hurst, Pál, & Lercher, 2004). Importantly, cancer genes identified by Davoli et al. (2013) and used in our analysis appear to be significantly enriched for a handful of functions expected to aid in carcinogenesis, such as cell-cycle control and apoptosis. In light of this observation, it seems plausible that many cancer genes may be related due to their common origin by duplication or share combined regulation and are more clustered than expected as a result. For example, two tumor suppressors frequently inactivated in colorectal cancer, *SMAD2* and *SMAD4* (Fearon, 2011), belong to the same protein family and are located within several megabases of each other. Moreover, we would speculate that the evolution of reduced genomic clustering may be less likely than the evolution of other, perhaps more accessible, mechanisms to suppress cancer (such as additional tumor suppressor genes). As a result, genomic clustering of cancer genes may have persisted despite its potential role in CIN evolution and carcinogenesis.

Quantifying the true variance in fitness effects of potential human SCNAs requires precise fitness measurements of mutations beneficial to a neoplastic cell's fitness. Theoretical studies have produced estimates of mean effects of such mutations from ~0.4% to ~60% (Beerenwinkel et al., 2007; Bozic et al., 2010; McFarland et al., 2014). Vermeulen et al. (2013) were also able to empirically measure the selective effects of three mutations in *p53*, *APC*, and *Kras* in cells of a mouse intestine. Fitness contributions of mutations in most of the known oncogenic and tumor-suppressive loci are yet to be estimated empirically. By considering the distribution of all known driver mutations discovered in different cancer types, we also have implicitly assumed such mutations would be equally beneficial in all cancers. However, the fitness contribution of a mutation in a particular locus should depend greatly on the selective environment and the potential interactions with other mutations present in the genome and will thus likely differ between cancer types. A more rigorous test of the susceptibility of the genome

to CIN hitchhiking would correlate the distribution of beneficial mutations, given their actual fitness contributions in a particular cancer, with the rate of CIN in that cancer. The necessary data are, however, currently unavailable.

Understanding when and how CIN evolves by indirect selection and the potential role of genomic organization in CIN evolution can have important practical implications. While the necessity of CIN for cancer is debatable, our simulations show that once it evolves, CIN can rapidly produce beneficial variation and accelerate carcinogenesis. There is also some evidence that CIN may evolve early in cancer progression (Olaharski et al., 2006; Rajagopalan et al., 2004; Shih et al., 2001; Tonini, 2017). Thus, a potential therapy to prevent CIN evolution in the first place may be able to severely inhibit cancer development. Building on the earlier work of McFarland et al. (2013), we also used our model to examine whether increasing the rate of CIN or the cost of deleterious mutations could prevent evolution of CIN. Both therapeutic strategies aim to increase the deleterious mutational load and may theoretically be expected to inhibit mutator evolution. Intriguingly, we discovered that in genomes characterized by high variance in the distribution of beneficial mutations, increasing the deleterious effects of individual mutations was considerably more effective than increasing the rate of CIN. The disparity in the effectiveness of the two strategies is likely due to the mechanics of CIN hitchhiking in these genomes. Stronger CIN mutators can still produce the rare but very beneficial SCNAs available in these genomes, which allows them to rapidly spread despite the associated deleterious load. Correspondingly, in genomes characterized by lower variance, stronger CIN mutators become less successful. On the other hand, exacerbating the cost of deleterious mutations dramatically decreases the fitness effect of all available SCNAs, amplifying the deleterious load associated with any increase in the rate of CIN and effectively inhibiting CIN mutators.

In the study of McFarland et al. (2013), both increasing the overall mutation rate of a tumor and magnifying the effects of deleterious mutations successfully led to cancer regression. In their model, both strategies work by strengthening selection against deleterious mutations accumulated by neoplastic populations during carcinogenesis. However, magnifying the deleterious effects of these mutations proved to be a more effective therapy in simulation than increasing the mutation rate. Our results agree that exacerbating deleterious effects could also be a more effective strategy to prevent CIN evolution and slow carcinogenesis, assuming the potentially high clustering of carcinogenic beneficial mutations in the human genome. In the clinic, exacerbating deleterious mutations could be potentially achieved, as suggested by McFarland et al., by targeting cellular mechanisms that act to ameliorate their effects in newly made proteins; examples of such mechanisms include chaperones that may help proteins destabilized by deleterious mutations maintain activity (Karras et al., 2017; Rutherford & Lindquist, 1998) and proteasomes that degrade such proteins (Crawford, Walker, & Irvine, 2011).

In summary, extending earlier computational models of cancer progression, we have developed a new model incorporating both

SCNA mutations and CIN-inducing mutators to investigate the role of indirect selection in CIN evolution. Our model predicts that CIN mutators are strongly favored by indirect selection in genomes in which mutations that affect neoplastic fitness are clustered within potential SCNAs. Interestingly, preliminary examination of the distribution of human oncogenes and tumor suppressors suggests that human genomes may, indeed, be organized in such a way as to be susceptible to CIN hitchhiking via indirect selection, although more data on actual fitness contributions of these mutations are needed. Understanding whether CIN evolves by hitchhiking and the role that genome organization plays in CIN evolution may help in future therapeutic efforts aimed at selecting against CIN to inhibit carcinogenesis.

## ACKNOWLEDGEMENTS

We thank members of the Weinreich laboratory for helpful discussion. We also thank two anonymous reviewers for their constructive comments. This work was supported by the postdoctoral fellowship from the Center for Computational Molecular Biology at Brown University.

## CONFLICT OF INTEREST

None declared.

## DATA ARCHIVING

Julia code for the simulation and the simulated data sets are available at <https://github.com/yraynes/Chromosomal-Instability>.

## ORCID

Yevgeniy Raynes  <http://orcid.org/0000-0003-1608-9479>

## REFERENCES

- Bakhoun, S. F., & Compton, D. A. (2012). Chromosomal instability and cancer: A complex relationship with therapeutic potential. *The Journal of Clinical Investigation*, 122(4), 1138–1143. <https://doi.org/10.1172/JCI59954>
- Beckman, R. A., & Loeb, L. A. (2006). Efficiency of carcinogenesis with and without a mutator mutation. *Proceedings of the National Academy of Sciences of the United States of America*, 103(38), 14140–14145. <https://doi.org/10.1073/pnas.0606271103>
- Bedard, P. L., Hansen, A. R., Ratain, M. J., & Siu, L. L. (2013). Tumour heterogeneity in the clinic. *Nature*, 501(7467), 355–364. <https://doi.org/10.1038/nature12627>
- Beerenwinkel, N., Antal, T., Dingli, D., Traulsen, A., Kinzler, K. W., Velculescu, V. E., ... Nowak, M. A. (2007). Genetic progression and the waiting time to cancer. *PLoS Computational Biology*, 3(11), e225. <https://doi.org/10.1371/journal.pcbi.0030225>
- Beroukhi, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., ... Meyerson, M. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature*, 463(7283), 899–905. <https://doi.org/10.1038/nature08822>

- Boland, C. R., & Goel, A. (2010). Microsatellite Instability in Colorectal Cancer. *Gastroenterology*, 138(6), 2073–2087.e3. <https://doi.org/10.1053/j.gastro.2009.12.064>
- Bozic, I., Antal, T., Ohtsuki, H., Carter, H., Kim, D., Chen, S., ... Nowak, M. A. (2010). Accumulation of driver and passenger mutations during tumor progression. *Proceedings of the National Academy of Sciences of the United States of America*, 107(43), 18545–18550. <https://doi.org/10.1073/pnas.1010978107>
- Burrell, R. A., McGranahan, N., Bartek, J., & Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*, 501(7467), 338–345. <https://doi.org/10.1038/nature12625>
- Cahill, D. P., Kinzler, K. W., Vogelstein, B., & Lengauer, C. (1999). Genetic instability and darwinian selection in tumours. *Trends in Cell Biology*, 9(12), M57–M60. [https://doi.org/10.1016/S0168-9525\(99\)01874-0](https://doi.org/10.1016/S0168-9525(99)01874-0)
- Cahill, D. P., Lengauer, C., Yu, J., Riggins, G. J., Willson, J. K., Markowitz, S. D., ... Vogelstein, B. (1998). Mutations of mitotic checkpoint genes in human cancers. *Nature*, 392(6673), 300–303. <https://doi.org/10.1038/32688>
- Carter, S. L., Eklund, A. C., Kohane, I. S., Harris, L. N., & Szallasi, Z. (2006). A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nature Genetics*, 38(9), 1043–1048. <https://doi.org/10.1038/ng1861>
- Chao, L., & Cox, E. C. (1983). Competition between high and low mutating strains of *Escherichia coli*. *Evolution*, 37(1), 125–134.
- Crawford, L. J., Walker, B., & Irvine, A. E. (2011). Proteasome inhibitors in cancer therapy. *Journal of Cell Communication and Signaling*, 5(2), 101–110. <https://doi.org/10.1007/s12079-011-0121-7>
- Datta, R. S., Gutteridge, A., Swanton, C., Maley, C. C., & Graham, T. A. (2013). Modelling the evolution of genetic instability during tumour progression. *Evolutionary Applications*, 6(1), 20–33. <https://doi.org/10.1111/eva.12024>
- Davoli, T., Andrew Wei, X. u., Mengwasser, K. E., Sack, L. M., Yoon, J. C., Park, P. J., & Elledge, S. J. (2013). Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell*, 155(4), 948–962. <https://doi.org/10.1016/j.cell.2013.10.011>
- Demuth, J. P., & Hahn, M. W. (2009). The life and death of gene families. *BioEssays*, 31(1), 29–39. <https://doi.org/10.1002/bies.080085>
- Ewens, W. (2004). *Mathematical population genetics*. New York, NY: Springer.
- Fearon, E. R. (2011). Molecular genetics of colorectal cancer. *Annual Review of Pathology: Mechanisms of Disease*, 6(1), 479–507. <https://doi.org/10.1146/annurev-pathol-011110-130235>
- Futreal, P., Andrew, L. C., Marshall, M., Down, T., Hubbard, T., Wooster, R., ... Stratton, M. R. (2004). A census of human cancer genes. *Nature Reviews Cancer*, 4(3), 177–183. <https://doi.org/10.1038/nrc1299>
- Gerrish, P. J., Colato, A., Perelson, A. S., & Sniegowski, P. D. (2007). Complete genetic linkage can subvert natural selection. *Proceedings of the National Academy of Sciences of the United States of America*, 104(15), 6266–6271. <https://doi.org/10.1073/pnas.0607280104>
- Halazonetis, T. D., Gorgoulis, V. G., & Bartek, J. (2008). An oncogene-induced DNA damage model for cancer development. *Science*, 319(5868), 1352–1355. <https://doi.org/10.1126/science.1140735>
- Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell*, 144(5), 646–674. <https://doi.org/10.1016/j.cell.2011.02.013>
- Heng, H. H., Bremer, S. W., Stevens, J. B., Horne, S. D., Liu, G., Abdallah, B. Y., ... Ye, C. J. (2013). Chromosomal instability (CIN): What it is and why it is crucial to cancer evolution. *Cancer and Metastasis Reviews*, 32(3–4), 325–340. <https://doi.org/10.1007/s10555-013-9427-7>

- Hurst, L. D., Pál, C., & Lercher, M. J. (2004). The evolutionary dynamics of eukaryotic gene order. *Nature Reviews Genetics*, 5, 299. <https://doi.org/10.1038/nrg1319>
- Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., & James Kent, W. (2004). The UCSC table browser data retrieval tool. *Nucleic Acids Research*, 32 (suppl\_1):D493–D496.
- Karras, G. I., Yi, S., Sahni, N., Fischer, M., Xie, J., Vidal, M., ... Lindquist, S. (2017). HSP90 shapes the consequences of human genetic variation. *Cell*, 168(5), 856–866.e12. <https://doi.org/10.1016/j.cell.2017.01.023>
- Kimura, M. (1967). On the evolutionary adjustment of spontaneous mutation rates. *Genetical Research*, 9(1), 23–34. <https://doi.org/10.1017/S0016672300010284>
- Kwei, K. A., Kung, Y., Salari, K., Holcomb, I. N., & Pollack, J. R. (2010). Genomic instability in breast cancer: Pathogenesis and clinical implications. *Molecular Oncology*, 4(3), 255–266.
- Lee, A. J., Endesfelder, D., Rowan, A. J., Walther, A., Birkbak, N. J., Futreal, P. A., ... Swanton, C. (2011). Chromosomal instability confers intrinsic multidrug resistance. *Cancer Research*, 71(5), 1858–1870. <https://doi.org/10.1158/0008-5472.CAN-10-3604>
- Lengauer, C., Kinzler, K. W., & Vogelstein, B. (1997). Genetic instability in colorectal cancers. *Nature*, 386(6625), 623–627. <https://doi.org/10.1038/386623a0>
- Lengauer, C., Kinzler, K. W., & Vogelstein, B. (1998). Genetic instabilities in human cancers. *Nature*, 396(6712), 643–649. <https://doi.org/10.1038/25292>
- Loeb, L. A. (2001). A mutator phenotype in cancer. *Cancer Research*, 61(8), 3230–3239.
- Loeb, L. A. (2011). Human cancers express mutator phenotypes: Origin, consequences and targeting. *Nature Reviews Cancer*, 11(6), 450–457. <https://doi.org/10.1038/nrc3063>
- Lord, C. J., & Ashworth, A. (2012). The DNA damage response and cancer therapy. *Nature*, 481(7381), 287–294. <https://doi.org/10.1038/nature10760>
- Marusyk, A., Almendro, V., & Polyak, K. (2012). Intra-tumour heterogeneity: A looking glass for cancer? *Nature Reviews Cancer*, 12(5), 323–334. <https://doi.org/10.1038/nrc3261>
- Masuda, A., & Takahashi, T. (2002). Chromosome instability in human lung cancers: Possible underlying mechanisms and potential consequences in the pathogenesis. *Oncogene*, 21(45), 6884–6897. <https://doi.org/10.1038/sj.onc.1205566>
- McFarland, C. D., Korolev, K. S., Kryukov, G. V., Sunyaev, S. R., & Mirny, L. A. (2013). Impact of deleterious passenger mutations on cancer progression. *Proceedings of the National Academy of Sciences of the United States of America*, 110(8), 2910–2915. <https://doi.org/10.1073/pnas.1213968110>
- McFarland, C. D., Mirny, L. A., & Korolev, K. S. (2014). Tug-of-war between driver and passenger mutations in cancer and other adaptive processes. *Proceedings of the National Academy of Sciences of the United States of America*, 111:15138–15143.
- McGranahan, N., Burrell, R. A., Endesfelder, D., Novelli, M. R., & Swanton, C. (2012). Cancer chromosomal instability: Therapeutic and diagnostic challenges. *EMBO Reports*, 13(6), 528–538. <https://doi.org/10.1038/embor.2012.61>
- Michor, F. (2005). Chromosomal instability and human cancer. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1455), 631–635. <https://doi.org/10.1098/rstb.2004.1617>
- Negrini, S., Gorgoulis, V. G., & Halazonetis, T. D. (2010). Genomic instability—an evolving hallmark of cancer. *Nature Reviews Molecular Cell Biology*, 11(3), 220–228. <https://doi.org/10.1038/nrm2858>
- Nowak, M. A., Komarova, N. L., Sengupta, A., Jallepalli, P. V., Shih, I. E., M., Vogelstein, B., & Lengauer, C. (2002). The role of chromosomal instability in tumor initiation. *Proceedings of the National Academy of Sciences of the United States of America*, 99(25), 16226–16231. <https://doi.org/10.1073/pnas.202617399>
- Olaharski, A. J., Sotelo, R., Solorza-Luna, G., Gensebatt, M. E., Guzman, P., Mohar, A., & Eastmond, D. A. (2006). Tetraploidy and chromosomal instability are early events during cervical carcinogenesis. *Carcinogenesis*, 27(2), 337–343. <https://doi.org/10.1093/carcin/bgi218>
- Pikor, L., Thu, K., Vucic, E., & Lam, W. (2013). The detection and implication of genome instability in cancer. *Cancer and Metastasis Reviews*, 32(3–4), 341–352. <https://doi.org/10.1007/s10555-013-9429-5>
- Rajagopalan, H., Jallepalli, P. V., Rago, C., Velculescu, V. E., Kinzler, K. W., Vogelstein, B., & Lengauer, C. (2004). Inactivation of hCDC4 can cause chromosomal instability. *Nature*, 428(6978), 77–81. <https://doi.org/10.1038/nature02313>
- Raynes, Y., & Sniegowski, P. D. (2014). Experimental evolution and the dynamics of genomic mutation rate modifiers. *Heredity*, 113(5), 375–380. <https://doi.org/10.1038/hdy.2014.49>
- Raynes, Y., Wylie, C. S., Sniegowski, P. D., & Weinreich, D. M. (2018). Sign of selection on mutation rate modifiers depends on population size. *Proceedings of the National Academy of Sciences of the United States of America*, 115(13), 3422–3427. <https://doi.org/10.1073/pnas.1715996115>
- Rutherford, S. L., & Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. *Nature*, 396, 336. <https://doi.org/10.1038/24550>
- Salk, J. J., Fox, E. J., & Loeb, L. A. (2010). Mutational heterogeneity in human cancers: Origin and consequences. *Annual Review of Pathology: Mechanisms of Disease*, 5(1), 51–75. <https://doi.org/10.1146/annurev-pathol-121808-102113>
- Shih, I.-M., Zhou, W., Goodman, S. N., Lengauer, C., Kinzler, K. W., & Vogelstein, B. (2001). Evidence that genetic instability occurs at an early stage of colorectal tumorigenesis. *Cancer Research*, 61(3), 818–822.
- Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favorable gene. *Genetical Research*, 23, 23–25.
- Sniegowski, P. D., Gerrish, P. J., Johnson, T., & Shaver, A. (2000). The evolution of mutation rates: Separating causes from consequences. *BioEssays*, 22(12), 1057–1066. [https://doi.org/10.1002/1521-1878\(200012\)22:12aabb1057:AID-BIES3aaabb3.0.CO;2-W](https://doi.org/10.1002/1521-1878(200012)22:12aabb1057:AID-BIES3aaabb3.0.CO;2-W)
- Solomon, D. A., Kim, T., Diaz-Martinez, L. A., Fair, J., Elkahoul, A. G., Harris, B. T., ... Waldman, T. (2011). Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science*, 333(6045), 1039–1043. <https://doi.org/10.1126/science.1203619>
- Sprouffske, K., Merlo, L. M., Gerrish, P. J., Maley, C. C., & Sniegowski, P. D. (2012). Cancer in light of experimental evolution. *Current Biology*, 22(17), R762–R771. <https://doi.org/10.1016/j.cub.2012.06.065>
- Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P. H., & Godelle, B. (1997). Role of mutator alleles in adaptive evolution. *Nature*, 387(6634), 700–702. <https://doi.org/10.1038/42696>
- Thibodeau, S. N., Bren, G., & Schaid, D. (1993). Microsatellite instability in cancer of the proximal colon. *Science*, 260(5109), 816–819. <https://doi.org/10.1126/science.8484122>
- Tonini, G. P. (2017). Growth, progression and chromosome instability of Neuroblastoma: A new scenario of tumorigenesis? *BMC Cancer*, 17(1), 20.
- Vermeulen, L., Morrissey, E., van der Heijden, M., Nicholson, A. M., Sottoriva, A., Buczacchi, S., ... Winton, D. J. (2013). Defining stem cell dynamics in models of intestinal tumor initiation. *Science*, 342(6161), 995–998. <https://doi.org/10.1126/science.1243148>
- Vilar, E., & Gruber, S. B. (2010). Microsatellite instability in colorectal cancer—the stable evidence. *Nature Reviews Clinical Oncology*, 7(3), 153–162. <https://doi.org/10.1038/nrclinonc.2009.237>
- Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A. Jr, & Kinzler, K. W. (2013). Cancer genome landscapes. *Science*, 339(6127), 1546–1558. <https://doi.org/10.1126/science.1235122>
- Walther, A., Houlston, R., & Tomlinson, I. (2008). Association between chromosomal instability and prognosis in colorectal

cancer: A meta-analysis. *Gut*, 57(7), 941–950. <https://doi.org/10.1136/gut.2007.135004>

Yates, A., Wasu Akanni, M., Amode, R., Barrell, D., Billis, K., Carvalho-Silva, D., ... Flicek, P. (2016). Ensembl 2016. *Nucleic Acids Research*, 44(D1), D710–D716. <https://doi.org/10.1093/nar/gkv1157>

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Raynes Y, Weinreich DM. Genomic clustering of fitness-affecting mutations favors the evolution of chromosomal instability. *Evol Appl*. 2019;12:301–313. <https://doi.org/10.1111/eva.12717>