Weakly deleterious natural genetic variation amplifies probability of resistance in multiplexed gene drive systems

Short title: Resistance in multiplexed gene drive systems

Bhavin S. Khatri[1,2*] and Austin Burt[1]

[1]Department of Life Sciences, Imperial College London,
Silwood Park, Ascot, SL5 7PY, United Kingdom
[2]The Francis Crick Institute, 1 Midland Road, London, NW1 1AT, United Kingdom

*To whom correspondence should be addressed; E-mail: bkhatri@imperial.ac.uk

Evolution of resistance is a major barrier to successful deployment of gene drive systems to suppress natural populations. Multiplexed guide RNAs that require resistance mutations in all target cut sites is a promising strategy to overcome resistance. Using novel stochastic simulations that accurately model evolution at very large population sizes, we explore the probability of resistance due to three important mechanisms: 1) non-homologous end-joining mutations, 2) single nucleotide mutants arising de novo or, 3) single nucleotide polymorphisms pre-existing as standing variation. If the fraction of functional end-joining mutants is rare, we show that standing variation dominates, via a qualitatively new phenomenon where weakly deleterious variants significantly amplify the probability of multi-site resistance. This means resistance can be probable even with many target sites in not very large populations. This result has broad application to resistance arising in multi-site evolutionary scenarios including the evolution of vaccine escape mutations in large populations.

Introduction

Suppression gene drive systems have the potential to be highly effective for population control of many vectors of human disease, including malaria carrying mosquitoes, as well as for invasive species control (*1*). With the recent discovery of the CRISPR-Cas9 gene editing system (*2*), such drives can be realised by targeting specific sequences for insertion of a drive construct into a gene of importance; during meiosis this converts a very large fraction gametes carrying wild type alleles to this drive construct, thereby increasing transmission of drive over and above random Mendelian segregation. If the drive construct only has significant fitness effects as a homozygote, it can rise to high frequencies; in particular, if the target gene is necessary for female fertility this provides an effective way to suppress the population (*3, 4*). An important question is how resilient gene drive systems are to the evolution of resistance (*1, 5–7*), which could arise via single nucleotide mutations or polymorphisms (SNPs) or by an imperfect end-joining repair process called non-homologous end joining (NHEJ) (*3, 8*). Although, in caged populations resistance to suppression has been observed via indels induced by NHEJ (*3*), drive systems where the chosen target site is critical to female fertility have been found to be resilient (*4*). However, it is an open question as to how resilient gene drive systems are to the evolution of resistance in natural populations, which can have very large population sizes (*9*).

Although empirically we understand well the rate at which these processes occur in individuals, from characterisation of the mutation rate (μ) and the net NHEJ rate (ν), both per generation, it is also crucially important to know 1) the population size, and 2) the fraction of potential resistance mutants that are actually functional. Rather than the per individual rate, the population-level mutation rate is critical, as it measures how likely such mutations are to arise in the population (*5–7*). For example, if a resistance mutant is produced per individual at rate 10^{-7} and the population size is $N = 10^4$, the population level mutation rate is 10^{-3} and it will typically take 1000 generations before a single resistance mutant is generated in a population, and so resistance is unlikely to arise given that drive typically acts to suppress a population on a timescale of much less than 100 generations; on the other hand if $N = 10^8$, 10 such mutants are generated every generation, and so resistance mutants with a selective advantage compared to drive are very likely to arise and fix before population elimination.

One way to reduce the potential for resistance to drive is to exploit multiple, redundant target sites for cleavage by the nuclease (*1, 7, 10*). If there are multiple guide RNAs (gRNAs) or target cut sites within a single locus, since a single successful cut is sufficient for drive to be copied, all sites must develop resistance mutants in order for the non-Mendelian transmission of drive to fail. Using a deterministic model, Noble et al. (*10*), showed that multiplexed strategies can allow drive to overcome resistance on short time scales, however, as population dynamics were not considered, in such infinite population size models, resistance will always arise eventually for any fertile allele. Beaghton et al, explored explicit population dynamics and finite population stochasticity and the probability of resistance using branching process theory in the context of the Y-drive suppression strategy (*5*) and Marshall et al (*7*) used more complex and realistic stochastic simulations to examine how resistance interplays with population size and the rate of generation of NHEJ mutants. Both these works showed that large natural population sizes reduce the maximum permissible rate of resistance mutants. Using heuristic arguments, Marhsall et al, then argued that relatively modest degrees of multiplexing m could prevent resistance from arising, as the rate that multiplexed resistance arises diminishes as a

power law of the rate of a single cut site with exponent m; in other words the critical population size at which resistance arises increases as a power law in m. These works have not explored the role of standing variation, but more recently Lanzaro et al (*11*) showed that higher levels of standing variation of resistance variants increases the probability of resistance to various gene drive strategies including suppression drives, though crucially they did not look at the role of population size and multiplexing.

On a theoretical level there are a number of works in the population genetic literature of relevance to resistance in suppression drives (*12–15*) which predict that population rescue or adaptation in response to an environmental change, in general, is more likely from standing variation than compared to de novo mutants arising. Although simple theoretical arguments for the role of standing variation with multi-site evolution have been considered previously using branching process theory (*16*), they assumed the initial frequency of mutants are fixed at their deterministic mutation-balance frequencies. However, stochastic theory (*15*) shows that this overestimates the role of standing variation, since by averaging over the allele frequency distribution before the change, the advantage is only weakly logarithmic in the fitness cost, and in general not very significant. There is currently no theory that addresses the question of the role that standing variation plays, with its full allele frequency spectrum, in determining the probability of resistance for a multi-site evolutionary system like multiplexed gene-drive, and this is an important question we address here.

In this work, we develop fully stochastic Wright-Fisher simulations of multiplexed drive to analyse the role of population size in more detail, for different degrees of multiplex m and for more general scenarios of resistance not previously considered, including, as we shall see, the critical role of standing variation in multiplexed drives. We exclusively explore the population rescue scenario, where drive is introduced at sufficiently large frequency that the population will be eliminated without resistance. Multiplexing means the critical population sizes required for resistance can be very large, yet the resistance mutants still arise at very small frequency, requiring a stochastic description of their establishment; to this end, we develop a new and very accurate Gaussian-Poisson approximation to the multinomial distribution to allow Wright-Fisher sampling in almost arbitrarily large population sizes. We use these simulations to assess the modes of resistance arising for both NHEJ mutants and SNP mutants, where the later can arise by de novo base-pair mutation or pre-exist in the population before the introduction of drive as standing variation. Importantly, we also allow each type to be functional or non-functional.

Our results show that multiplexing with multiple gRNAs in general increases the minimum population size at which resistance can arise, as previous works have found, but rationalise these results to find that the probability follows a universal sigmoidal form, which is motivated from heuristic arguments based on the probability of fixation given a number of resistance mutants generated before drive eliminates the population. This analysis shows that in multiple gRNA systems, resistance via de novo SNPs arises through sequential accumulation of resistance alleles, whilst for NHEJ, as the cleavage efficiency is very high, resistance arises from generation of resistance at all gRNAs concurrently. On the other hand, with respect to pre-existing variants we find a new theoretical result: the more complex structure of standing variation with multiple gRNAs leads to a power law dependence on the fitness cost of standing variation, compared to the weak logarithmic fitness effects for a single gRNA. This leads to an extreme sensitivity of resistance on the fitness cost, significantly amplifying the probability of resistance, compared to de novo, for weakly selected alleles as the number of target sites and gRNAs increases. As well as being important to predict the success of multiplexed drives

systems in preventing resistance at a given population size, we suggest this result has wide applicability to understanding resistance due to standing variation for a range of multi-site evolutionary systems, including the question of vaccine escape.

Simulations

We use non-spatial Wright-Fisher stochastic simulations of drive with separate sexes throughout, but with coupling to population dynamics using density-dependent Beverton-Holt growth. The details of these simulations are given in the Supplementary Information. The simulations are stochastic, even when population sizes are very large, because resistant mutations always arise at small frequency, so genetic drift needs to be explicitly considered. Resistance corresponds to the establishment and then fixation of functionally resistant alleles, which we define to be a frequency greater than 0.95. The simulations entail stochastic dynamics with one, two, or three gRNAs, where at each *target* site there are 4 alleles that can occur W (wild type), D (drive), R (functional resistance), N (non-functional resistance); this means that across an m-fold target site, we have an m-locus, 4-allele population genetic system with no recombination. However, as we assume the drive construct is copied over as a whole, in practice, it is an m-locus 3-allele system $+1$ for the D allele. On a single chromosome the possible haplotypes (assuming no positional effects) for a 2-fold system are WW, WR, RR, WN, RN, NN, DD, which is a total of $n(n+1)/2+1 = (3\times 4)/2+1 = 7$ haplotypes, and an analogous calculation in the Supplementary Information for $m = 3$ gives a total of 11 haplotypes.

The alleles are assumed to have no effect on male fitness, while fitness effects in females are as follows. W alleles have zero fitness cost, and D and N alleles are deleterious with homozygous fitness cost $s = 1$ and heterozygous fitness costs (when paired with a W allele) of $h = 0.3$ and $h_N = 0.02$, respectively; the larger dominance coefficient for drive represents potential somatic fitness costs to heterozygotes females, from leaky expression of the Cas-9 protein (*4*). The heterozygous fitness cost of R, before the introduction of drive, is σ, which we vary in the simulations. When combined in haplotypes we assume each target site has independent fitness effects, which means a single occurrence of an N allele is deleterious. Fitness costs are manifest as reduced female survival. Further details are given in the Supplementary Information.

In W/D heterozygotes (and their multiplexed equivalents discussed in the Supplementary Information) we assume a cleavage efficiency $\epsilon = 0.95$, and a non-homologous end joining (NHEJ) rate of $\nu = 0.05$ approximately representative of the target site in the gene *doublesex* in *Anopheles gambiae* (*4*). We assume NHEJ mutants produce functional resistant alleles R with probability β and nonfunctional resistant alleles N with probability $1-\beta$. As a result D gametes are generated from conversion of W at a rate $(1-\nu)\epsilon$, functional resistance alleles R at rate $\epsilon\beta\nu$, and non-functional alleles N at rate $(1-\beta)\nu$, while the fraction that remain wild type is $1-\epsilon$.

We assume the probability of cleavage at each available target site given by efficiency ϵ, occurs independently at different sites, so that the fraction of non-driving gametes produced from a genotype with a non-driving allele/haplotype paired with a driver is $\frac{1}{2}[1-\epsilon(1-\nu)]^{m-r}$, where m is the number of gRNAs and $r \leq m$ is the number of resistant (R or N) sites in the non-driving allele. Thus when there are multiple gRNAs the presence of a resistant sites give some protection to the chromosome from being cut, even if one or more cleavable sites remain.

In addition, functional resistance alleles R are generated de novo at rate $\xi\mu$, where μ is the mutation rate for the length of site of interest and ξ is the fraction of SNP mutations that are functional, and non-functional resistance alleles at rate $(1-\xi)\mu$. We assume $\mu = 18\mu_0 = 5.4\times10^{-8}$, where 18 is the length of each target site and $\mu_0 = 3\times10^{-9}$ is the base-pair mutation rate measured for *Drosophila* (*17*).

Standing variation

Before drive is introduced into a population there may be pre-existing functionally resistant SNPs in the population which would affect the probability of resistance arising. To study the effect of standing variation we run replicate simulations where we allow a burn-in period of $1/\sigma$ generations to allow for the population to come to a mutation-selection balance equilibrium. The initial frequency of the various resistance alleles/haplotypes when drive is introduced at $t = 0$ is then implicitly drawn from the mutation-selection balance equilibrium. Note that in the case of $m > 1$ the mutation-selection balance distribution will be complex with different frequencies for haplotypes carrying different numbers of resistance mutations R.

Gaussian-Poisson hybrid approximation to generating multinomial random numbers

In this paper we run simulations to very large effective population sizes. Whilst it is typical in such a scenario to ignore the stochastic part of the evolutionary dynamics by using deterministic dynamics, this is only accurate if the allele frequencies are large themselves, or equivalently the number of copies in the population are large ($\gg 1$). We are interested in the dynamics of resistance, which by definition means we need to study situations where the allele arises by de novo mutation as a single copy in a single individual, where it must survive genetic drift, or exist at very low frequency as standing variation. When there are multiple alleles, particularly when simulating Wright-Fisher evolutionary dynamics, this is accomplished simply by drawing multinomial random numbers. However, when the effective population size is large this can become increasingly slow. In addition, the maximum population size is restricted to the largest integer that can be stored in a computer; for the GNU scientific library's implementation of multinomial random number generators, this is limited to 32-bit, which gives a limit of roughly 4×10^9.

An alternative approach is to use the multivariate Gaussian approximation to the multinomial distribution:

$$p(n_1, n_2, ..., n_K | x_1, x_2, ..., x_K, N) \tag{1}$$

$$= \frac{N!}{n_1! n_2! ... n_K!} x_1^{n_1} x_2^{n_2} ... x_K^{n_K} \tag{2}$$

$$\approx \frac{\exp\left(-\frac{1}{2}(\boldsymbol{n} - N\boldsymbol{x})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{n} - N\boldsymbol{x})\right)}{\sqrt{(2\pi)^K \det(\boldsymbol{\Sigma})}} \tag{3}$$

where $\boldsymbol{n} = (n_1, n_2, ..., n_K)^T$, $\boldsymbol{x} = (x_1, x_2, ..., x_K)^T$, are the vectors of the numbers drawn of K alleles, and their expected frequency, respectively, and $\boldsymbol{\Sigma}$ is the scaled covariance matrix of the multinomial distribution, where $\Sigma_{ij} = N x_i(\delta_{ij} - x_j)$. However, this approximation is poor when

for any of the alleles $Nx_k \sim 1$. In this limit, these rare alleles are well-approximated by a Poisson distribution. The approach taken in this paper is therefore to partition the alleles into a rare category $\mathcal{R}$, if $Nx_k \leq 10$, and non-rare if $Nx_k > 10$, where the former is drawn from independent Poisson distributions, while the latter from a multivariate Gaussian distribution conditioned on a smaller total population size $N' = N - \sum_{k \in \mathcal{R}} n_k$:

$$p(n_k|x_k) \approx \frac{(Nx_k)^{n_k} e^{-Nx_k}}{n_k!} \quad \forall k \in \mathcal{R} \tag{4}$$

$$p(\boldsymbol{n}'|\boldsymbol{x}') \approx \frac{\exp\left(-\frac{1}{2}(\boldsymbol{n}' - N'\boldsymbol{x}')^T \boldsymbol{\Sigma}'^{-1}(\boldsymbol{n}' - N'\boldsymbol{x}')\right)}{\sqrt{(2\pi)^{K'} \det(\boldsymbol{\Sigma}')}} \tag{5}$$

where the vectors $\boldsymbol{n}'$ and $\boldsymbol{x}'$ only take elements $k \notin \mathcal{R}$, whose length is $K' = K - |\mathcal{R}|$, and the covariance matrix $\Sigma'_{ij} = N'x'_i(\delta_{ij} - x'_j)$. We can assume independent Poisson distributions for each of the rare alleles precisely because they are rare and the effects of drift are approximately independent of each other, and the constraint of constant population size is imposed on the non-rare alleles through the modified population size N' and the correlation structure in the covariance matrix $\boldsymbol{\Sigma}'$.

Results

Our main results are the probability of resistance arising as a function of the effective population size N, for one, two, or three gRNAs. We will explore the probability of resistance as a function of three main parameters, the fraction of NHEJ mutants that are functional (β), the fraction of de novo mutations that are functional (ξ), and the heterozygous fitness cost (in the absence of drive) of functional resistance (σ) (assumed to be the same whether it was derived from de novo mutation or NHEJ); the remaining fraction $1 - \beta$ of NHEJ mutants, or $1 - \xi$ of de novo SNPs are assumed nonfunctional. We find that the probability of resistance has a universal sigmoidal behaviour, which is of the form $p = 1 - e^{-N/N^*}$, where for small population sizes $N \ll N^*$ resistance is very unlikely, whilst for sufficiently large population sizes ($N \gg N^*$) resistance arises with near 100% certainty. We use the simulations to develop a heuristic theory to characterise how N^* depends on the parameters β, ξ, σ, N and then focus in particular on contours in the parameter space of these variables where the probability of resistance is $p = 0.05$, or 5%.

The results are organised by 4 different conditions: 1) NHEJ mutations only, 2) de novo SNP mutants only, 3) standing variation of SNP mutants in mutation-selection balance, and 4) where all of the first three mechanisms are possible, where for each we vary the fraction of functional NHEJ mutants β, the fraction of de novo SNPs ξ, and the fitness costs of functional mutants (NHEJ or SNP) as appropriate.

All the results are determined from running 500 replicate simulations. For each set of parameters we determine whether resistance to the drive construct arose using a resistance criterion that the sum of the frequency of resistance allele/haplotypes reaches 0.9; for example, when the frequency of R and N, for a single gRNA, is greater than 90%, or for 2 gRNAs, when the frequency of RR, RN and NN is greater than 90%. We also examine the time to resistance, given that resistance arises, in the Supplementary material. In all simulations drive is introduced at a frequency of 0.1 in males.

Typical trajectories of allele/haplotype frequencies

In Fig.1 we show for $m = 1$, $m = 2$ and $m = 3$ the typical trajectories of the different alleles/haplotypes at population sizes for $N < N^*$ (left hand plots of the figure), where resistance does not arise and $N > N^*$ (right hand plots of the figure), where resistance does arise, due to de novo mutation or NHEJ. Below each is the corresponding time series of population size, showing population elimination (left) and rescue (right). These simulations assume that the fraction of functional NHEJ mutations is $\beta = 10^{-4}$ and that all de novo SNPs are functional ($\xi = 1$). We see in all cases that initially drive replaces the wild type in less than 10 generations, which causes the mean population fitness to decrease and the population size itself to decrease; following this, if $N > N^*$ we see that functional resistance mutants arise and then fix giving rise to population rescue, while for $N < N^*$, resistance mutants do not arise on the timescale at which the population is eliminated. As we increase the degree of multiplexing m, the same story unfolds, except there are many more resistance haplotypes, which all arise at small frequency, and except for the single functional mutant for each value of m (R, RR and RRR, respectively for $m = 1, 2, \& 3$ gRNAs), none of these rise to sufficiently high frequency to achieve population rescue as they are deleterious as homozygotes, and as drive heterozygotes, haplotypes containing W can still be converted to D. The resistance haplotypes that have m copies of R do increase to high frequencies giving resistance and population rescue.

Probability of resistance: NHEJ mutants only

In this section we examine simulations where only NHEJ mutants are possible ($\mu = 0$). In Fig.2 we plot the probability of resistance as a function of population size for a) 1 gRNA, b) 2 gRNAs and c) 3 gRNAs, for the standard parameters outlined above, but where the cost of functional resistance mutants is $\sigma = 0.01$ (different values of σ did not change these results – see Fig.3 below). Each set of simulation data with square symbols of a given colour corresponds to a different value of β, the fraction of functional NHEJ mutants; we see a characteristic sigmoidal shape for all the curves, but each offset, where for decreasing values of β the curves shift to larger population sizes, indicating that the probability of resistance is decreasing. These results (square symbols) are obtained using the hybrid Poisson-Gaussian approximation to the multinomial distribution (see Methods), which allows Wright-Fisher simulations at very large population sizes; to check this approximation works well, the results with pentagram symbols represent simulations at smaller population sizes using multinomial generated random numbers, for which we see very good agreement with the Poisson-Gaussian approximation.

Each of these curves is fit with a single free parameter γ_n, using the functional form $p_n = 1 - e^{-N/N_n^*}$, where

$$N_n^* = \frac{1}{4\gamma_n(\epsilon\nu\beta)^m}; \tag{6}$$

we find that $\gamma_n \approx 0.2$ for all values of m (see Supplementary Materials for details) and we see the fit is excellent, as shown in Fig. 2. This form of N^* indicates that once the population level rate of producing functional resistance mutants $\sim N(\epsilon\nu\beta)^m$, which requires m copies of R, is sufficiently large, then the probability of resistance is large. As we show in the Supplementary Information this

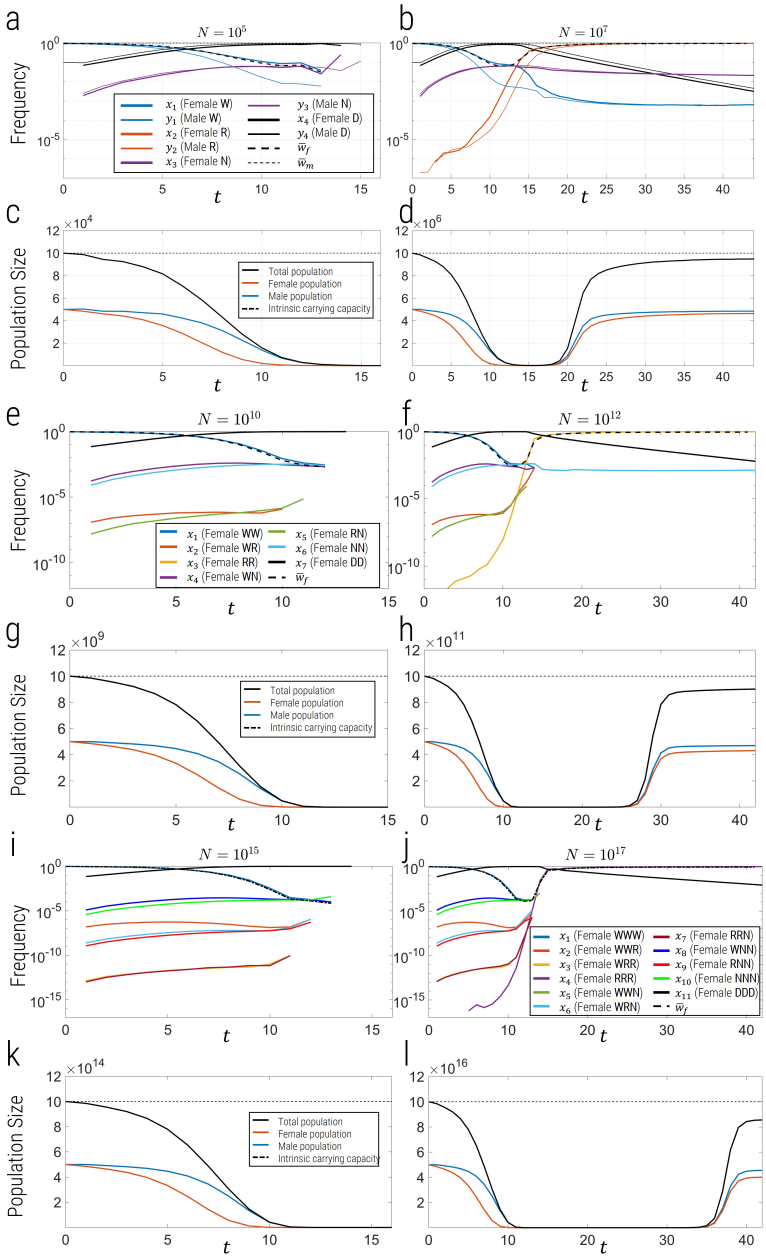


Figure 1: Plot of frequency of different alleles (a,b, e, f, i & j) and population dynamics (c, d, g, h, k & l) vs generation time for $N < N^*$ (a, c, e, g, i & k) and $N > N^*$ (b, d, h, j & l) for $m = 1$ gRNA (top panel), $m = 2$ gRNAs (middle panel) and $m = 3$ gRNAs (bottom panel), where N^* is the critical population size at which the probability of resistance (rescue) is $1 - e^{-1}$. For $m = \{2, 3\}$, we only plot the female allele frequencies for clarity.

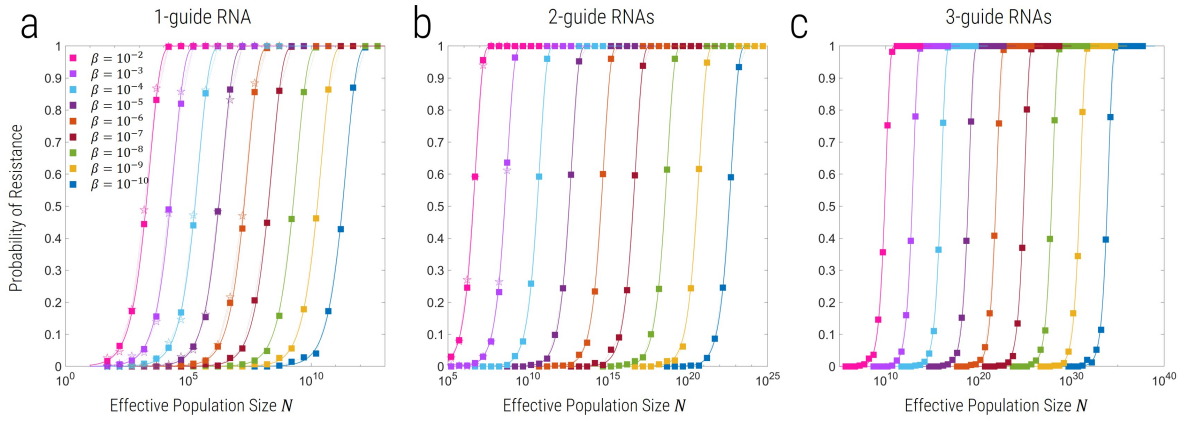


Figure 2: Probability of resistance evolving as a function of effective population size (N) for $m = 1$, 2, or 3 gRNAs, where resistance can only arise due to functional NHEJ mutants. R alleles have a fitness cost of $\sigma = 0.01$ relative to the wild type. Curves of different colour correspond to different values of β, where squares correspond to simulations using the hybrid Poisson-Gaussian approximation and open pentagrams to simulations using exact multinomial sampling.

particular form with a single fixed γ_n independent of m arises when the rate of generation of m-fold resistance mutants concurrently is faster than generating them sequentially, which we show is generally the case when cleavage is efficient $(1 - \epsilon \ll 1)$.

In Fig.3, we can summarise this information by plotting contours of $p = 0.05$ on axes of β vs N, where the region to the left of the contour indicates values of β and N for which we expect the probability of resistance to be ≤ 0.05. We see as above that as N increases the value of β required to keep $p \leq 0.05$ decreases, but that it decreases as a power law $\sim \beta^{-m}$, as indicated by the form for N_n^* above; broadly this shows that increasing the number of gRNAs increases the range of population sizes where resistance is unlikely, with the trade-off between β and N described. The solid lines are plots of $p = 1 - e^{-N/N_n^*}$ for $p = 0.05$ and $\gamma_n = 0.2$. The different symbols correspond to different fitness costs in the simulations, and we see that here this has no effect on contours, which is intuitive as these mutants only become advantageous once drive has nearly fixed, and their effective frequency dependent selection coefficient has only a weak dependence on σ, the cost in the presence of wild type only.

Probability of resistance: de novo SNPs only

In this section we examine the probability of resistance for simulations where resistance can only arise due to de novo mutations, with NHEJ turned off $(\nu = 0)$ and assuming there are no preexisting functional resistance SNPs. Plots of the probability of resistance vs effective population size have a similar sigmoidal behaviour as in Fig.2, where as we decrease ξ, the fraction of functional SNP mutants, the curves shift to larger population sizes, meaning resistance is less likely for a given population size (not shown). Each of these curves can be fit with the same generic functional form $p_d = 1 - e^{-N/N_d^*}$, where

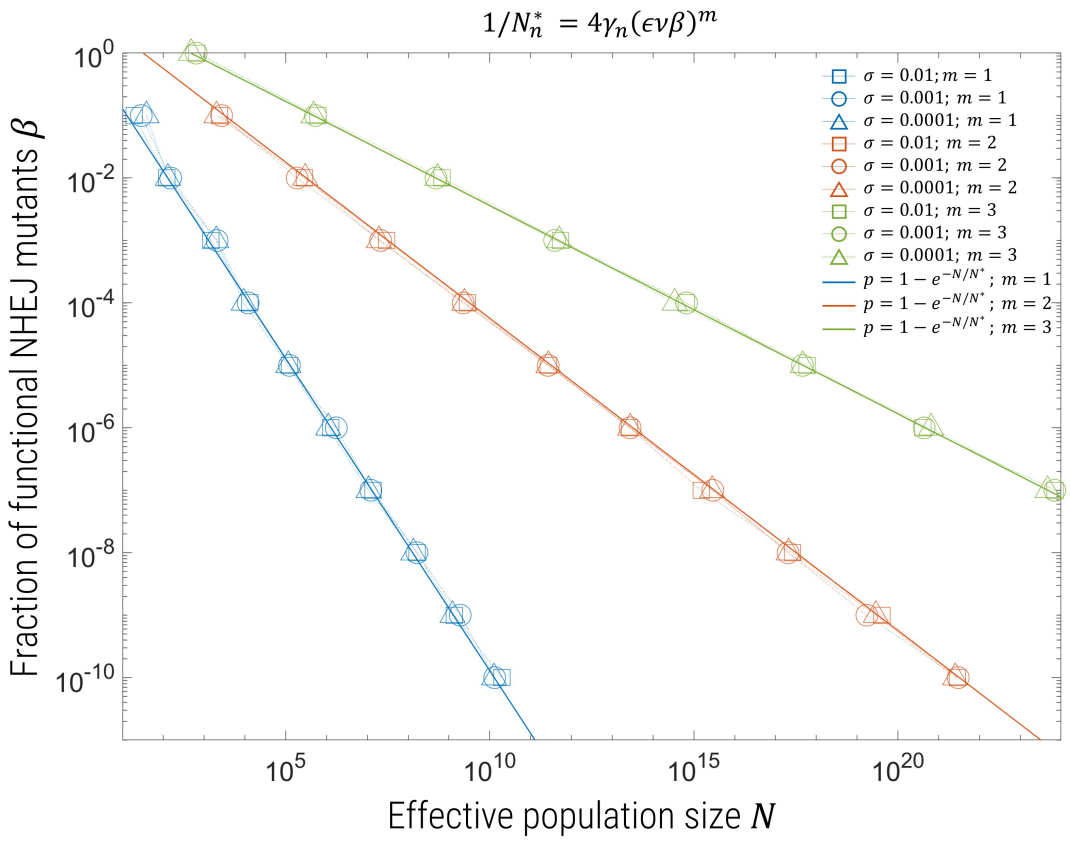


Figure 3: Contours for probability of resistance $p = 0.05$ as a function of N and β for $m = 1$ (blue), 2 (red) or 3 (yellow) gRNAs. Different symbols correspond to interpolated values from simulations for different fitness costs for functional mutants: square symbols $\sigma = 0.01$, triangle symbols $\sigma = 0.001$, circle symbols $\sigma = 0.0001$, and the solid line is a plot of $p = 1 - e^{-N/N_n^*}$ for $\gamma_n = 0.2$.

$$N_d^* = \frac{1}{4\gamma_m(\xi\mu)^m}, \tag{7}$$

but unlike for NHEJ mutants, the fitting constant γ_m is not independent of m and we show in the Supplementary material is of the form $\gamma_m = s_b\tau^m$, which arises when multi-fold resistance alleles arise sequentially rather than concurrently; here, s_b is an effective/aggregate beneficial selection coefficient for the resistant haplotype with m functional resistant alleles and τ is the timescale over which these resistant mutants accumulate. By fitting the simulation data we find $s_b \approx 0.09$ and $\tau \approx 8.5$ generations, which pleasingly is a time consistent with the time scale over which drive increases to high frequency.

This information is summarised in Fig.4a, by plotting contours of $p = 0.05$ as a function of N and ξ, where regions to the left of the contour indicate combinations of N and ξ for which the probability of resistance is less than 0.05. We see that as N increases, the fraction of mutants that

are functional must be smaller to maintain the probability of resistance at 0.05. The solid lines are plots of $p = 1 - e^{-N/N_d^*}$, for $p = 0.05$, and $\gamma_m = \{0.76, 6, 55\}$, coloured blue, red, and yellow, for $m = \{1, 2, 3\}$, respectively. The different symbols correspond to different fitness costs of mutants in the presence of wild type, and as for NHEJ mutants, we see this has no effect on the probability of resistance.

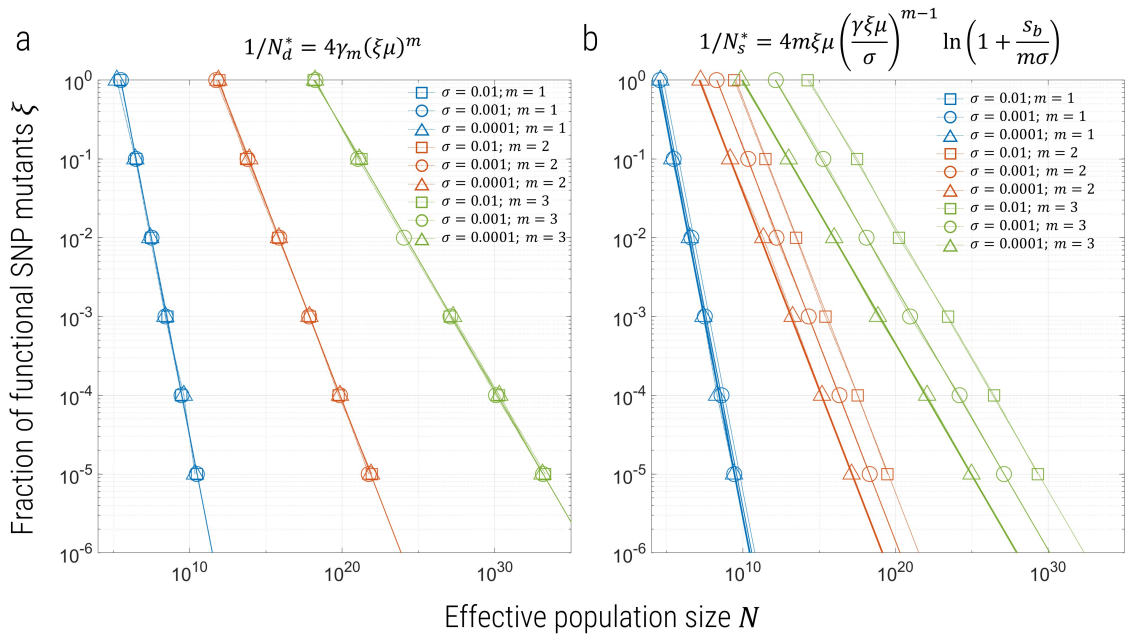


Figure 4: Contours for probability of resistance $p = 0.05$ as a function of N and ξ for $m = 1$ (blue), 2 (red) or 3 (yellow) gRNAs. The left panel (a) corresponds to simulations with de novo generation of SNP mutants only, while the right panel (b) are simulations with both de novo and preexisting SNP mutants. Different symbols correspond to simulations for different fitness costs for functional mutants: square symbols $\sigma = 0.01$, triangle symbols $\sigma = 0.001$, circle symbols $\sigma = 0.0001$, and the solid lines are a plot of $p = 1 - e^{-N/N^*}$, with $N^* = \{N_d^*, N_s^*\}$ for a) & b) given by the Eqns.7 & 8, respectively, where for (a) $\gamma_m = \{0.76, 6, 55\}$ and for (b) $\gamma = 2$ and $s_b = 0.5$.

Probability of resistance: preexisting and de novo SNP mutants only

We now allow the possibility that functional resistant (R) mutants may exist in a mutation-selection balance before the introduction of drive, by running each replicate simulation for a period of time $1/\sigma$ before the introduction of drive, which gives sufficient time for the frequency distribution of the SNP mutants at the time of introduction of drive to have equilibrated (see Methods). As in the previous sections, we can very accurately fit curves of the probability of resistance vs N, using the same functional form $p_s = 1 - e^{-N/N_s^*}$, where now

$$N_s^* = \frac{1}{4\xi\mu m\left(\frac{\gamma\xi\mu}{\sigma}\right)^{m-1}\ln(1+\frac{s_b}{m\sigma})}; \tag{8}$$

for $m = 1$ this corresponds to the theory of Pennings & Hermisson (*15*), assuming a fixed beneficial selection coefficient of s_b for the resistance mutant in the presence of the drive allele, which is derived by calculating the average probability of fixation over the distribution of allele frequencies in mutation selection balance, where the fitness cost is σ, before drive is introduced. For sufficiently strong selection $4N\sigma > 1$, this allele frequency distribution is closely approximated by a gamma distribution with shape parameter $\theta = 4N\xi\mu$ and rate parameter $\alpha = 4N\sigma$. In the case of $m > 1$ gRNAs, the functional form for N^* can be heuristically motivated by assuming the frequency distribution is of the same form (gamma distributed), but with a modified rate parameter $\alpha_m = 4Nm\sigma$ and shape parameter $\theta_m \sim 4N\xi\mu r^*_{m-1} \sim 4N\xi\mu(\xi\mu/\sigma)^{m-1}$, where r^*_{m-1} is the frequency of mutants with $m-1$ R alleles and a single W in mutation-selection balance before drive is introduced. For example, if there are $m = 2$ gRNAs then r^*_1 is the frequency of RW, while for $m = 3$ RNAs r_2 is the frequency of RRW. We show in the Supplementary material that $r^*_{m-1} \sim (\xi\mu/\sigma)^{m-1}$. θ_m represents the effective mutational flow from $m-1$ mutants to m mutants, which is balanced by negative selection, where each m mutant has fitness cost $m\sigma$, before the introduction of drive. By averaging the probability of fixation over this distribution we recover the scaling law for N^*_s stated above. We do not attempt to calculate an exact theory, but use a fitting parameter γ to represent the scaling between these heuristic considerations and an exact theory.

We first fit the curves of probability of resistance vs N for $m = 1$, which gives an approximate value of $s_b \approx 0.5$ for all values of ξ (Supplementary Material), where γ does not appear in N^*_s. For $m > 1$, we fix s_b to this value and fit the curves of $p(N)$ for the single fitting parameter γ, and find that $\gamma \approx 2$ for both $m = 2$ and $m = 3$, as shown in the Supplementary material. In all cases, we find the curves fit the simulation data very well.

We summarise these results by plotting contours of constant probability of resistance $p = 0.05$ as a function of ξ and N as shown in Fig.4b; regions to the left of each curve represent combinations of ξ and N, for which the probability of resistance is less than 5%. The major effect of preexisting mutations is to very significantly reduce the effective population size needed before resistance arises for more than $m > 1$ gRNAs. Importantly, unlike for de novo SNPs or NHEJ mutations, there is a significant effect for changing selection, where the probability of resistance increases for decreasing σ, because less harmful mutations will segregate at a higher frequency before release. Though Hermisson and Pennings (*15*) showed this to be the case for a single site ($m = 1$) (*6,15*), the effect, as we see is only weak and logarithmic in σ; here however, for $m > 1$, $N^* \sim \sigma^{m-1}$, which represents a significant amplification of more weakly selected standing variation.

Probability of resistance: NHEJ, de novo and preexisting SNPs combined

Finally, we combine all three mechanisms by which resistance can arise to assess their relative importance as a function of varying β, ξ and σ. The summary of these results are shown in Fig.5, for contours of $p = 0.05$ for β vs N, for the case of a) all SNPs being functional ($\xi = 1$) and b) 1% functional SNPs ($\xi = 0.01$). The same broad trends as observed in the simulations of the previous sections are seen, where decreasing β increases the range of population sizes for which resistance

is improbable. However, in the presence of preexisting SNPs, we find that for sufficiently small β, this effect plateaus and the 5% contours are constant with respect to N, as the dominant and more rapid mechanism of resistance becomes SNPs from standing variation. We also find that increasing ξ decreases the population size at which resistance arises at 5%, given by a uniform shift of these curves to the left, as seen by comparing panel a) to b). The solid lines in each plot correspond to assuming the probability of resistance from NHEJ mutations and preexisting (& de novo) mutations are independent: $p_{res} = 1 - (1 - p_n)(1 - p_s) = 1 - e^{-N/N^*}$, where

$$\begin{aligned} \frac{1}{N^*} &= \frac{1}{N_n^*} + \frac{1}{N_s^*} \\ &= 4\gamma_n(\epsilon\nu\beta)^m + 4\xi\mu m \left(\frac{\gamma\xi\mu}{\sigma}\right)^{m-1} \ln\left(1 + \frac{s_b}{m\sigma}\right). \end{aligned} \tag{9}$$

The plots use the parameters $\gamma_n = 0.2$ and $\gamma_s = 2$, as derived from fitting in the previous sections, where the different line thicknesses correspond to different values of σ; we see that asymptotically for either $\beta \to 1$ $(1/N_n^* \gg 1/N_s^*)$ or $\beta \to 0$ $(1/N_n^* \ll 1/N_s^*)$, where NHEJ mutations are dominant and SNPs are dominant respectively, the solid lines match the simulation data very well, as expected. However, for the intermediate regime where NHEJ mutations and SNPs from standing variation are equally dominant, there is a mismatch, which indicates the assumption that resistance from NHEJ and SNPs are independent is relatively poor. This is as expected, since this heuristic ignores the mosaic nature in which resistance arises at multiple target sites, where for example at one site resistance may have arisen by standing variation and the other two by NHEJ; nonetheless we see this heuristic gives a good guide to the probability of resistance in the presence of both NHEJ and standing SNP variation.

Time to Resistance

In the Supplementary material we also examine the mean time taken for resistance to arise, *given* that resistance arose, where resistance corresponds to the establishment and then fixation of a functionally resistant allele. We find across all simulations, very consistently, the mean time is between 70 and 100 generations, where as expected in general for large values of β we have shorter times, since the rate of producing NHEJ mutants is larger and as the population size is increased the time to resistance increases. Broadly, we find that if variants are more deleterious before the introduction of drive and/or as the population size increases, that the time to resistance is longer. These results point to the fact that the time for resistance to arise is conditioned on resistance arising, and so all the variants that are destined to give resistance, must be generated, or pre-exist, before the population is eliminated; population elimination happens on a timescale set by the dynamics of drive replacing the wild type — which does not change for different simulations performed in this paper — and by the size of the initial population, where larger populations take a logarithmically longer time to elimination (on average).

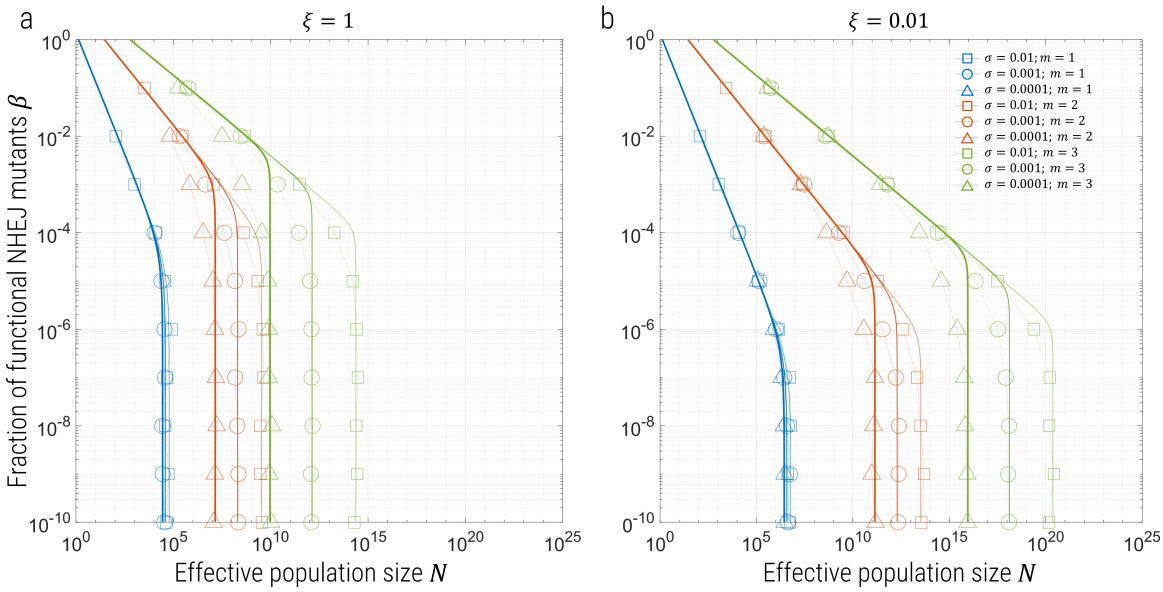


Figure 5: Contours for probability of resistance $p = 0.05$ for β vs N and a) $\xi = 1$ and b) $\xi = 0.01$, for simulations including NHEJ, de novo and pre-existing SNPs. Blue symbols represent $m = 1$, red $m = 2$ and yellow $m = 3$, where the different symbols present different fitness cost σ of mutants in the presence of wild type, as shown in the legend. The solid lines are plots of $1 - e^{-N/N^*}$, for $1/N^* = 1/N_n^* + 1/N_s^*$ — which assumes the probability of resistance from NHEJ and SNPs are independent — for exactly the same values of the γ parameters in the previous sections/plots, where the thick lines correspond to $\sigma = 0.0001$, intermediate thickness to $\sigma = 0.001$ and thin lines to $\sigma = 0.01$.

Discussion

Understanding and overcoming resistance in suppression drive systems is a major obstacle to successful control of natural populations that are vectors for disease such as malaria (*3, 4*). As this work and previous research (*5–7*) shows, population size is a key determinant of the probability of resistance to arise. Here, we find universally across all simulations that the probability of resistance evolving takes the form $p = 1 - e^{-N/N^*}$, where N^* is a summary measure of how resilient suppression drive systems are to resistance; if for natural populations $N \ll N^*$ then resistance is unlikely, while if $N \gg N^*$ then resistance and population rescue is very likely. We find that in general, $\frac{1}{N^*}$ corresponds roughly to the inverse of the per individual rate of generating mutants; in other words, when the population is sufficiently large such that there are ~ 1 resistance mutations generated per generation in the whole population, then the probability of resistance is large. We have explored how N^* depends on the underlying mechanisms of resistance using stochastic Wright-Fisher simulations for up to 3 gRNAs (cut sites) using a novel approximation to the multinomial distribution which allows simulations for arbitrarily large population sizes. In particular, we have looked at the effect of two important parameters, β, the fraction of functional NHEJ resistance mutants, and ξ, the fraction

of functional SNP resistance mutants, as well as the role of standing variation vs de novo mutation.

We heuristically motivate this universal form for the probability of resistance $p = 1 - e^{-N/N^*}$, for NHEJ and de novo SNPs by considering what is the probability of fixation of a certain number of mutants generated in some sufficient time before drive eliminates the population, where each mutant has the same probability of fixation π. As detailed in the Supplementary Information this looks like $p = 1 - (1-\pi)^n \approx 1 - e^{-n\pi}$, where $n \sim N \times (\text{rate of generation of mutants})$. In reality, the probability of fixation is a time-varying quantity (*6, 14*), as selection for resistance depends on the frequency of drive, but our results show the correct functional form can be obtained for multiplexed targets by these simple considerations up to a scaling constant, which we determine empirically from simulation. For standing variation, we also find the same simple form, but following Hermisson & Pennings (*15*), we show in the Supplementary Information it arises from averaging the probability of fixation of mutants, once drive is introduced, given a frequency distribution of multi-site mutants in mutation-selection balance. A key result of the paper is to show that in the presence of standing variation for multiplex drive the probability of resistance is extremely sensitive to how deleterious resistance mutants are before the introduction of drive.

When we examine how resistance arises for multiplexed gRNAs and allowing only NHEJ mutants, we find a critical population size $N_n^* = \frac{1}{4\gamma_n(\epsilon\nu\beta)^m}$ (Eqn.6), where a single value $\gamma_n \approx 0.2$ for the scaling constant, fits the simulation data very well for all values of m and with no dependence on σ, the fitness cost of functional mutants before the introduction of drive. The rate per individual of generating functional NHEJ mutants that are resistant at all m sites is $(\epsilon\nu\beta)^m$, and so this result says that at the critical population size $N = N_n^*$, the population level rate of generating functional NHEJ mutants is $2N(\epsilon\nu\beta)^m \approx \frac{1}{2\gamma_n} = 2.5 \sim 1$ per generation, and resistance via population rescue occurs with high probability. As we show in the Supplementary Information, this particular form of N_n^*, which has a single constant γ_n for all m, is consistent with the concurrent generation of mutants that are resistant at all m sites being much faster than the accumulation of resistance at each target site sequentially over generations. To be clear, both pathways, concurrent and sequential generation of m-fold mutants, have a rate that scale as $(\epsilon\nu\beta)^m$, but they have different pre-factors determining their relative importance. Using this result we can find the required number of gRNAs to keep the probability of resistance to less than or equal to p:

$$m > \frac{\ln\left(\frac{1}{4\gamma_n}\ln\left(\frac{1}{1-p}\right)\right) - \ln(N)}{\ln(\epsilon\nu\beta)}. \tag{10}$$

This result is similar to Equation 2 of (*7*), the main difference being that it is derived by approximately solving the dynamics of resistance allele generation, using their probability of fixation (Supplementary Information) and explicitly accounting for a fraction β of functional NHEJ mutants.

A similar analysis of how de novo SNPs give rise to resistance with m gRNAs gives rise to a critical population size $N_n^* = \frac{1}{4\gamma_m(\xi\mu)^m}$ (Eqn.7), which is of similar form to NHEJ-only mutants, but now where the scaling constant γ_m varies as we vary m. We find in fitting to the simulation data that $\gamma_m = \{0.76, 6, 55\}$ for $m = \{1, 2, 3\}$, respectively. As for NHEJ mutants, there is no dependence on σ the fitness cost of SNP mutants before the introduction of drive. As the rate per individual of generating functional resistance at all m sites is $(\xi\mu)^m$, this means the population level rate needed for the probability of resistance to be large is $2N(\xi\mu)^m \approx \frac{1}{2\gamma_m} \approx \{0.658, 0.083, 0.009\}$ per generation, for $m = \{1, 2, 3\}$, respectively; this indicates that for the same individual rate of

producing functional mutants at all m sites, the critical population size for de novo generation of mutants is smaller than for NHEJ generation of mutants and the disparity increases as the number of gRNAs m increases. That is, for the same rate of generation of functional mutants, resistance arises more easily for de novo SNPs vs NHEJ mutants. This can be rationalised by understanding the underlying mode of generation of de novo mutants: as we show in the Supplementary Information the particular m dependence of the constant is $\gamma_m = s_b \tau^m$, where $s_b \approx 0.09$ is an effective aggregate beneficial selection coefficient of resistance mutants in the presence of drive, which reasonably we would expect to be independent of m, and $\tau \approx 8.5$ generations, is the timescale over which mutants are generated. As we show, this form arises if sequential accumulation of resistance alleles till all m sites are resistant is faster than their concurrent generation m-fold resistance; this is typically true if $\xi\mu \ll 1$, which given a target site mutation rate of $\mu = 5.4 \times 10^{-8}$ and that $\xi \leq 1$ is clearly satisfied.

When we allow there to be standing variation of functional resistance mutants in mutation selection balance with fitness cost σ before the introduction of drive, we find $1/N_s^* = 4\xi\mu m \left(\frac{\gamma\xi\mu}{\sigma}\right)^{m-1} \ln(1+\frac{s_b}{m\sigma})$ (Eqn.8), which unlike for NHEJ and de novo SNPs, now has a dependence on σ. For $m = 1$, this critical population size corresponds to the theory of Hermisson & Pennings (*15*), which arises by averaging the probability of fixation of mutants with selective advantage s_b in the presence of drive, over their equilibrium frequency distribution, before drive is introduced; if we were instead to take the deterministic frequency, as is common (*16*) then this leads to an overestimate of the probability of fixation. Fitting to our simulation data for $m = 1$, we find $s_b \approx 0.5$, where it should be understood that this is an effective aggregate selection coefficient, which could in principle be derived from a more detailed time-dependent theory (*6, 14*). It also makes the assumption that selection for resistance is strong enough for fixation to happen rapidly, such that averaging the ultimate probability of fixation is appropriate. For $m > 1$, we see the critical population size picks up a much stronger power law dependence $N_s^* \sim \sigma^{m-1}$ than the logarithmic dependence for $m = 1$. As we show in the Supplementary Information this can be derived heuristically if we assume that the frequency distribution of m-fold resistance mutants, arises from an effective mutation-selection balance, before drive, between mutations from $m-1$-fold mutants and selection against these mutants before the introduction of drive. As the frequency of the $m-1$-fold mutants scale as $\sim (\xi\mu/\sigma)^{m-1}$, this leads to the power law seen with respect to σ, where the initial frequency of resistance mutants is strongly reduced as σ increases, which is then reflected in a reduction in the probability of resistance. For $m > 1$, if we fix s_b to its value for $m = 1$, then there is only a single fitting parameter for the functional form given, which we find is $\gamma \approx 2$, and independent of m.

Given the expression for N_s^* and that the probability of resistance $p = 1 - e^{-N/N_s^*}$, we can in principle calculate the minimum number of gRNAs m required to prevent resistance to a specified probability. Eqn.8 is transcendental, but if we replace the weak m dependence in the logarithm $(\ln(1+\frac{s_b}{m\sigma}) \to \ln(1+\frac{s_b}{\sigma}))$ then for $\sigma \gg \xi\mu$, we find the approximate expression:

$$m > \frac{1}{\ln(\frac{\gamma}{4N\sigma})} W_{-1}\left(\frac{\gamma\xi\mu}{\sigma}\frac{\ln(\frac{1}{1-p})\ln(\frac{\gamma\xi\mu}{\sigma})}{\ln(1+\frac{s_b}{\sigma})}\right), \tag{11}$$

where $W_k(x)$ is the Lambert W function, which are solutions of the equation $we^w = x$.

By inspecting Figs.4a&b we can see that in general the contours where the probability of resistance $p = 0.05$ for standing variation and de novo mutation (b) are shifted to smaller population sizes compared to de novo resistance only for all values of σ, which indicates that resistance arises more

easily and at smaller population sizes with standing variation. This has been previously noted for single sites $m = 1$ in various evolutionary contexts including the question of which is more important, de novo vs standing variation for adaptive evolution (*15,18*), population rescue (*12*) or in the context of the evolution of gene drive resistance (*6*). All these studies show that changing the magnitude of the fitness cost σ has a relatively weak effect on the probability of resistance, as borne out by the logarithmic dependence of N_s^* on σ for $m = 1$ in Eqn.8. However, a major new finding is that for the multiplex drive case, where resistance alleles must evolve at all m target sites in order for resistance to arise, there is in fact a marked dependence on the deleteriousness of resistance alleles before the introduction of drive. This result means that there is a significant amplification of weakly deleterious alleles in their contribution to resistance in a multiplex scenario, and a significant reduction in the critical population size with standing variation compared to de novo mutation, as seen in Fig.4b. We can quantify this amplification by calculating the ratio

$$\frac{N_d^*}{N_s^*} = \frac{m\gamma^{m-1}\ln\left(1+\frac{s_b}{m\sigma}\right)}{\sigma^{m-1}\gamma_m}, \tag{12}$$

which has values $\frac{N_d^*}{N_s^*} \approx \{5, 217, 6266\}$ for $m = \{1, 2, 3\}$ and $\sigma = 0.01$, and $\frac{N_d^*}{N_s^*} \approx \{11, 5.2 \times 10^4, 1.6 \times 10^8\}$ for $\sigma = 10^{-4}$, which are very large amplification factors for $m > 1$ and with consequent implications for the prediction of the probability of resistance for multiplexed gene drives. (Note that these very large differences are somewhat hidden in Fig.4 as the results are plotted on a log scale and over many orders of magnitude.)

Simulations when all three mechanisms are allowed to give rise to resistance (NHEJ, de novo SNPs and SNP standing variation) show that the dominant underlying mechanism of resistance depends on the relative importance of NHEJ mutants vs standing variation, embodied in a comparison between N_n^* and N_s^*, where we have already established that in general the role of de novo SNPs will be weak, particularly when $m > 1$. We find that when $N_n^* \gg N_s^*$, $N^* \approx N_s^*$ – which occurs when β is sufficiently small – the role of standing variation is dominant, and the value of σ critically determines the probability of resistance, as discussed in the previous paragraph. On the other hand, when $N_n^* \ll N_s^*$, $N^* \approx N_n^*$, and there is no dependence on σ. In the intermediate regime, where $N_n^* \sim N_s^*$, both mechanisms contribute roughly equally to resistance. Interestingly, we find that this intermediate regime is not very accurately described by assuming that each mechanism acts independently, and requires calculation of the probability of mosaic resistance, where resistance at different sites can arise from either NHEJ or SNPs. However, Fig.5 shows the qualitative behaviour is correct, and that broadly, as we increase the number of gRNAs m, the critical value of β, which corresponds to when both NHEJ and standing variation contribute roughly equally to resistance tends to increase as m increases, and decreases as ξ increases.

Our modelling has highlighted 5 key parameters determining the probability of resistance evolving: β, ξ, σ, m, and N. Researchers developing gene drive constructs for population suppression will have some level of control over all but the last of these, by choice of target site. NHEJ typically produces insertion or deletion mutations (*3*), so choosing a target site that is unable to tolerate length variation will be one way to reduce β, and this would usually be a top priority in choosing a target site. Hammond et al. (*3*) have demonstrated that having a single target site that can tolerate length variation quickly leads to the evolution of resistance, even in small populations ($N \approx 600$). Though data are scarce, the next most frequent type of mutation produced by NHEJ is presumably single

nucleotide changes at the cut site, and ensuring those are nonfunctional would be a second priority. Assuming a target population of $N = 10^6$ and the baseline parameter values, Eqn.10 indicates that under the worst case scenario of $\beta = 1$, $m = 6$ gRNAs will be needed to have the probability of resistance arising due solely to NHEJ be less than 0.05, whereas if β can be reduced to 0.01, then only $m = 3$ will be needed, and if β can be reduced further to 10^{-4}, then only 2 will be needed. If the target population is $N = 10^9$, then the corresponding values are $m = \{8, 4, 2\}$ gRNAs for $\beta = \{1, 0.01, 10^{-4}\}$, respectively.

Single nucleotide mutations can also arise spontaneously, and, if the population is large, may be already present before release. These may also provide resistance against cleavage, particularly if they are near the cut site (*19*). The likelihood that resistance evolves from these mutations depends on ξ, the fraction of them that are functional (since it is only these that can spread through a population) and on σ, the extent to which those functional resistant mutations reduce fitness (since that affects their frequency in the pre-release population). Again, all else being equal, more functionally constrained sequences will be preferred. In principle, genetic surveys of sequence variation at the target site may provide useful information on ξ and σ and the probability of resistance evolving through standing variation. Assuming a target population of $N = 10^6$, baseline parameter values and $\sigma = 10^{-6}$, then using Eqn.11 under the worst case scenario of $\xi = 1$, $m = 4$ gRNAs are needed to ensure the probability of resistance evolving from de novo mutations and pre-existing variation is less than 0.05, whereas just $m = 2$ gRNAs are needed if $\sigma = 10^{-4}$ or $\sigma = 0.01$, which demonstrates the sensitivity to weakly deleterious standing variation. However, if $\xi = 0.01$ then only 1 gRNA is needed, irrespective of the value of σ. If the target population is $N = 10^9$ the corresponding values are $m = \{7, 3, 2\}$ gRNAs are needed respectively, for $\xi = 1$ and $\sigma = \{10^{-6}, 10^{-4}, 0.01\}$, whereas for $\xi = 0.01$, $m = 2$ gRNAs are needed, regardless of σ.

An important application of suppression drives is to control populations of mosquitoes to reduce the burden of malaria on human populations. Recently, the contemporary effective population size for *Anopheles gambiae* in sub-Saharan Africa was estimated as $N \sim 10^9$, using a new method based on analysing soft sweeps (*9*). The above considerations of NHEJ and standing variation separately give an indication of how many gRNAs are required to achieve a probability of resistance less than 0.05, in each scenario. However, as Fig.5 demonstrate, when both mechanisms of resistance are possible, the effective critical population size N^* is a combination of N_n^* and N_s^*. Eqn.9 is a reasonable approximation, but it is difficult to invert this expression to find how many gRNAs are needed to prevent resistance to a certain probability p. Here, we instead estimate this critical population size and compare to our target population size of $N = 10^9$ for *An. gambiae*, for different values of m, β, ξ and σ by reading off Fig.5 or by extrapolation using the approximation in Eqn.9. As mentioned, we do not currently have strong empirical evidence for the value of β, but we can examine two different extremes: 1) $\beta = 10^{-4}$, which corresponds to assuming functional NHEJ mutants are quite rare, as we might expect given the expectation that NHEJ will tend to produce significant genetic changes like multiple base-pair insertions and deletions and 2) $\beta = 10^{-2}$, which is more pessimistic, should in fact NHEJ more readily produce less significant genetic changes, which are more likely to be functional. In a worst case that $\xi = 1$, we can read off from Fig.5 and our extrapolation of the simulations using Eqn.9 to different values of σ, to calculate the critical population sizes when resistance is equal to 5% in Table.1. These numbers indicate that if resistance alleles are strongly deleterious ($\sigma = 0.01$) to moderately deleterious ($\sigma = 10^{-4}$) then resistance can be prevented with $m = 3$ gRNAs if $\beta = 10^{-4}$, or $m = 4$ gRNAs if $\beta = 10^{-2}$. However, if the resistance alleles are

$\boldsymbol{\xi = 1, \beta = 10^{-4}}$

	$m = 1$	$m = 2$	$m = 3$	$m = 4$(*)
$\sigma = 10^{-6}$(*)	8×10^3	9×10^4	6×10^5	4×10^6
$\sigma = 10^{-4}$	10^4	3×10^6	4×10^9	7×10^{12}
$\sigma = 10^{-2}$	10^4	4×10^8	2×10^{13}	2×10^{19}

$\boldsymbol{\xi = 1, \beta = 10^{-2}}$

	$m = 1$	$m = 2$	$m = 3$	$m = 4$(*)
$\sigma = 10^{-6}$(*)	100	7×10^4	6×10^5	4×10^6
$\sigma = 10^{-4}$	100	7×10^4	3×10^7	10^{12}
$\sigma = 10^{-2}$	100	3×10^5	5×10^8	10^{12}

$\boldsymbol{\xi = 0.01, \beta = 10^{-4}}$

	$m = 1$	$m = 2$	$m = 3$	$m = 4$(*)
$\sigma = 10^{-6}$(*)	10^4	7×10^8	6×10^{11}	4×10^{14}
$\sigma = 10^{-4}$	10^4	4×10^8	3×10^{13}	10^{20}
$\sigma = 10^{-2}$	10^4	3×10^9	3×10^{17}	10^{20}

$\boldsymbol{\xi = 0.01, \beta = 10^{-2}}$

	$m = 1$	$m = 2$	$m = 3$	$m = 4$(*)
$\sigma = 10^{-6}$(*)	100	3×10^5	6×10^8	10^{12}
$\sigma = 10^{-4}$	100	3×10^5	3×10^9	2×10^{13}
$\sigma = 10^{-2}$	100	3×10^5	6×10^8	2×10^{13}

Table 1: Effective population sizes N at which a probability of resistance $p = 0.05$ is obtained from simulations and theory for different numbers of gRNAs m and different selection coefficients σ, for the fixed value of fraction of functional mutants $\xi = 1$ and $\xi = 0.01$. The starred values indicate population sizes extrapolated from simulations results using heuristic theory (Eqn.9).

very weakly deleterious ($\sigma = 10^{-6}$) then even 4 gRNAs is not sufficient, for both values of β, which exemplifies the strong amplification of the probability of resistance in the presence of weakly selected standing variation.

On the other hand, we can examine the scenario that $\xi = 0.01$, which roughly corresponds to on average a single functional mutant in all the $3L$ one-step mutants about the wild type sequence in a target site of size $L = 18$bp. The effective population sizes at which a 5% probability of resistance is attained is shown in Table.1 for both $\beta = 10^{-4}$ and $\beta = 10^{-2}$. Again if we assume $N = 10^9$, we see that now resistance can be prevented to a probability less than 5% for all values of σ and β with $m = 3$ gRNAs, except for very weakly deleterious mutants ($\sigma = 10^{-6}$) and $\beta = 10^{-2}$, which requires at least $m = 3$ gRNAs. These considerations highlight the need to empirically determine both β and ξ for putative target sites for multiplex suppression drive applications, using genetic screens which determine the functionality of NHEJ mutants and by examining SNP variants and their fitness effects from population genomic data.

These simulations make a number of simplifying assumptions that are necessitated by modelling the already relatively complex situation of multiplex drive. We assume that all of the nucleotides ξ corresponding to the fraction of functional mutations in each target site, once mutated completely abolish cleavage and give complete resistance. These assumptions can be relaxed in various ways,

such that each mutation contributes to partial reduction in cleavage efficiency; within our model this would correspond to a different effective value of ξ. We also assume there is no recombination between sites and related to this that resistance only arises due to target site effects; recombination and off-target resistance has been explored using deterministic modelling (*20, 21*). We also have not explored the role of varying a number of parameters like cleavage efficiency ϵ, the intrinsic growth rate R_m, or the fitness parameters of drive; some of these have been previously explored in the context of single gRNAs (*6, 7*) and these works and ours in general indicate these will have a secondary role compared to the population scaled rate of NHEJ and de novo mutation. For example, increasing ϵ will tend to increase the rate that drive replaces wild type, and we may expect one consequence is a quantitative reduction in the probability of resistance as there is less time to generate mutants and establish before population elimination, but the qualitative results we have presented will not change. There is also the possibility that homing events may be more error-prone than normal DNA replication and could lead to the loss of function of one or more gRNAs, which we leave to future modelling efforts, but could increase the probability of resistance evolving depending on the rate of such events occurring.

On one hand, it is not clear that spatial structure will strongly affect these results, since the population size we consider here should be very close to the census size; whether spatially separated or in a well-mixed system, the number of mutations arising per generation will depend on the total number of individuals in the population. However, what is likely to be different is how quickly a resistance mutation establishes and spreads; in spatially structured populations selection is effectively weaker (*22*), reduced by factor $1 - F_{ST}$, where $F_{ST} = 1/(1 + 4Nm)$ is Wright's fixation index for the island model, which means fixation would be less rapid in a geographically dispersed population. This could lead to an increase in the critical population size N^*, since at a given population size the probability of resistance is smaller, as elimination is more likely before the resistance allele becomes sufficiently prevalent in the population. However, this is likely to only have significant effect in very highly structured populations (very limited migration), since typically selection for resistance mutants is very large once drive has risen to large frequency; however, if spatial structure is important, it would mean a systematic underestimation of the critical population size N^*. Importantly, in models with spatial structure even arbitrarily strong gene drives may not eliminate a target population (*23–25*). This effect can arise if the gene drive causes reductions in population density, which leads to increased inbreeding, which in turn reduces the efficacy of the drive (*26, 27*). If the population is not eliminated, then eventually one would expect resistance to evolve, though if the population is substantially suppressed this may take a long time.

Our finding of the amplification of standing variation and in particular the expression for the critical population size for resistance to arise (Eqn.8) should have general applicability to evolution of multi-site resistance in many different contexts (*28*). This includes multi-drug resistance to antibiotics (*29, 30*), anti-viral therapies (*31*) and cancer (*16, 32*). The theory is particular applicable to the evolution of antibody/vaccine escape mutants, where multiple epitopes on a virus structure are targeted by antibodies to reduce infection, and where recombination is likely to play a small role. Although each of these evolutionary scenarios carry with it particular idiosyncracies, which would require proper attention, qualitatively, it is clear that very weakly deleterious standing variation can very easily give rise to small critical population sizes even for moderately large m.

A key case in point is the evolution of vaccine/antibody escape of Sars-Cov2, and viruses in general, where a large number of epitope sites are targeted by neutralising antibodies (*33–38*). If

infections are persistent, as currently with Sars-Cov2 and other endemic diseases, we do not strictly have a population rescue problem, but Eqn.8 can be viewed as the critical population if a long time was allowed for standing variation to develop. In the case of Sars-Cov2 as many as $m = 20$ (*36*) epitope sites may need to evolve to prevent antibody binding, whilst for measles only $m = 5$ may be required (*38*). The results here (Eqn.8) would broadly suggest that higher levels of infection could increase levels of standing variation and make vaccine escape very probable. Although it is too early to assign specific selective causes, if it is the escape of natural or vaccine-induced immunity that has led to the recent emergence of the B.1.1.529 (Omicron) variant with roughly 30 mutations in the spike-ACE2 receptor binding protein, then these mutations could potentially have accumulated via this mechanism, prior to whatever change in selective pressures that allowed its establishment; this may have occurred in a single prolonged infection in an immunocompromised individual, and/or through chains of transmission between a number of acutely infected individuals. We can understand qualitatively how such a large number of mutations *could* arise, if they were weakly deleterious, even for moderate effective populations sizes, since we can re-arrange Eqn.8 (ignoring the weak logarithmic term and assuming $4\mu\xi m N_s^* > 1$) to give $\sigma \approx \gamma\mu\xi(4\mu\xi m N_s^*)^{\frac{1}{m-1}}$, which for large m decreases extremely slowly; in other words the value of σ required for say $m = 30$ mutations to give rise to resistance will be marginally smaller than that required for $m = 10$. However, the applicability to Sars-Cov2 is arguable as the time since the start of the pandemic may not be sufficient for multi-site variants to reach their equilibrium frequency, although this modelling and theory provides a framework to address this question.

Overall, our results provide a foundation to understand how resistance arises in multiplexed suppression-drive systems and the paramount role that standing variation plays in amplifying weakly deleterious variation to give resistance at even moderate population sizes for a large number m of gRNAs. The results highlight the need to characterise important unknown parameters such as the fraction of functional mutants at drive target sites, due both to non-homologous end joining mutations and single nucleotide mutations, which can significantly affect the probability that resistance arises. Finally, we suggest the multi-site evolutionary rescue problem studied here and the role of standing variation has general applicability to multi-drug antibiotic, anti-viral and cancer resistance, as well as the evolution of vaccine escape.

Acknowledgments

We acknowledge useful discussions with Ioanna Morianou on the genetics of resistance and comments on the manuscript by John Connolly and John Mumford. This work was supported by grants from the Bill & Melinda Gates Foundation and the Open Philanthropy Project.

References

1. A. Burt, *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**, 921 (2003).

2. M. Jinek, *et al.*, *Science* **337**, 816 (2012).

3. A. M. Hammond, *et al.*, *PLOS Genetics* **13**, e1007039 (2017).

4. K. Kyrou, *et al.*, *Nature Biotechnology* (2018).

5. A. Beaghton, P. J. Beaghton, A. Burt, *Malaria Journal* **16**, 286 (2017).

6. R. L. Unckless, A. G. Clark, P. W. Messer, *Genetics* **205**, 827 (2017).

7. J. M. Marshall, A. Buchman, H. M. S. C., O. S. Akbari, *Scientific Reports* **7**, 3776 (2017).

8. J. Champer, *et al.*, *PLOS Genetics* **13**, e1006796 (2017).

9. B. S. Khatri, A. Burt, *Molecular Biology and Evolution* (2019).

10. C. Noble, J. Olejarz, K. M. Esvelt, G. M. Church, M. A. Nowak, *Science Advances* **3**, e1601964 (2017).

11. G. C. Lanzaro, H. M. S. C., T. C. Collier, J. M. Marshall, A. A. James, *BioEssays* **43**, 2000282 (2021).

12. H. A. Orr, R. L. Unckless, *The American naturalist* **172**, 160 (2008).

13. H. A. Orr, R. L. Unckless, *PLoS genetics* **10**, e1004551 (2014).

14. H. Uecker, J. Hermisson, *Genetics* **188**, 915 (2011).

15. J. Hermisson, P. S. Pennings, *Genetics* **169**, 2335 (2005).

16. Y. Iwasa, F. Michor, M. A. Nowak, *Journal of Theoretical Biology* **226**, 205 (2004).

17. P. D. Keightley, R. W. Ness, D. L. Halligan, P. R. Haddrill, *Genetics* **196**, 313 (2014).

18. T. Karasov, P. W. Messer, D. A. Petrov, *PLoS genetics* **6**, e1000924 (2010).

19. T. Zheng, *et al.*, *Scientific Reports* **7**, 40638 (2017).

20. R. Gomulkiewicz, M. L. Thies, J. J. Bull, *Genetics* **217**, iyaa040 (2021).

21. F. Cook, J. J. Bull, R. Gomulkiewicz, *bioRxiv* p. 2021.08.30.458221 (2021).

22. J. L. Cherry, J. Wakeley, *Genetics* **163**, 421 (2003).

23. A. R. North, A. Burt, H. C. J. Godfray, *BMC Biology* **17**, 26 (2019).

24. A. R. North, A. Burt, H. C. J. Godfray, *BMC Biology* **18**, 98 (2020).

25. J. Champer, I. K. Kim, S. E. Champer, A. G. Clark, P. W. Messer, *Molecular Ecology* **30**, 1086 (2021).

26. J. J. Bull, C. H. Remien, R. Gomulkiewicz, S. M. Krone, *PeerJ* **7**, e7921 (2019).

27. P. Beaghton, A. Burt, *bioRxiv* p. 2021.11.11.468225 (2021).

28. B. A. Wilson, N. R. Garud, A. F. Feder, Z. J. Assaf, P. S. Pennings, *Molecular Ecology* **25**, 42 (2016).

29. B. Tepekule, H. Uecker, I. Derungs, A. Frenoy, S. Bonhoeffer, *PLoS Computational Biology* **13**, e1005745 (2017).

30. D. V. McLeod, S. Gandon, *eLife* **10**, e65645 (2021).

31. P. S. Pennings, *PLoS Computational Biology* **8** (2012).

32. I. Bozic, *et al.*, *eLife* **2**, e00747 (2013).

33. M. Yuan, *et al.*, *Science* **369**, 1119 (2020).

34. L. Piccoli, *et al.*, *Cell* **183**, 1024 (2020).

35. A. J. Greaney, *et al.*, *Nature Communications* **12**, 4196 (2021).

36. F. Schmidt, *et al.*, *Nature* pp. 1–5 (2021).

37. D. V. Egeren, *et al.*, *PLOS ONE* **16**, e0250780 (2021).

38. M. Muñoz Alía, R. A. Nace, L. Zhang, S. J. Russell, *Cell Reports Medicine* **2**, 100225 (2021).

39. A. Deredec, A. Burt, H. C. J. Godfray, *Genetics* **179**, 2013 (2008).

Supplemental Materials

Methods

Single guide RNA

Simulations with a single guide RNA (gRNA) or target site require specification of a 4×4 fitness matrix for both males $\boldsymbol{W}_{\mathrm{m}}$ and females $\boldsymbol{W}_{\mathrm{f}}$, whose elements w_{ij} are the fitness of the genotype i/j, where for example genotype $1/3 \equiv$ W/N. In the absence of mutation and drive the allele frequency vector in next generation of females $\boldsymbol{x}_{t+1}$ and males $\boldsymbol{y}_{t+1}$ can be expressed succinctly in matrix notion in terms of the frequency vectors in the current generation $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$ as follows

$$\boldsymbol{x}_{t+1} = \frac{(\boldsymbol{W}_{\mathrm{f}}\boldsymbol{y}_t)\odot \boldsymbol{x}_t + (\boldsymbol{W}_{\mathrm{f}}\boldsymbol{x}_t)\odot \boldsymbol{y}_t}{2\bar{w}_{\mathrm{f}}}, \tag{S1}$$

$$\boldsymbol{y}_{t+1} = \frac{(\boldsymbol{W}_{\mathrm{m}}\boldsymbol{x}_t)\odot \boldsymbol{y}_t + (\boldsymbol{W}_{\mathrm{m}}\boldsymbol{y}_t)\odot \boldsymbol{x}_t}{2\bar{w}_{\mathrm{m}}}, \tag{S2}$$

where the mean fitness of males and females is given by the quadratic forms

$$\bar{w}_{\mathrm{f}} = \boldsymbol{x}_t^T \boldsymbol{W}_{\mathrm{f}} \boldsymbol{y}_t \tag{S3}$$

$$\bar{w}_{\mathrm{m}} = \boldsymbol{y}_t^T \boldsymbol{W}_{\mathrm{m}} \boldsymbol{x}_t \tag{S4}$$

and where the $\odot$ operation is the element by element (Hadamard) multiplication operation, i.e. $\boldsymbol{z} = \boldsymbol{x} \odot \boldsymbol{y}$ is equivalent to $z_i = x_i y_i$.

The fitness matrix for males, $\boldsymbol{W}_{\mathrm{m}}$ is a 4×4 matrix, where each entry, $(W_{\mathrm{m}})_{ij} = 1$. For females, we assume all functional genotypes (exc N and D) have fitness $w = (1-\sigma)^{n_R}$, where n_R is the number of R alleles in the genotype, so $w_{11} = w$(W/W) $= 1$, $w_{12} = w_{21} = w$(W/R) $= (1-\sigma)$ and $w_{22} = w$(R/R) $= (1-\sigma)^2$. On the other hand for non-functional heterozygotes, we assume a dominance coefficient h with drive to represent somatic leaky expression affecting viability, with a typical value $h \approx 0.3$, and a very recessive dominance coefficient ($h_N = 0.02$) with non-functional resistance alleles. W/D heterozygotes; heuristically, this leads to the fitness of genotypes that include functional and non-functional alleles of the form $w = (1-\sigma)^{n_R}(1-h_j s)$, where as before n_R is the number of R alleles in the genotype and if $j =$ N, $h_j = h_N$ and if $j =$ D, $h_j = h$. Genotypes with only non-functional alleles N and D are very deleterious and have fitness $w = 1 - s$, where $s = 1$. Altogether, this gives the following fitness matrix for females:

$$\boldsymbol{W}_{\mathrm{f}} = \begin{pmatrix} 1 & 1-\sigma & 1-h_N s & 1-hs \\ 1-\sigma & (1-\sigma)^2 & (1-\sigma)(1-h_N s) & (1-\sigma)(1-hs) \\ (1-\xi)\mu & (1-\sigma)(1-h_N s) & 1-s & 1-s \\ 1-hs & (1-\sigma)(1-h_N s) & 1-s & 1-s \end{pmatrix}. \tag{S5}$$

Including mutations is straightforward using a mutation probability matrix $\boldsymbol{M}$, whose elements M_{ij} represent the probability of generating allele i per generation from allele j, where this stochastic matrix, which acts on probability/frequency *column* vectors must have the property $\sum_i M_{ij} = 1$:

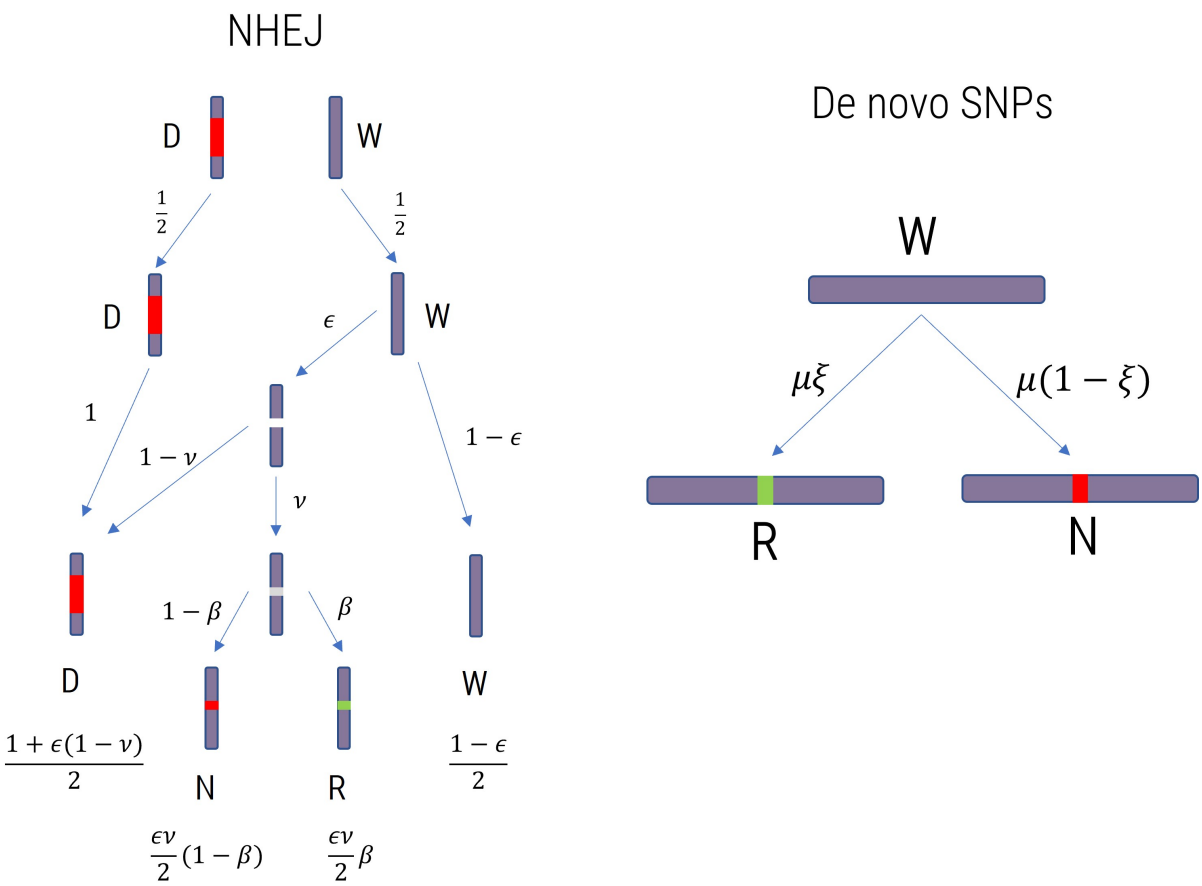


Figure S1: Diagram showing gametogenesis for $m = 1$ gRNAs with NHEJ for a W/D individual (left) and how de novo SNPs arise by mutation (right). On the right, wild type is the background colour, a white marker indicates the cleavage state, a long red bar represents drive inserted at the target site, a grey marker is an NHEJ event, which are functional (green) or non-functional (red).

$$\boldsymbol{x}_{t+1} = \boldsymbol{M}\frac{(\boldsymbol{W}_{\mathrm{f}}\boldsymbol{y}_t)\odot\boldsymbol{x}_t + (\boldsymbol{W}_{\mathrm{f}}\boldsymbol{x}_t)\odot\boldsymbol{y}_t}{2\bar{w}_{\mathrm{f}}}, \tag{S6}$$

$$\boldsymbol{y}_{t+1} = \boldsymbol{M}\frac{(\boldsymbol{W}_{\mathrm{m}}\boldsymbol{x}_t)\odot\boldsymbol{y}_t + (\boldsymbol{W}_{\mathrm{m}}\boldsymbol{y}_t)\odot\boldsymbol{x}_t}{2\bar{w}_{\mathrm{m}}}, \tag{S7}$$

We make the assumption that there are only two mutational paths W $\rightarrow$ R and W $\rightarrow$ N, which happen with rate $\xi\mu$ and $(1-\xi)\mu$, respectively and there are no backmutations to W (Fig.S1. This gives a mutation matrix:

$$\boldsymbol{M} = \begin{pmatrix} 1-\mu & 0 & 0 & 0 \\ \xi\mu & 1 & 0 & 0 \\ (1-\xi)\mu & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{S8}$$

Finally, we add the effect of drive by following (*S39*), where we assume drive acts during meiosis after viability selection. Drive is only produced by heterozygotes W/D, whose frequency after selection and mutation is $(1-\mu)w_{14}^{\mathrm{s}}/\bar{w}_{\mathrm{s}}(x_1y_4 + x_4y_1)$ for females and males, where $\mathrm{s} = \{\mathrm{f}, \mathrm{m}\}$, respectively, which gives the following allele frequency update equations

$$\begin{aligned} \boldsymbol{x}_{t+1} = \boldsymbol{M}\frac{(\boldsymbol{W}_{\mathrm{f}}\boldsymbol{y}_t)\odot\boldsymbol{x}_t + (\boldsymbol{W}_{\mathrm{f}}\boldsymbol{x}_t)\odot\boldsymbol{y}_t}{2\bar{w}_{\mathrm{f}}} \\ + (1-\mu)\frac{w_{14}^{\mathrm{f}}}{2\bar{w}_{\mathrm{f}}}\boldsymbol{\kappa}_{\mathsf{W}}H_t, \end{aligned} \tag{S9}$$

$$\begin{aligned} \boldsymbol{y}_{t+1} = \boldsymbol{M}\frac{(\boldsymbol{W}_{\mathrm{m}}\boldsymbol{x}_t)\odot\boldsymbol{y}_t + (\boldsymbol{W}_{\mathrm{m}}\boldsymbol{y}_t)\odot\boldsymbol{x}_t}{2\bar{w}_{\mathrm{m}}} \\ + (1-\mu)\frac{w_{14}^{\mathrm{m}}}{2\bar{w}_{\mathrm{m}}}\boldsymbol{\kappa}_{\mathsf{W}}H_t. \end{aligned} \tag{S10}$$

where $H_t = (x_1)_t(y_4)_t + (x_4)_t(y_1)_t$ is the drive-heterozygote frequency (before viability selection acts) and the vector $\boldsymbol{\kappa}$ is stoichiometry like vector describing the change in fractions of each alleles produced as gametes from the W/D heterozygote relative to no drive:

$$\boldsymbol{\kappa}_{\mathsf{W}} = (-\epsilon, \epsilon\nu\beta, \epsilon\nu(1-\beta), \epsilon(1-\nu))^T. \tag{S11}$$

Eqns S9 & S10 are then treated as the mean or expected frequency in a Wright-Fisher multinomial sampling process. To allow for fluctuations in relative numbers of males and females, we perform Wright-Fisher sampling with total population size constraint N_{t+1} on the population vector of female and male allele frequencies $\boldsymbol{z}_{t+1} = (\boldsymbol{x}_{t+1}^T F_{t+1}/N_{t+1}, \boldsymbol{y}_{t+1}^T M_{t+1}/N_{t+1})^T$, where now $\boldsymbol{z}_{t+1}$ is a 8 column vector, and for example z_3 is the frequency of female R alleles, and z_5 is the frequency of male W alleles, where F_{t+1}, M_{t+1} and N_{t+1} are the total number of females, males and their sum, which are calculated using a Beverton-Holt scheme described next.

To couple the population genetics calculated here to the population dynamics, we use the Beverton-Holt model of density-dependent growth for male and female populations separately, assuming only females give birth:

$$F_{t+1} = \frac{R_m\bar{w}_{\mathrm{f}}F_t}{1 + N_t/\alpha}, \tag{S12}$$

$$M_{t+1} = \frac{R_m\bar{w}_{\mathrm{m}}F_t}{1 + N_t/\alpha}, \tag{S13}$$

here R_m is the absolute mean number of offspring from females, and α is a parameter controlling density dependent growth, which we typically tune to give a carrying capacity $K = (R_m(\bar{w}_{\mathrm{f}} + \bar{w}_{\mathrm{m}}) - 1)\alpha = (R_m - 1)\alpha = N_0$, the initial population size we set for the population in the absence of drive. In these simulations we take a typical value of $R_m = 6$.

index	haplotype
1	WW
2	WR
3	RR
4	WN
5	RN
6	NN
7	DD

Table 1: Mapping between frequency vector index and haplotypes for 2-fold multiplex

index	haplotype
1	WWW
2	WWR
3	WRR
4	RRR
5	WWN
6	WRN
7	RRN
8	WNN
9	RNN
10	NNN
11	DDD

Table 2: Mapping between frequency vector index and haplotypes for 3-fold multiplex

Multiplexed drive

In general, notational complexity increases considerably when we consider 2-fold and higher orders of multiplexed drive. We will only consider 2-fold and 3-fold multiplex in this paper, but we will set out a general framework for any order of multiplex m.

Firstly, how many different haplotypes n_H are there for each degree m of multiplex? For $n = 3$ alleles at each site (excluding drive), for 2-fold multiplex, as we calculated above there are $n(n+1)/2 = 6$ haplotypes, which is the number of unordered pairs, corresponding to the number of upper diagonal elements for $n \times n$ matrix; including drive, we then need to track a total of $n_H = 7$ haplotype frequencies. For 3-fold multiplex, the number of haplotypes to be tracked is the number of unordered triplets, which is not so trivially calculated, but amounts to calculating the analogue of the number of "upper diagonal" elements of a $n \times n \times n$ tensor. The answer is that there are $n^{(m)}/m! = 3^{(3)}/3! = 10$ haplotypes, excluding drive, for $m = 3$ fold multiplex, where $n^{(m)} = n(n+1)(n+2)...(n+m-1)$ is the rising factorial — note this expression holds true for any m integer. So including drive, for 3-fold multiplex, there are a total of $n_H = 11$ haplotypes frequencies to be tracked.

We also need to decide on the ordering of the haplotypes in the frequency vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. The convention we use is column ordering and its generalisation for higher dimensional tensors, which gives the mapping of indices to haplotypes as shown in Table 1 and Table 2

The fitness matrices for 2-fold multiplex are 7×7, while for 3-fold they are 11×11. For males

their fitness matrices are filled with all ones. We do not explicitly write out the fitness matrix of female genotypes for multiplex drive, but its construction follows the same rules as for 1-fold drive described above, where there are 3 categories of genotypes, where now any haplotype with an N at any site becomes non-functional: 1) only functional haplotypes on both chromosomes, where fitness $w = (1-\sigma)^{n_R}$, where n_R is number of R alleles across both chromosomes (e.g. genotype WR/RR has $n_R = 3$, and $w_{23} = w_{32} = (1-\sigma)^3$); 2) functional/non-functional genotypes have fitness of the form $w = (1-\sigma)^{n_R}(1-h_j s)$ (e.g. genotype WRN/RRR has $n_R = 3$ and $h_j = h_N$, so $w_{46} = w_{64} = (1-\sigma)^3(1-h_N s)$, or genotype WWR/DDD has $n_R = 1$ and $h_j = h$, so $w_{2,11} = w_{11,2} = (1-\sigma)(1-hs)$); 3) non-functional genotypes all have fitness $w = 1-s$ (e.g. w(WWN/DDD) $= w_{5,11} = w_{11,5} = 1-s$).

The mutation matrix, for mutiplexed drive of degree m, $\boldsymbol{M}^{(m)}$, is also a $n_H \times n_H$ matrix, the jth column representing the rate of mutations from the jth haplotype to each of the other haplotypes. We assume that the only mutations possible, at each target site, are W $\rightarrow$ R and W $\rightarrow$ N, with no backmutations. For this reason the only non-zero columns are haplotypes j that have at least a single W allele; of these each W alleles will mutate according the first column of the 1-fold mutation matrix, $\boldsymbol{M}$ (Eqn.S8), whilst the alleles that are not wild type are unaffected by mutation and so are given by the second and third columns of $\boldsymbol{M}$. For mutation probability from the jth haplotype to the ith, if we let n_{W} be the number of W alleles in the jth haplotype, and n'_{W}, n'_{R} and n'_{N} be the number of W, R, N alleles in the subset of the ith haplotype that only has W alleles, then for $m = 2$ the probability of mutation is

$$M^{(2)}_{ij} \equiv M^{(2)}_{k'_1 k'_2, k_1 k_2} = \frac{n_{\mathsf{W}}!}{n'_{\mathsf{W}}! n'_{\mathsf{R}}! n'_{\mathsf{N}}!} M_{k'_1 k_1} M_{k'_2 k_2} \tag{S14}$$

where haplotype $k_1 k_2$ maps to j and haplotype $k'_1 k'_2$ maps to i, using Table 1 and k_1, k_2, k'_1, k'_2 can take values $\{1, 2, 3\}$ corresponding to alleles $\{\mathsf{W}, \mathsf{R}, \mathsf{N}\}$, respectively. The multinomial coefficient accounts for the redundancy in the number of ways haplotype $k'_1 k'_2$ can be obtained by mutation of $k_1 k_2$. Explicitly, evaluating this for $m = 2$ we arrive at the following mutation matrix

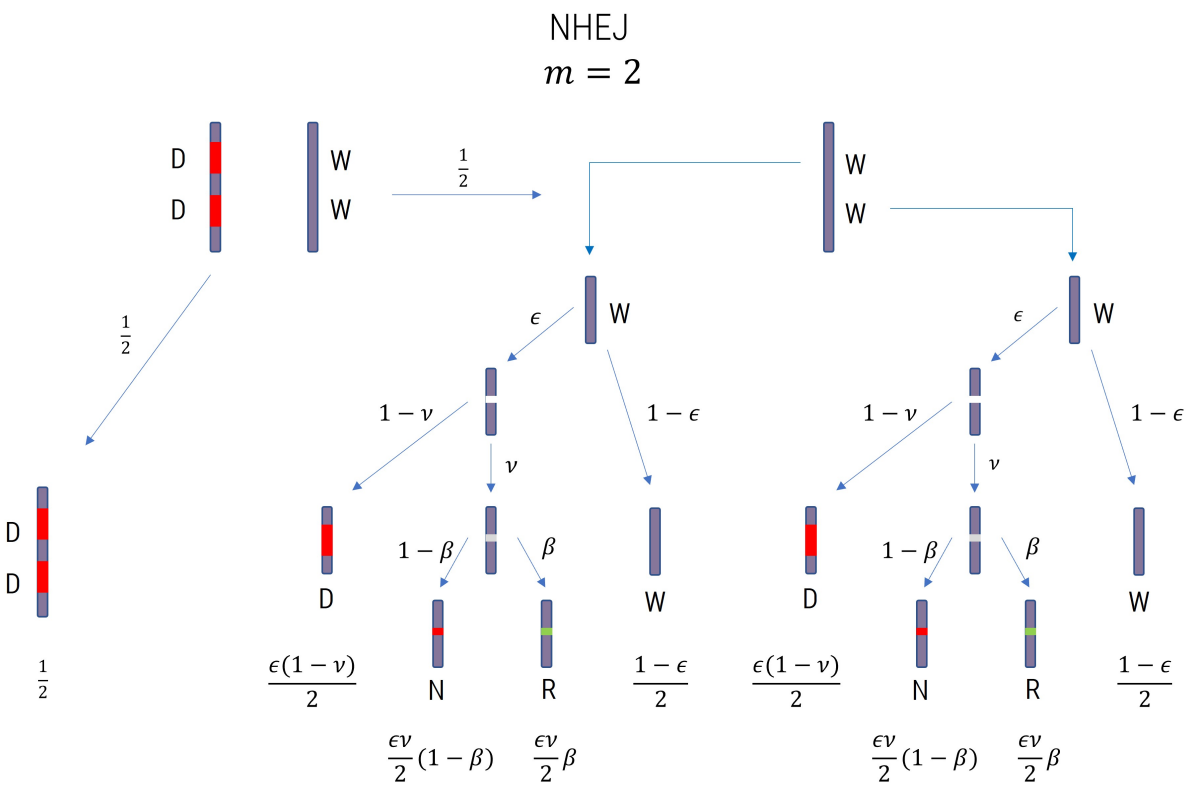
$$\boldsymbol{M}^{(2)} = \begin{pmatrix} (1-\mu)^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2\xi\mu(1-\mu) & 1-\mu & 0 & 0 & 0 & 0 & 0 \\ \xi^2\mu^2 & \xi\mu & 1 & 0 & 0 & 0 & 0 \\ 2(1-\xi)\mu(1-\mu) & 0 & 0 & 1-\mu & 0 & 0 & 0 \\ 2\xi(1-\xi)\mu^2 & (1-\xi)\mu & 0 & \xi\mu & 1 & 0 & 0 \\ (1-\xi)^2\mu^2 & 0 & 0 & (1-\xi)\mu & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{S15}$$

For $m = 3$, equivalent considerations give the following mutation probability from the jth to i^{th} haplotype

$$M^{(3)}_{ij} \equiv M^{(3)}_{k'_1 k'_2 k'_3, k_1 k_2 k_3} = \frac{n_{\mathsf{W}}!}{n'_{\mathsf{W}}! n'_{\mathsf{R}}! n'_{\mathsf{N}}!} M_{k'_1 k_1} M_{k'_2 k_2} M_{k'_3 k_3}, \tag{S16}$$

where haplotype $k_1 k_2 k_3$ maps to j and haplotype $k'_1 k'_2 k'_3$ maps to i, using Table 2 and $k_1, k_2, k_3, k'_1, k'_2, k'_3$ can take values $\{1, 2, 3\}$ corresponding to alleles $\{\mathsf{W}, \mathsf{R}, \mathsf{N}\}$, respectively.

For multiplex drive, we make the assumption that drive acts independently at each target site, as shown in Fig.S2 for $m = 2$ gRNAs, and the genotype WW/DD, where we assume the same



	W	R	N	D
W	$\frac{1}{4}(1-\epsilon)^2$	$\frac{1}{4}(1-\epsilon)\epsilon\nu\beta$	$\frac{1}{4}(1-\epsilon)\epsilon\nu(1-\beta)$	$\frac{1}{4}(1-\epsilon)\epsilon(1-\nu)$
R	$\frac{1}{4}(1-\epsilon)\epsilon\nu\beta$	$\frac{1}{4}(\epsilon\nu\beta)^2$	$\frac{1}{4}(\epsilon\nu)^2\beta(1-\beta)$	$\frac{1}{4}\epsilon^2\nu(1-\nu)\beta$
N	$\frac{1}{4}(1-\epsilon)\epsilon\nu(1-\beta)$	$\frac{1}{4}(\epsilon\nu)^2\beta(1-\beta)$	$\frac{1}{4}(\epsilon\nu(1-\beta))^2$	$\frac{1}{4}\epsilon^2\nu(1-\nu)(1-\beta)$
D	$\frac{1}{4}(1-\epsilon)\epsilon(1-\nu)$	$\frac{1}{4}\epsilon^2\nu(1-\nu)\beta$	$\frac{1}{4}\epsilon^2\nu(1-\nu)(1-\beta)$	$\frac{1}{4}\epsilon^2(1-\nu)^2$

Total fraction of DD gametes $\frac{1}{2}$ + all terms shaded in red

Figure S2: Diagram showing gametogenesis for $m = 2$ gRNAs with NHEJ for a WW/DD individual, where the model assumes the action of drive is independent on each target site. wild type is the background colour, a white marker indicates the cleavage state, a long red bar represents drive inserted at the target site, a grey marker is an NHEJ event, which are functional (green) or non-functional (red).

genotype	heterozygote frequency
WW/DD	$H_1 = x_1y_7 + x_7y_1$
WR/DD	$H_2 = x_2y_7 + x_7y_2$
WN/DD	$H_4 = x_4y_7 + x_7y_4$

Table 3: Drive heterozygotes for 2-fold multiplex and their frequencies. Time dependence is implicit.

genotype	heterozygote frequency
WWW/DDD	$H_1 = x_1y_{11} + x_{11}y_1$
WWR/DDD	$H_2 = x_2y_{11} + x_{11}y_2$
WRR/DDD	$H_3 = x_3y_{11} + x_{11}y_3$
WWN/DDD	$H_5 = x_5y_{11} + x_{11}y_5$
WRN/DDD	$H_6 = x_6y_{11} + x_{11}y_6$
WNN/DDD	$H_8 = x_8y_{11} + x_{11}y_8$

Table 4: Drive heterozygotes for 3-fold multiplex and their frequencies. Time dependence is implicit.

values of ϵ, ν, and β at each target site. The fraction of gametes is then effectively calculated by a multiplicative table of the fractions of W, R, N, D produced at each target site; since only a single successful cleavage and repair is needed (signified by a single D), we need to account for all fractions shown shaded in red in the table, to calculate the total fraction of DD gametes. This is formalised mathematically below and extended to any number of m of gNRAs.

To consider how multiplexed drive affects haplotype frequencies, we need to consider now that there is more than just a single genotype that is affected by drive. All genotypes with drive on one chromosome and at least a single W allele on the other can be converted to homozygous drive; for $m = 2$ there are three which are enumerated in Table 3 and for $m = 3$ there are six enumerated in Table 4.

Each of these heterozygotes produces different fractions of the n_H possible haplotypes, depending on the number of W alleles in each haplotype. We create an analogous matrix to $\boldsymbol{M}$ (Eqn.S8), whose j^{th} column gives the fraction of gametes produced by the jth allele assuming it is in genotype with D:

$$\boldsymbol{\kappa} = \begin{pmatrix} 1-\epsilon & 0 & 0 & 0 \\ \epsilon\nu\beta & 1 & 0 & 0 \\ \epsilon\nu(1-\beta) & 0 & 1 & 0 \\ \epsilon(1-\nu) & 0 & 0 & 1 \end{pmatrix}. \tag{S17}$$

With this matrix the fraction of gametes that have the ith haplotype from drive heterozygote j can be calculated for any value of m. For haplotypes $i < n_H$ (i.e. not including drive) this is given by

$$\boldsymbol{\kappa}_{ij}^{(m)} = \frac{n_{\mathsf{W}}!}{n'_{\mathsf{W}}!n'_{\mathsf{R}}!n'_{\mathsf{N}}!} \prod_{\ell}^{m} \kappa_{k'_\ell k_\ell} - \delta_{ij} \qquad (i < n_H), \tag{S18}$$

where n_{W} is the number of W alleles in the drive heterozygote j, and $n'_{\mathsf{W}}, n'_{\mathsf{R}}, n'_{\mathsf{N}}$ are the number of $\mathsf{W}, \mathsf{R}, \mathsf{N}$ alleles in haplotype i, and as before the haplotypes k_1k_2 and $k'_1k'_2$ map to indices using

Table 1. For $i = n_H$, the drive haplotype (i.e. DD or DDD), we need to sum over all haplotypes i that produce at least a single D allele (which is effectively converted to, for example, DD or DDD); if these set of haplotypes is denoted by $\mathcal{D}_j$ then

$$\boldsymbol{\kappa}^{(m)}_{n_H,j} = \sum_{i\in\mathcal{D}_j} \frac{n_{\mathsf{W}}!}{n'_{\mathsf{W}}!n'_{\mathsf{R}}!n'_{\mathsf{N}}!n_{\mathsf{D}}!} \prod_{\ell}^{m} \kappa_{k'_\ell k_\ell} - \delta_{n_H,j}, \tag{S19}$$

where on the RHS the relation between $i \in \mathcal{D}_j$ and $k'_\ell,\ k'_\ell,\ n'_{\mathsf{W}}, n'_{\mathsf{R}}, n'_{\mathsf{N}}$ and n'_{D} is implicit, and straightforward to enumerate.

Together these equations allow us to write down expressions for the update equations for haplotype frequencies between generations for $m-$fold multiplex

$$\begin{aligned} \boldsymbol{x}_{t+1} = \boldsymbol{M}^{(m)} & \frac{(\boldsymbol{W}_{\mathrm{f}}\boldsymbol{y}_t)\odot \boldsymbol{x}_t + (\boldsymbol{W}_{\mathrm{f}}\boldsymbol{x}_t)\odot \boldsymbol{y}_t}{2\bar{w}_{\mathrm{f}}} \\ & + (1-\mu)\sum_{i\in\mathcal{H}_m} \frac{w^{\mathrm{f}}_{j,n_H}}{2\bar{w}_{\mathrm{f}}} \boldsymbol{\kappa}^{(m)}_{:,j} H_j, \end{aligned} \tag{S20}$$

$$\begin{aligned} \boldsymbol{y}_{t+1} = \boldsymbol{M}^{(m)} & \frac{(\boldsymbol{W}_{\mathrm{m}}\boldsymbol{x}_t)\odot \boldsymbol{y}_t + (\boldsymbol{W}_{\mathrm{m}}\boldsymbol{y}_t)\odot \boldsymbol{x}_t}{2\bar{w}_{\mathrm{m}}} \\ & + (1-\mu)\sum_{i\in\mathcal{H}_m} \frac{w^{\mathrm{m}}_{j,n_H}}{2\bar{w}_{\mathrm{m}}} \boldsymbol{\kappa}^{(m)}_{:,j} H_j, \end{aligned} \tag{S21}$$

where $\mathcal{H}_m$ is the set of indices of haplotypes that have at least a single W allele, which is a function of m the degree of multiplexing (Tables 3&4), H_j are the drive-heterozygote frequencies given in Tables 3&4, where the time-dependence on t is implicit, and where the notation $\boldsymbol{\kappa}^{(m)}_{:,j}$ is the jth column of the matrix $\boldsymbol{\kappa}^{(m)}$.

Concurrent vs Sequential accumulation of resistance alleles for NHEJ vs de novo mutation

When there are multiple gRNAs, whether by NHEJ or de novo SNPs, functional resistance mutations can arise sequentially (e.g. WWW $\rightarrow$ WWR $\rightarrow$ WRR $\rightarrow$ RRR) or concurrently through m-fold mutations (e.g. WWW $\rightarrow$ RRR) across the target sites, or some combination of the two (e.g. WWW $\rightarrow$ WWR $\rightarrow$ RRR). Here, using a simple heuristic analysis, we examine based on calculating the probability of fixation when there are a certain number of copies of functional resistance haplotypes, what signature we would expect in the scaling of N^* with respect to the base rate of generation of each type of mutant, which is $\epsilon\beta\nu$ for NHEJ and $\xi\mu$ for de novo SNPs. The approach essentially deterministically calculates the number of mutants generated over a certain timescale and ignores all fitness effects, assuming drift dominates at small frequencies, and only accounts for the fitness through the probability of establishment.

Generation of functional resistance non-homologous end joining mutants (NHEJ)

For a single gRNA, per generation and per each heterozygous W/D individual, non-homologous end joining mutants arise at a rate proportional to $\epsilon\nu$, and a fraction of β of these we assume to be functional (R), and so the rate of producing functional NHEJ mutants is proportional to $\epsilon\nu\beta$. Hence, of r_t is the frequency of the R allele in the population in generation t, then the difference equation describing the dynamics is given by

$$r(t+1) = r(t) + \epsilon\nu\beta q(t)(1-q(t)), \tag{S22}$$

where q_t is the frequency of the D allele and to a good approximation $q(t)(1-q(t))$ is the frequency of the W/D heterozygotes, assuming all other alleles apart from the W and D are at much smaller frequencies initially. Mutants generated earlier will generally have contribute most to the ultimate probability of establishment and population rescue, and so we make the approximation that the frequency of the heterozygotes is roughly $q(t)$ and so

$$r(t+1) \approx r(t) + \epsilon\nu\beta q(t), \tag{S23}$$

which has solution $r(t) = r(0) + \epsilon\nu\beta Q(t) = \epsilon\nu\beta Q(t)$, since we assume at $t = 0$ when drive is introduced there are no mutants, and where $Q(t) = \sum_{t'=0}^{t-1} q(t')$. If we assume there is some characteristic time over which mutants are generated, related to the time to fixation of drive τ, then the number of copies of the mutant generated in this time is τ is therefore $2Nr(\tau) = 2N\epsilon\nu\beta Q(\tau)$. If each mutant has probability of fixation of $\pi = 2s_b$, where s_b is an effective selection coefficient of the benefit the functional resistance allele has in the presence of drive, then given this number of mutants, we calculate the probability of fixation by 1- probability that none of these mutants establishes:

$$\begin{aligned} p &= 1-(1-\pi)^{Nr(\tau)} \\ &\approx 1-e^{-\pi N r(\tau)} \\ &= 1-e^{-4Ns_b\epsilon\nu\beta Q(\tau)} \\ &= 1-e^{-N/N_n^*} \end{aligned} \tag{S24}$$

where $N_n^* = (4s_b\epsilon\nu\beta Q(\tau))^{-1}$, which gives us the result that we expect resistance to arise significantly, when $N > N_n$ or the rate of generation of NHEJ mutants is $2N\epsilon\nu\beta \sim (2s_b Q(\tau))^{-1}$, then the probability of resistance is large.

For $m = 2$ gRNAs, a single resistance mutation (WR or RW) is not sufficient to prevent the copying of drive and so only haplotypes with two functional resistance mutations (RR) have a selective advantage in the presence of drive. Assuming the frequency of all mutants is small, and so assuming the fact that single mutants are selected against can be ignored (due to drift), the difference equations describing this scenario are

$$r_1(t+1) = r_1(t) + 2(1-\epsilon)\epsilon\nu\beta q(t)(1-q(t)) \tag{S25}$$

$$r_2(t+1) = r_2(t) + \epsilon\nu\beta r_1(t)(1-q(t)) + (\epsilon\nu\beta)^2 q(t)(1-q(t)) \tag{S26}$$

where r_1 is the frequency of mutants with a single functional resistance allele at either of the two target sites and r_2 is the frequency of mutants that have a functional resistance allele at both target sites. If we make the same assumption that initially the frequency of drive will be small, such that $q(1-q) \approx q$, then the solution for the frequency of single-fold functional resistance mutants is $r_1(t) = 2(1-\epsilon)\epsilon\nu\beta Q(t)$, and so we have a difference equation for only r_2:

$$r_2(t+1) = r_2(t) + 2(1-\epsilon)(\epsilon\nu\beta)^2 Q(t) + (\epsilon\nu\beta)^2 q(t). \tag{S27}$$

The second term represents the generation of two-fold mutants from single-fold heterozygotes with drive (WR/DD or RW/DD), whilst the third term represents generation directly from wild type-drive heterozygotes. We can see that if $1-\epsilon$ is small then later direct mechanism of generation will be dominant, as long as the cumulative frequency of drive $Q(t)$ is not significantly larger than the current frequency $q(t)$; as $Q(t)$ is a cummulative frequency it will be of order $1/s_D$, where $s_D = (1+\epsilon)(1-hs/2) - 1$ is the effective selection coefficient of drive, which gives the condition $s_D \gg 2(1-\epsilon)$ such that direct generation of double mutants dominates over sequential generation. For typical parameters of suppression drive as used in this paper $\epsilon = 0.95$, $s = 1$ and $h = 0.3$, we get $s_D \approx 0.66$ and $2(1-\epsilon) = 0.1$, and so we would expect direct generation of double mutants to dominate. If this is the case then our difference equation for double functional mutants is

$$r_2(t+1) \approx (\epsilon\nu\beta)^2 q(t). \tag{S28}$$

and so comparing to Eqn.S22, we have the same difference equation but $\epsilon\nu\beta \to (\epsilon\nu\beta)^2$ and so would expect the probability of resistance to be of the form $p = 1 - e^{-N/N_n^*}$ with $N_n^* = (4s_b(\epsilon\nu\beta)^2 Q(\tau))^{-1}$. A similar analysis can be performed for $m > 2$ gRNAs to give in general $N_n^* = (4s_b(\epsilon\nu\beta)^m Q(\tau))^{-1}$, for m gRNAs. Here, we see that if we let $\gamma_n = s_b Q(\tau)$ then we have a single unknown constant, which is the same for all m, and fitting to the simulation data of p vs N for different values of β (Fig.2 in main text), we find indeed this is the case and $\gamma \approx 0.2$ for $m = \{1, 2, 3\}$, as shown in Fig.S3.

De novo generation of SNP functional resistant mutants

In the case of de novo generation of mutants the difference equation for a single gRNA is

$$r(t+1) = r(t) + \xi\mu(1 - q(t)) \approx r(t) + \xi\mu, \tag{S29}$$

where we assume that initially $q \ll 1$. The solution is simply $r(t) = \xi\mu t$. Using the same simple heuristic as for NHEJ, this then leads to the probability of fixation of mutants generated up to some time τ, related to the timescale for fixation of drive, to be:

$$\begin{aligned} p &= 1 - (1-\pi)^{Nr(\tau)} \\ &\approx 1 - e^{-\pi N r(\tau)} \\ &= 1 - e^{-4Ns_b\xi\mu\tau} \\ &= 1 - e^{-N/N_d^*} \end{aligned} \tag{S30}$$

with $N_d^* = (4s_b\xi\mu\tau)^{-1}$.

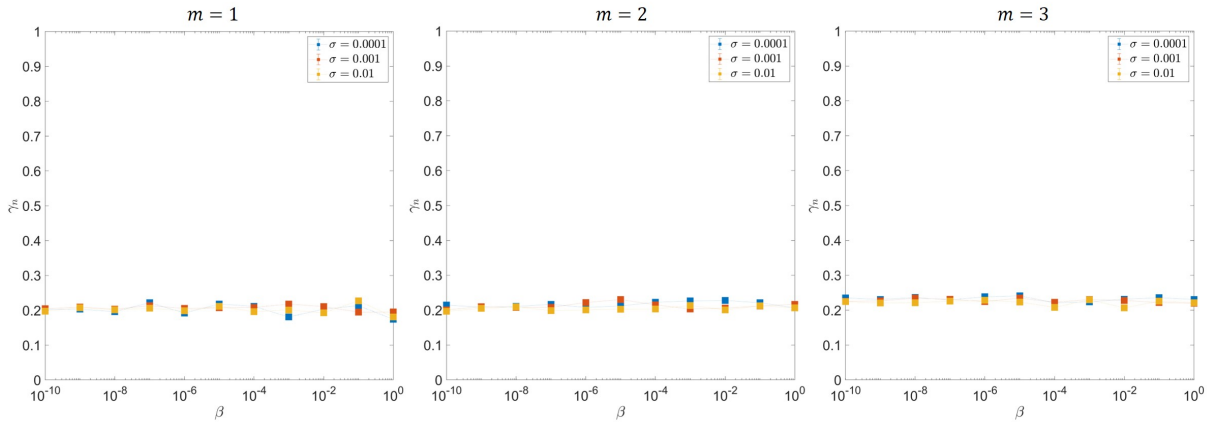


Figure S3: Values of γ_n fitting parameter from fits to simulation results of probability of resistance p vs N for different values of β, where $p = 1 - e^{-N/N_n^*}$ with $N_n^* = (4\gamma_n(\epsilon\nu\beta)^m)^{-1}$, where each panel are for different numbers of gRNAs m as indicated.

For $m = 2$ gRNAs, the difference equations for de novo generation of single and double functional resistance mutants is

$$r_1(t+1) = r_1(t) + 2\xi\mu(1 - q(t)) \approx r_1(t) + \xi\mu \tag{S31}$$

$$r_2(t+1) = r_2(t) + \xi\mu r_1(t) + (\xi\mu)^2(1 - q(t)) \approx r_2(t) + \xi\mu r_1(t) + (\xi\mu)^2 \tag{S32}$$

where again we assume the frequency of drive is initially small and we ignore any fitness effects for mutants at small frequency in calculating the frequency of mutants. As above the frequency of single mutants will be $r_1(t) \approx 2\xi\mu t$, where now as there are two paths from the wild type to generate a single mutant there is a factor of 2, and so the difference equation for double functional mutants will be

$$r_2(t+1) \approx r_2(t) + 2(\xi\mu)^2 t + (\xi\mu)^2. \tag{S33}$$

Now here the second term, representing the sequential generation of mutants, will dominate for any sufficiently number of generations $t > 1$ and so ignoring the third term which represents concurrent generation of mutants, the solution is $r_2(t) = 2(\xi\mu)^2 t(t+1)/2 \sim (\xi\mu)^2 t^2$ for large t. This means on a time scale τ the number of functional double-mutants generated will be $2Nr_2(\tau) \sim 2N \times (\xi\mu)^2 t^2 = 2N(\xi\mu)^2 t^2$. Following the previous arguments, the probability of fixation will then be $p \approx 1 - e^{-N/N_d^*}$ with $N_d = (4(\xi\mu)^2 s_b \tau^2)^{-1}$. Again the arguments can be extended to $m = 3$ gRNAs and we find $N_d = (4(\xi\mu)^2 s_b \tau^3)^{-1}$, to leading order in τ.

We can fit the simulation data of the de novo probability of resistance vs N as ξ is varied, using this heuristic theory with a single fitting constant γ_m and $N_d^* = (4(\xi\mu)^m \gamma_m)^{-1}$, for each value of m and we find $\gamma_1 = 0.76 \pm 0.004$, $\gamma_2 = 6 \pm 0.04$ and $\gamma_3 = 55 \pm 0.3$, as shown in Fig.S4; using the relation $\gamma_m = s_b \tau^m$, we find these data are consistent with $\tau \approx 8.5$ generations and a beneficial selection coefficient of $s_b \approx 0.09$.

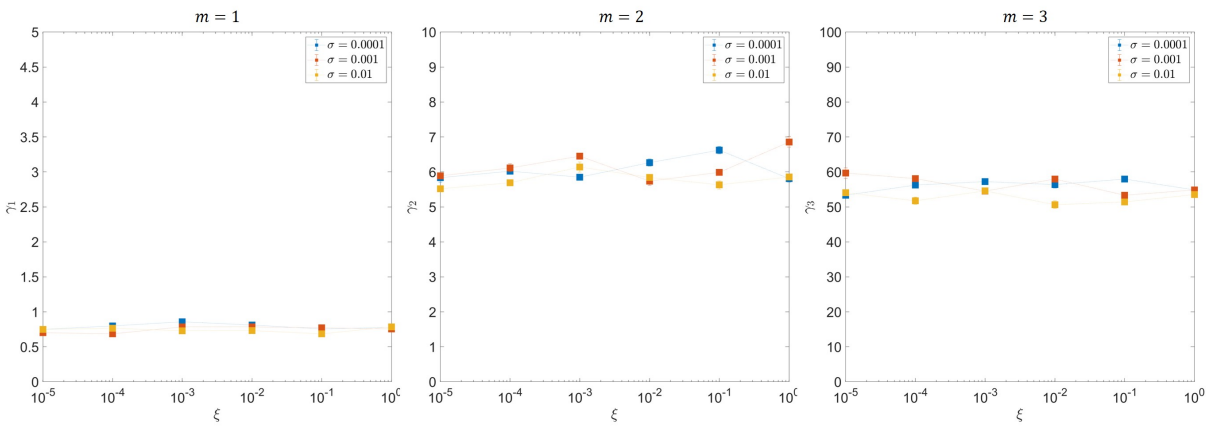


Figure S4: Values of γ_m fitting parameter from fits to simulation results of probability of resistance p vs N for different values of ξ, where $p = 1 - e^{-N/N_d^*}$ with $N_d^* = (4\gamma_m(\xi\mu)^m)^{-1}$, where each panel are for different numbers of gRNAs m as indicated.

Multiplex resistance from standing variation of SNP mutants

In this section, we show that the significant amplification of standing variation or pre-existing mutants for $m > 1$ gRNAs can be heuristically understood by calculating approximately the average probability of fixation of m-fold functionally resistant, assuming that in the presence of drive it has a beneficial selection coefficient s_b and that before the introduction of drive it is mildly selected against, giving a distribution of the frequency of such mutants in mutation-selection balance.

For $m = 1$ gRNAs the relevant theory has already been calculated in detail by Hermisson & Pennings (*S15*), which we reproduce here in brevity and slightly simplified form. In mutation-selection balance, the diffusion approximation can be used to calculate the equilibrium distribution of the frequency of such variants. Here a single functional resistance mutant has selection coefficient $-\sigma$ before the introduction of drive and the distribution of frequency r is approximately gamma distributed, as long as $\alpha = 4N\sigma \gg 1$:

$$\rho(r) = \frac{\alpha^\theta}{\Gamma(\theta)} r^{\theta-1} e^{-\alpha r} \tag{S34}$$

which has shape parameter $\theta = 4N\xi\mu$ and rate α. The mean of this distribution is $\langle r \rangle = \theta/\alpha = \xi\mu/\sigma$, which is the classic mutation-selection balance frequency. Now the probability of fixation given an initial frequency r_0 and beneficial selection coefficient s_b is

$$\pi(r_0) = \frac{1 - e^{-\alpha_b r_0}}{1 - e^{-\alpha_b}} \approx 1 - e^{-\alpha_b r_0}, \tag{S35}$$

where the last approximation assumes $\alpha_b = 4Ns_b \gg 1$. The average probability of fixation assuming a distribution $\rho(r_0)$ is then simply calculated by averaging over the initial frequency r_0, which gives

$$\langle \pi \rangle = 1 - \frac{Z(\alpha_b + \alpha, \theta)}{Z(\alpha, \theta)}, \tag{S36}$$

where Z is the normalisation constant of the gamma distribution with population scaled fitness f: $Z(f,\theta) = \int_0^\infty r^{\theta-1}e^{-fr}\mathrm{d}r$, where extending the integral to ∞ will have small penalty in the strong selection limit, i.e. as long as $f \gg 1$. Evaluating this and re-arranging we find the result of Hermisson & Pennings in the strong selection limit:

$$\langle\pi\rangle = 1 - e^{-4N\xi\mu\ln\left(1+\frac{s_b}{\sigma}\right)} \tag{S37}$$

And so the probability of resistance $p = 1 - e^{-N/N_s^*}$, where $N_s^* = \left(4\xi\mu\ln\left(1+\frac{s_b}{\sigma}\right)\right)^{-1}$, which is of the same form as for NHEJ and de novo SNP mutations. There is only a single unknown parameter s_b, and so fitting this form to the simulation data for the p vs N, for varying ξ we find $s_b \approx 0.5$ as shown in Fig. ($m = 1$ panel), which is larger than found from de novo simulations, and likely reflects some mechanistic aspect not reflected in these simple heuristic considerations.

For $m = 2$, we could try to calculate exactly the joint steady-state distribution of the frequency of all mutants, under the sequential mutation scheme WW $\rightarrow$ {WR, RW} $\rightarrow$ RR, but this is in general an unsolved problem in population genetics, since there is a net flux or current of probability through the states of the system (detailed balance is not obeyed) and so calculating this joint distribution is not straightforward. Here, instead we appeal to a simple heuristic, which reproduces the scaling with σ seen in the simulation data; we assume the distribution of the double-mutant frequency r_2 follows the same gamma distribution, but that mutation to this allele is solely from single mutants whose frequency is assumed fixed at it's equilibrium frequency $\langle r_1\rangle = 2\xi\mu/\sigma$, where the factor 2 arises from the multiplicity 2 of paths to single mutants from the wild type; this is reasonable given that sequential mutation rather than concurrent is dominant, as demonstrated for de novo mutation. Given this and that the fitness cost of double mutants, before drive is introduced, is -2σ, the distribution of the double functional resistance mutants is

$$\rho_2(r_2) = \frac{{\alpha_2}^{\theta_2}}{\Gamma(\theta_2)} r^{\theta_2-1}e^{-\alpha_2 r}, \tag{S38}$$

where $\alpha_2 = 8N\sigma$, and $\theta_2 = 4N\xi\mu\langle r_1\rangle = 8N(\xi\mu)^2/\sigma$. The average frequency of double functional mutants is $\theta_2/\alpha_2 = (\xi\mu)^2/\sigma^2$, which is simple to show is what we would expect solving the deterministic equations for the frequency of double-mutants. Using the form of Eqn.S36, we find

$$\langle\pi\rangle = 1 - e^{-\frac{8N(\xi\mu)^2}{\sigma}\ln\left(1+\frac{s_b}{2\sigma}\right)}, \tag{S39}$$

which is of the form $1 - e^{-N/N_s^*}$, with $N_s^* = (\frac{8(\xi\mu)^2}{\sigma}\ln\left(1+\frac{s_b}{2\sigma}\right))^{-1}$, which we see means $N_s^* \sim \sigma$ for $m = 2$, instead of the much weaker logarithmic dependence for $m = 1$. A similar analysis can be done for $m = 3$, which gives $N_s^* \sim \left(\frac{(\xi\mu)^3}{\sigma^2}\ln\left(1+\frac{s_b}{3\sigma_b}\right)\right)^{-1} \sim \sigma^2$, which is consistent with the power law scaling we see in the simulations (Fig.4b). In practice, we find that the following form $N_s^* = (4\xi\mu m\left(\frac{\gamma_s\xi\mu}{\sigma}\right)^{m-1}\ln\left(1+\frac{s_b}{m\sigma_b}\right))^{-1}$ fits the simulation data of p vs N for different values of ξ and m with a single fitting parameter $\gamma_s \approx 2$ (for $m = 2$ and $m = 3$) as shown in Fig.S5.

Note that there is an implicit assumption in all these calculations, that establishment must arise before the population is removed, which will be true if $\sim 1/s_b$ is much smaller than the timescale over which the population is removed.

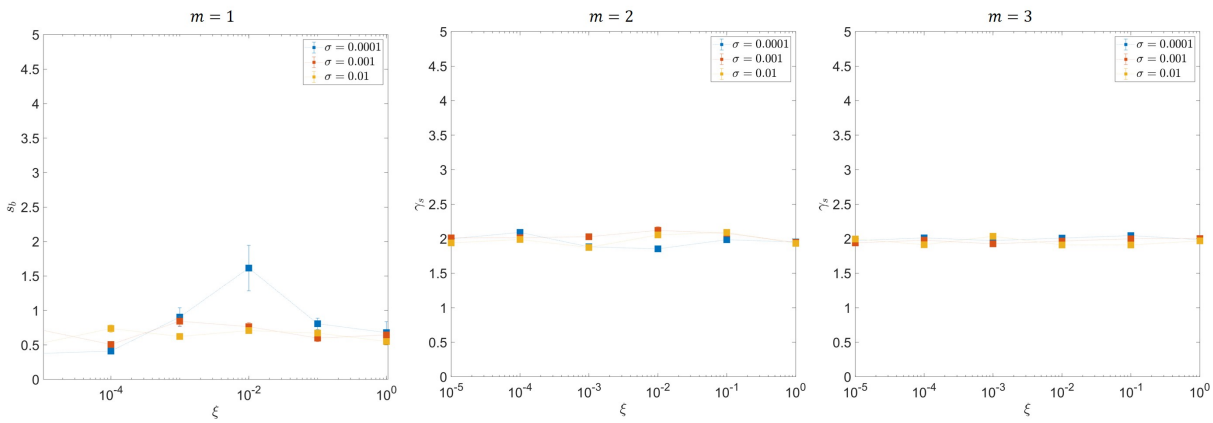


Figure S5: Values of s_b ($m = 1$) and γ_s ($m = 2$ and $m = 3$) fitting parameter from fits to simulation results of probability of resistance p vs N for different values of ξ with standing variation and de novo SNPs and NHEJ turned off, where $p = 1 - e^{-N/N_s^*}$ with $N_s^* = (4\xi\mu m \left(\frac{\gamma_s \xi \mu}{\sigma}\right)^{m-1} \ln\left(1 + \frac{s_b}{m\sigma_b}\right))^{-1}$, where each panel are for different numbers of gRNAs m as indicated.

Time to Resistance

NHEJ only

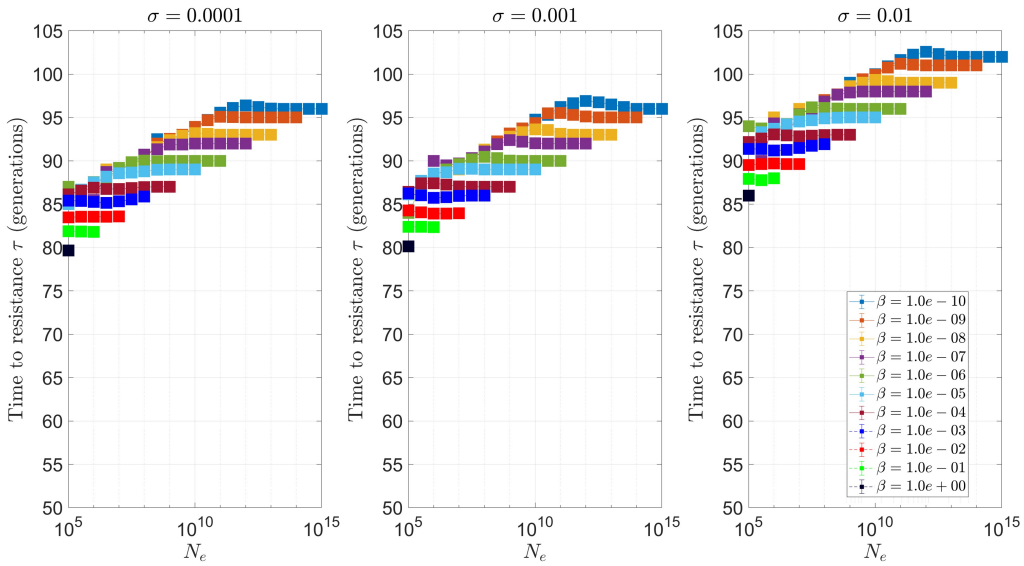


Figure S6: Mean time to resistance over $M = 500$ replicate simulations for $m = 1$ gRNA. Standard parameters of simulations.

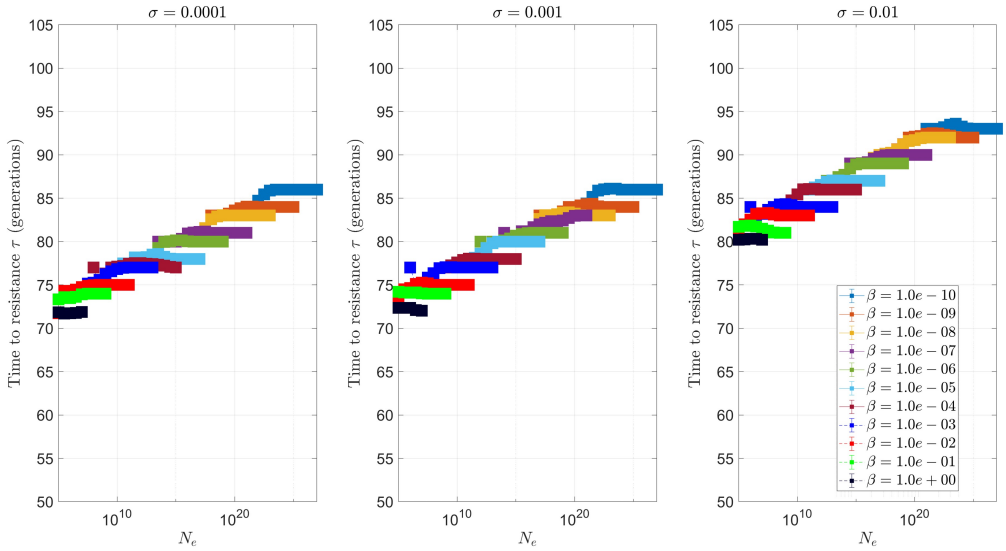


Figure S7: Mean time to resistance over $M = 500$ replicate simulations for $m = 2$ gRNAs. Standard parameters of simulations.

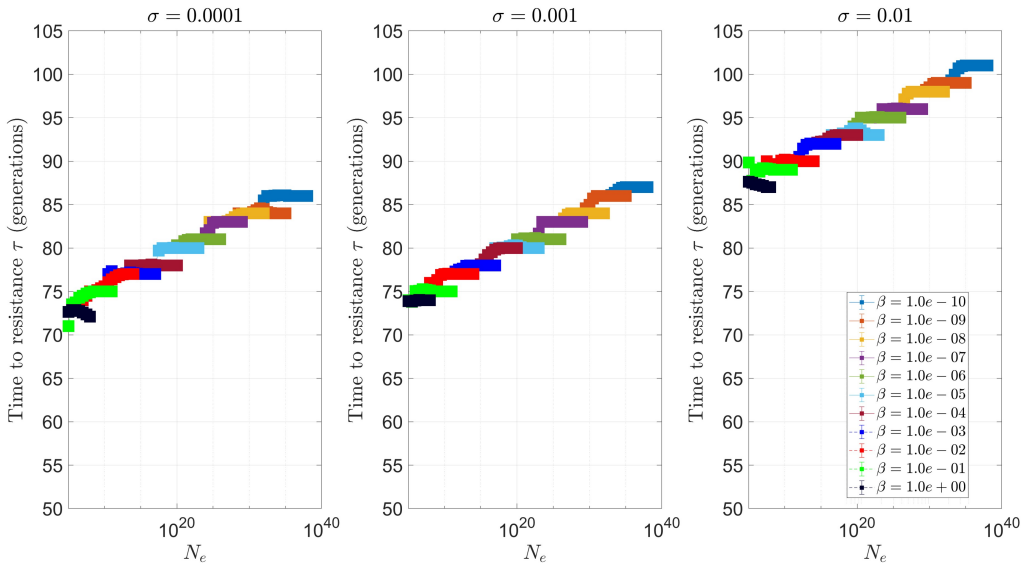


Figure S8: Mean time to resistance over $M = 500$ replicate simulations for $m = 3$ gRNAs. Standard parameters of simulations.

Single nucleotide mutations only (de novo)

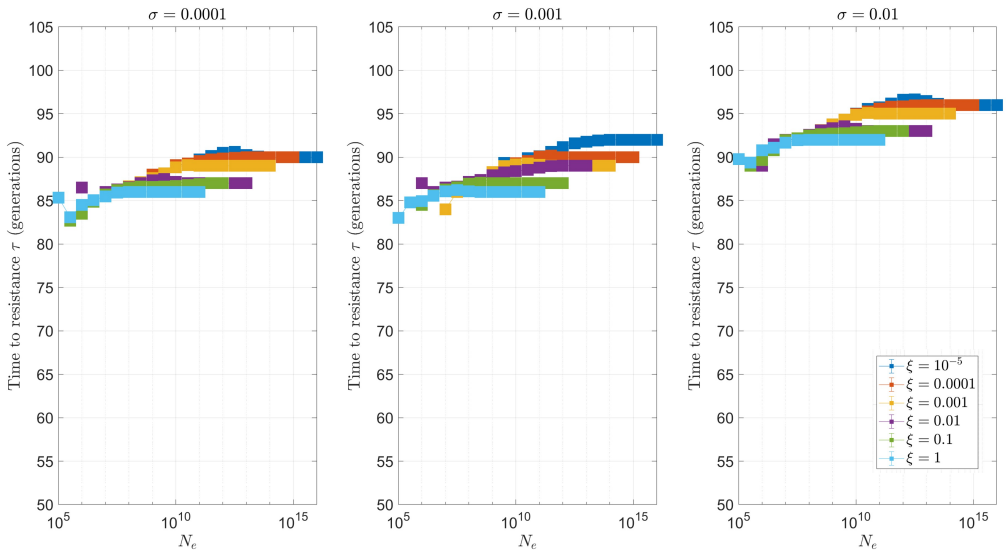


Figure S9: Mean time to resistance over $M = 500$ replicate simulations for $m = 1$ gRNA. Standard parameters of simulations.

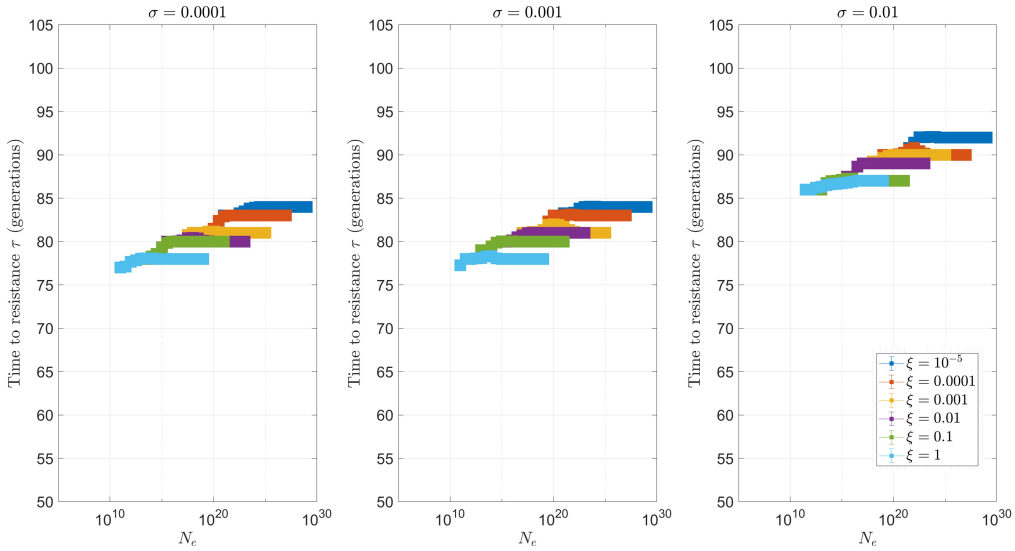


Figure S10: Mean time to resistance over $M = 500$ replicate simulations for $m = 2$ gRNAs. Standard parameters of simulations.

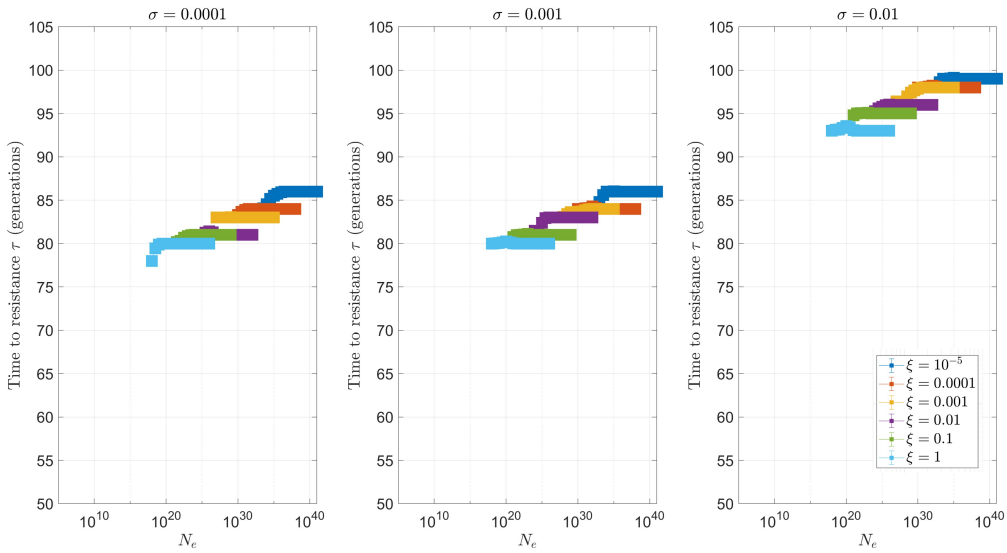


Figure S11: Mean time to resistance over $M = 500$ replicate simulations for $m = 3$ gRNAs. Standard parameters of simulations.

Single nucleotide mutations only (pre-existing + de novo)

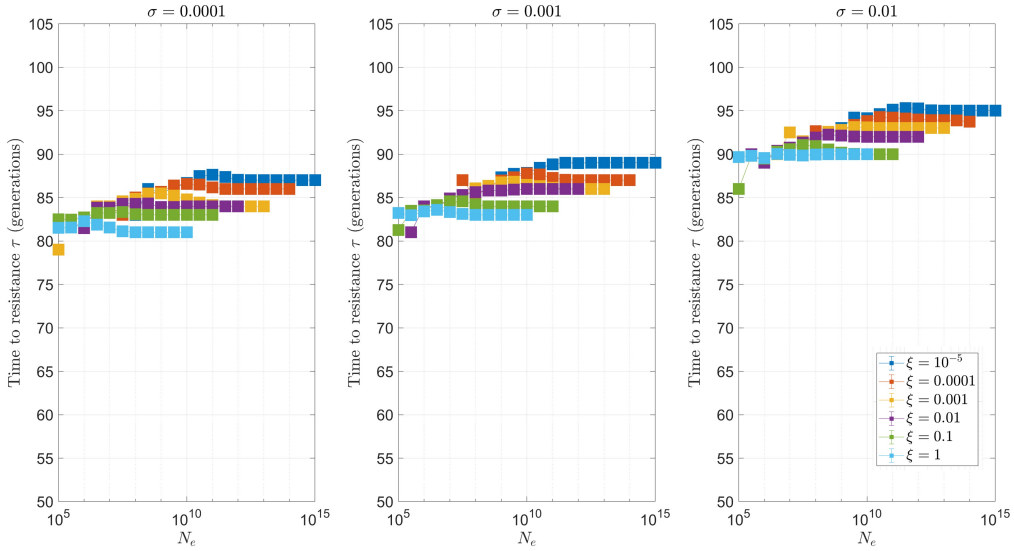


Figure S12: Mean time to resistance over $M = 500$ replicate simulations for $m = 1$ gRNA. Standard parameters of simulations.

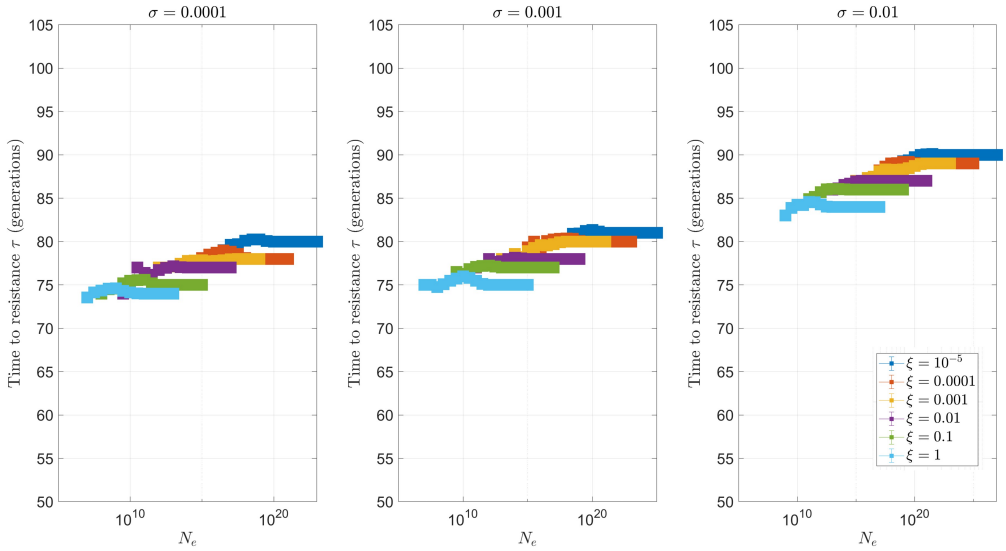


Figure S13: Mean time to resistance over $M = 500$ replicate simulations for $m = 2$ gRNAs. Standard parameters of simulations.

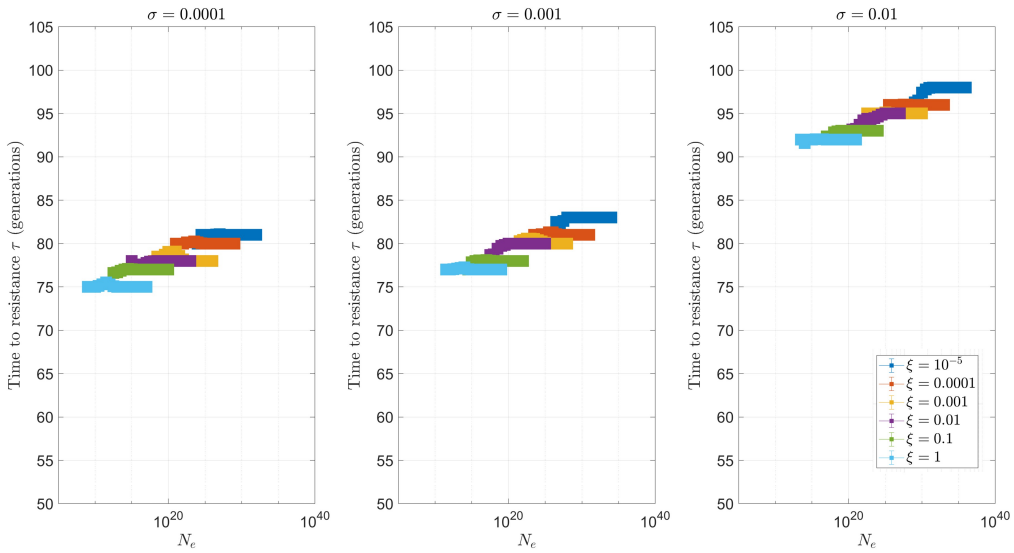


Figure S14: Mean time to resistance over $M = 500$ replicate simulations for $m = 3$ gRNAs. Standard parameters of simulations.