Supporting Information

Bozic and Nowak 10.1073/pnas.1412075111

SI Text

The Model. We model the growth of a metastatic lesion as a branching process (1) that starts from a single cell sensitive to treatment. Sensitive cells divide with rate b and die with rate d. The net growth rate of sensitive cells is $r=b-d$. During division, one of the daughter cells receives a resistance mutation with probability u. Resistant cells have birth and death rates b_R and d_R. Resistant mutations can be neutral in the absence of treatment, which means they have the same birth and death rates as sensitive cells, and we initially focus on this case. Alternatively, if $c=(b_R-d_R)/(b-d)>1$, then resistance mutations are advantageous before treatment; if $c<1$, they are deleterious. We assume that mutation rate u is small, final lesion size M is large, $Mu \gg 1$, and we are mostly interested in the behavior of the early surviving clones. Furthermore, since mutation rate u is small, we assume that the size of the resistant population is much smaller than the size of the sensitive population, and approximate the size of the sensitive population with the size of the lesion.

Rate of Production of Mutants. We use the result (2) that the collection of tumor sizes at which resistance mutations are produced can be viewed (approximated) as a homogeneous Poisson process on $[1,M]$ with intensity $u/(1-d/b)$. The reasoning follows from the fact that the average total number of resistance mutations produced when there are exactly x sensitive cells in the population is given by

$$R_x = \frac{bux}{1-d/b}\int_0^\infty f_x(t)dt, \qquad \text{[S1]}$$

where $f_x(t)$ is the probability that there are exactly x sensitive cells after time t and the factor $1-d/b$ comes from only looking at lineages in which the tumor population did not go extinct. In the small mutation rate u limit, one can neglect the production of mutants to calculate $f_x(t)$ as pertaining to a single type branching process on the sensitive cells. Even though it seems that Iwasa et al. (2) were not aware of it, $f_x(t)$ was derived by Bailey (1)

$$f_x(t) = (1-\alpha)(1-\beta)\beta^{x-1}, \qquad \text{[S2]}$$

with $\alpha=(de^{rt}-d)/(be^{rt}-d)$ and $\beta=(be^{rt}-b)/(be^{rt}-d)$. Plugging in the expression for f_x into Eq. **1** leads to the rate at which mutants are produced when there are x sensitive cells

$$R_x = u/(1-d/b). \qquad \text{[S3]}$$

A more intuitive way to prove this result is as follows: when the population contains exactly x sensitive cells, the probability that they will produce a mutant before going to $x-1$ or $x+1$ sensitive cells is $bu/(b+d)$. The (average) number of occurrences of exactly x cells in the process is one plus the (average) number of returns of a biased random walk with $p=b/(b+d)$. Multiplying the probability of producing a mutant cell while at state x with the number of occurrences of that state leads to $R_x=u/(1-d/b)$.

Each mutant cell survives stochastic drift with probability $1-d/b$, so the tumor sizes at which mutations that survive stochastic drift are produced can be viewed as a Poisson process on $[1,M]$ with intensity u. Because M is large and u is small, we can replace the interval $[1,M]$ by $[0,M]$ without losing much accuracy (3).

Size of the *k*th Resistant Clone. Let M_k be the size of the sensitive population when the kth successful resistant mutant appears. Furthermore, let Y_k denote the size of the kth resistant subpopulation that survives stochastic drift when there are M-sensitive cells, conditioned on $M_k \le M$. By the time the sensitive population reached size M, Y_k can be approximated by MV/M_k, where V is an exponentially distributed random variable with mean $b/(b-d)$ (4). If $F_k(y)=\Pr[Y_k \le y]$ is the cumulative distribution of Y_k, then, expanding on the reasoning in ref. 3, we have

$$\begin{aligned} F_k(y) &\approx 1-\Pr[MV/M_k \ge y | M_k \le M] \\ &= 1-\int_0^M \text{Prob.Density}[M_k=z|M_k\le M]\times \Pr\left[V\ge \frac{yz}{M}\right]dz \\ &= 1-\int_0^M \frac{(zu)^{k-1}e^{-zu}u}{(k-1)!}\left(1-\sum_{l=0}^{k-1}\frac{(Mu)^l e^{-Mu}}{l!}\right)^{-1}\exp\left(-\frac{ryz}{Mb}\right)dz. \end{aligned}$$

Evaluating the integral above leads to

$$F_k(y) \approx 1-\left(\frac{Mu}{Mu+y-dy/b}\right)^k \frac{\Gamma(k)-\Gamma(k,Mu+y-dy/b)}{\Gamma(k)-\Gamma(k,Mu)}, \qquad \text{[S4]}$$

where $\Gamma(k)=(k-1)!$ and $\Gamma(a,z)=\int_z^\infty t^{a-1}e^{-t}dt$ is the incomplete Gamma function.

The probability density function for Y_k, $f_k=F_k'$, is given by

$$f_k(y) \approx rk(bMu)^k(bMu+ry)^{-1-k}\frac{\Gamma(k)-\Gamma(k,Mu+y-dy/b)}{\Gamma(k)-\Gamma(k,Mu)}. \qquad \text{[S5]}$$

In particular, for $k \ll Mu$, we have

$$F_k(y) \approx 1-\left(\frac{Mu}{Mu+y-dy/b}\right)^k, \qquad \text{[S6]}$$

and

$$f_k(y) \approx k(1-d/b)(Mu)^k(Mu+y-dy/b)^{-1-k}. \qquad \text{[S7]}$$

Comparison of Formula **S6** and the cumulative distribution function for Y_k obtained from 5,000 runs of the exact computer simulation of the branching process is shown in Fig. 1*B*.

Calculating the expected number of cells in the kth clone using Formula **S7** (integrating from 0 to M) and expanding it in the small u limit, we obtain

$$\mathrm{E}(Y_k) \approx \frac{bMu}{r(k-1)}+O(u^2), \qquad \text{[S8]}$$

for $k \ge 2$ and

$$\mathrm{E}(Y_1) \approx \frac{bMu}{r}\left(\log\frac{r}{bu}-1\right)+O(u^2). \qquad \text{[S9]}$$

We can also obtain the median for the number of cells in the kth clone, $Y_k^{1/2}$, from the cumulative distribution function (**S6**)

$$Y_k^{1/2} = \frac{bMu}{r}\left(2^{1/k}-1\right). \qquad \text{[S10]}$$

Ratio of Resistant Clone Sizes. To more precisely determine the relationship between the sizes of different subclones, we next

calculate the probability distribution of Y_1/Y_k: the ratio of sizes of first and kth clone. We again use the fact that the random variable describing the size of the kth clone, Y_k, can be approximated by $V_k M/M_k$, where $V_k \sim \mathrm{Exp}[b/(b-d)]$, and that M_k, size of the population on the arrival of the kth clone, is the sum of k exponential random variables with mean $1/u$. We have

$$\Pr[Y_1/Y_k \le x] = \Pr[V_1 M/M_1 \le x V_k M/M_k] = \Pr[V_1/V_k \le x M_1/M_k].$$

We note that $Z = V_1/V_k$ is the ratio of two independent, identically distributed exponential random variables and thus its probability density function is $f_Z(z) = 1/(z+1)^2$. Similarly, $W = M_1/M_k \sim \Gamma[1,\lambda]/(\Gamma[1,\lambda]+\Gamma[k-1,\lambda]) \sim \beta[1,k-1]$ is a β-distributed random variable with probability density function $f_W(w) = (k-1)(1-w)^{k-2}$. It follows that

$$\begin{aligned}\Pr[Y_1/Y_k \le x] &= \Pr[Z \le xW] \\ &= \int_0^1 (k-1)(1-w)^{k-2} \int_0^{wx} \frac{1}{(z+1)^2} dz dw \\ &= \frac{x}{1+x}\left(1 - \frac{k-1}{k} {}_2F_1[1,1,1+k,-x]\right),\end{aligned} \qquad \textbf{[S11]}$$

where ${}_2F_1$ is the hypergeometric function. Notably, this distribution depends only on k and not on any parameters of the process.

For example, the probability that the ratio of sizes of the first and the second clone is smaller than x is

$$\Pr[Y_1/Y_2 \le x] = 1 - \frac{\log(1+x)}{x}. \qquad \textbf{[S12]}$$

In particular, the first successful subclone that appears is smaller than the second appearing subclone in 31% of cases. The probability that the first appearing subclone is twice as large or larger than the second appearing clone is 55%, and the probability that it is 10 or more times larger is 24%.

Nonneutral Resistance. In this section, we will obtain similar results for the probability distributions of clone sizes and their ratios in the case in which resistant cells have birth and death rates b_R and d_R, respectively. Tumor sizes at which resistance mutations appear can still be viewed as a homogeneous Poisson process on $[0,M]$ with intensity $u/(1-d/b)$. However, the lineages of newly produced resistance mutations will escape extinction with probability $1-d_R/b_R$, so the successful resistant subclones in this scenario will arrive with rate $u_R = u(1-d_R/b_R)/(1-d/b)$. Another change from the neutral case is that the size of the kth clone when the total population size is M can be approximated with $Y_k \sim (M/M_k)^c U$, where M_k is the population size when the kth successful resistance mutation appeared and U is an exponentially distributed random variable with mean $b_R/(b_R-d_R)$. We recall that $c = (b_R-d_R)/(b-d)$.

With the above caveats, we can write the derivation of the cumulative distribution function for the size of the kth appearing resistant subclone, Y_k, similarly as before

$$\begin{aligned}F_k(y) &\approx 1 - \Pr[(M/M_k)^c U \ge y | M_k \le M] \\ &= 1 - \int_0^M \mathrm{Prob.Density}[M_k = z | M_k \le M] \times \Pr\left[U \ge \frac{yz^c}{M^c}\right] dz \\ &= 1 - \left(1 - \sum_{l=0}^{k-1} \frac{(Mu_R)^l e^{-Mu_R}}{l!}\right)^{-1} \int_0^M \frac{u_R (z u_R)^{k-1} e^{-z u_R}}{(k-1)!} \\ &\quad \times \exp\left(-\frac{r_R y z^c}{b_R M^c}\right) dz.\end{aligned} \qquad \textbf{[S13]}$$

The difference is that in the nonneutral case the integral has to be evaluated numerically.

For the ratio of clone sizes Y_k/Y_1 in the nonneutral case, we have

$$\begin{aligned}\Pr[Y_1/Y_k \le x] &= \Pr\left[U_1 (M/M_1)^c \le x U_k (M/M_k)^c\right] \\ &= \Pr\left[U_1/U_k \le x (M_1/M_k)^c\right] \\ &= \int_0^1 (k-1)\left(1-w^{1/c}\right)^{k-2} \frac{1}{c} w^{1/c-1} \int_0^{wx} \frac{1}{(z+1)^2} dz\, dw \\ &= \int_0^1 \frac{k-1}{c}\left(1-w^{1/c}\right)^{k-2} \frac{x w^{1/c}}{1+wx} dw.\end{aligned} \qquad \textbf{[S14]}$$

We see that even when resistance is not neutral, the ratio of clone sizes depends only on the order of appearance and the relative fitness c and not on M, u, and the specific birth and death rates of cells.

Our formulas rely on the approximation $Y \sim \exp[(b_R-d_R)t]U$ for the size of a resistant clone Y, where U is an exponentially distributed random variable with mean $b_R/(b_R-d_R)$ and time t is measured from the appearance of the founder cell of the clone. This approximation assumes large t and loses accuracy for t close to 0 [e.g., it predicts that the average size of a clone at time 0 is $b_R/(b_R-d_R)$ rather than 1]. Thus, our formulas lose accuracy when successful resistant clones are produced shortly before reaching size M. Successful resistant clones appear as a Poisson process on the number of sensitive cells with rate $u_R = u(1-d_R/b_R)/(1-d/b)$ when resistance is not neutral, and the first such clone will appear when there are $\sim 1/u_R$ sensitive cells. For our approximation to hold, M must be significantly larger than $1/u_R$, which is equivalent to $Muc(b/b_R) \gg 1$. In other words, we assume that the relative fitness c is such that the expected number of successful resistant clones, Mu_R, is much larger than 1.

1. Bailey NTJ (1964) *The Elements of Stochastic Processes with Applications to the Natural Sciences* (Wiley, New York).
2. Iwasa Y, Nowak MA, Michor F (2006) Evolution of resistance during clonal expansion. *Genetics* 172(4):2557–2566.
3. Diaz LA, Jr, et al. (2012) The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* 486(7404):537–540.
4. Durrett R, Moseley S (2010) Evolution of resistance and progression to disease during clonal expansion of cancer. *Theor Popul Biol* 77(1):42–48.