Supporting Information

Fischer et al. 10.1073/pnas.1409403112

SI Text

Bang-Bang Control of Wright–Fisher Evolution

Consider the Wright–Fisher model of evolution of a biallelic population of finite size $N \gg 1$ with (symmetric) mutation rate $\mu = N\mu_0 \geq 0$ and (intrinsic) selection coefficient $\sigma = Ns > 0$. In the diffusion approximation (1, 2), the probability distribution $P(x,t)$ for the allele frequency $x \equiv n_A/N$ of the A allele changes in time according to the following Fokker–Planck equation (3):

$$\partial_t P(x,t) = \left[-\partial_x(\sigma\, x(1-x) + \mu(1-2x)) + \frac{1}{2}\partial_x^2 x(1-x)\right] P(x,t), \quad \text{[S1]}$$

with boundary condition $P(x,0) = \delta(x - x_0)$. In the limit $\mu \to 0$, consider now the control task of maintaining an initial polymorphism $0 < x_0 < 1$ for as long as possible by linearly changing the selection coefficient instantaneously in response to and as a function of x_t:

$$\sigma \to \sigma + u(x_t), \quad u \in [-u_c, 0],\ u_c > \sigma. \quad \text{[S2]}$$

The optimal control strategy $\overline{u}(x)$ can also be expected to maximize the average survival time of the polymorphism,

$$\overline{u} = \underset{u}{\operatorname{argmax}} \left\langle T^{(0,1)} \right\rangle_{x_0}, \quad \text{[S3]}$$

where $\langle \ldots \rangle_{x_0}$ is the average over trajectories starting at x_0 and the maximization is over all functions $u : [0,1] \to [-u_c, 0],\ x \mapsto u(x)$. In this setting, using control itself does not incur a cost and does not enter the maximization objective. It can be shown (4) that maximization is then to be carried out in the smaller function space $u : [0,1] \to \{-u_c, 0\}$, where only the extremal control values are used. This type of control is called bang-bang. It is also evident that the optimal control strategy will be a piecewise constant function with a single step at a threshold $x_c \in (0,1)$,

$$\overline{u}(x) \equiv \begin{cases} 0, & x < x_c \\ -u_c, & x \geq x_c \end{cases}. \quad \text{[S4]}$$

If both the intrinsic selection coefficient σ and the control strength u_c are given, then we need to optimize only a single parameter, the threshold x_c.

Hamilton–Jacobi–Bellman Equation for Wright–Fisher Evolution. In the diffusion approximation, the above Fokker–Planck evolution equation for the Wright–Fisher model follows from the discrete time Master equation by a controlled expansion in the population size N (1). In the very same way, the cost-to-go recurrence relation in Eq. **2** leads to the Hamilton–Jacobi–Bellman equation **[3]**, as we will show in the following. In the cost-to-go equation **[2]**

$$J(n_t, t) = \min_{u_t} \sum_{n'=0}^{N} J(n', t+1)\ W(n'|n_t, u_t), \quad \text{[S5]}$$

the transition probability W is given by the binomial update rule, where we expand the rate in $\varepsilon = 1/N \ll 1$

$$W(n'|n) = \binom{N}{n'} [p(x)]^{n'}\ [1-p(x)]^{N-n'}, \quad x = \frac{n}{N}, \quad \text{[S6]}$$

$$p(x) = \frac{(1+s+u_0)x}{1+(s+u_0)x} = x + \varepsilon\, f(x) + \mathcal{O}(\varepsilon^2), \quad \text{[S7]}$$

$$f(x) = (\sigma + u)x(1-x), \quad \text{with} \quad \sigma \equiv N\, s,\ u \equiv N\, u_0. \quad \text{[S8]}$$

In the cost-to-go equation, we also switch to continuous variables and rescale J by

$$J(n,t) \to \tilde{J}(x,\tau) \equiv \frac{1}{N}\, J\left(Nx, \frac{t}{N}\right), \quad \text{[S9]}$$

$$\tilde{J}(x_\tau, \tau) = \min_{u_\tau} \sum_{\Delta x} \tilde{J}(x_\tau + \Delta x, \tau + \varepsilon) W(\Delta x | x_\tau, u_\tau). \quad \text{[S10]}$$

We drop the tilde, set $x_\tau = x$, and perform an expansion of above equation in ε:

$$\sum_{\Delta x} J(x + \Delta x, \tau + \varepsilon)\ W(\Delta x | x, u_\tau) \quad \text{[S11]}$$

$$= \left(1 + \varepsilon\partial_\tau + \mathcal{O}(\varepsilon^2)\right) \sum_{l=0}^{\infty} \frac{1}{l!} \left(\partial_x^{(l)} J(x,\tau)\right) \sum_{\Delta x} \Delta x^l\ W(\Delta x | x, u_\tau). \quad \text{[S12]}$$

The last term is the lth moment of the jump distribution and can be found via the moment generating function of the binomial distribution:

$$a_l(x,u) \equiv \sum_{\Delta x} \Delta\, x^l W(\Delta x | x, u) = \partial_z^{(l)}|_{z=0} \sum_{\Delta x} e^{z\Delta x} W(\Delta x | x, u) \quad \text{[S13]}$$

$$= \partial_z^{(l)}|_{z=0} \exp\left[\varepsilon\left(z\, f(x,u) + \frac{z^2}{2} x(1-x)\right) + \mathcal{O}(\varepsilon^2)\right]. \quad \text{[S14]}$$

Only the first two moments remain to leading order in ε.

$$a_1(x,u) = \varepsilon\, f(x,u) + \mathcal{O}(\varepsilon^2), \quad \text{[S15]}$$

$$a_2(x,u) = \varepsilon\, x(1-x) + \mathcal{O}(\varepsilon^2), \quad \text{[S16]}$$

$$a_{k\geq 3}(x,u) = \mathcal{O}(\varepsilon^2). \quad \text{[S17]}$$

Altogether, we have for the expansion above

$$\begin{aligned} J(x,\tau) &= \min_{u_\tau} \sum_{\Delta x} J(x + \Delta x, \tau + \varepsilon) W(\Delta x | x, u_\tau) \\ &= \min_{u_\tau} \left[1 + \varepsilon\left(\partial_\tau + f(x,u_\tau)\ \partial_x + \frac{1}{2} x(1-x)\partial_x^{(2)}\right) \right. \\ &\qquad \left. + \mathcal{O}(\varepsilon^2)\right] J(x,\tau), \end{aligned} \quad \text{[S18]}$$

and with that the Hamilton–Jacobi–Bellman equation for the Wright–Fisher process in the diffusion limit, i.e. to lowest order in ε:

$$-\partial_\tau J(x,\tau) = \min_{u_\tau} \left[f(x,u_\tau)\partial_x + \frac{1}{2} x(1-x)\partial_x^{(2)}\right] J(x,\tau). \quad \text{[S19]}$$

If the control strength u only appears linearly in the drift term $f(x,u)$ and is bounded, then the minimum is attained at one of the extremal values, i.e. is of bang-bang type.

Analytical Evaluation of the Mean First Passage Time. Under bang-bang control, the effective selection coefficient $\sigma+\overline{u}$ is frequency dependent but still piecewise constant. For $\sigma>0$ and $\sigma-u_c<0$, the population experiences an upward drift for $x<x_c$ and a downward drift for $x>x_c$. If the drift forces in both domains are strong ($\sigma\gg 1$ and $\sigma-u_c\ll -1$), then a typical population that is still polymorphic will most likely be in the vicinity of x_c at any one point in time (see also Fig. 1*A* in the main text). One can then try to calculate the mean first passage time for trajectories starting at x_c. The formula can be found using standard theory of stochastic processes (2). Let us momentarily reinsert an arbitrary initial frequency x_0,

$$\langle T\rangle_{x_0}=\frac{\left(\int_0^{x_0}\frac{\mathrm{d}y}{\psi(y)}\right)\int_{x_0}^{1}\frac{\mathrm{d}y'}{\psi(y')}\int_0^{y'}\frac{\mathrm{d}z\;\psi(z)}{z(1-z)}-\left(\int_{x_0}^{1}\frac{\mathrm{d}y}{\psi(y)}\right)\int_0^{x_0}\frac{\mathrm{d}y'}{\psi(y')}\int_0^{y'}\frac{\mathrm{d}z\;\psi(z)}{z(1-z)}}{\frac{1}{2}\int_0^1\frac{\mathrm{d}y}{\psi(y)}},\qquad\text{[S20]}$$

$$\text{with}\quad \psi(z)\equiv\exp[2\sigma z-2u_c\;(z-x_c)\Theta(z-x_c)].\qquad\text{[S21]}$$

The mean first passage time depends on the initial point x_0, which can be either below or above x_c. This will affect all of the integrals, so let us write

$$\langle T\rangle_{x_0}=\langle T\rangle_{x_0}^{+}\Theta(x_0-x_c)+\langle T\rangle_{x_0}^{-}\Theta(x_c-x_0),\qquad\text{[S22]}$$

$$\text{with}\quad \langle T\rangle_{x_0}^{\pm}=\frac{I_{0,x_0}^{\pm}\,J_{x_0,1}^{\pm}-I_{x_0,1}^{\pm}\,J_{0,x_0}^{\pm}}{\frac{1}{2}\,I_{0,1}}.\qquad\text{[S23]}$$

The integrals in this ratio can now be computed one by one. They all have analytical solutions.

$$I_{0,x_0}\equiv\int_0^{x_0}\frac{\mathrm{d}y}{\psi(y)}=I_{0,x_0}^{+}\Theta(x_0-x_c)+I_{0,x_0}^{-}\Theta(x_c-x_0),\qquad\text{[S24]}$$

$$I_{0,x_0}^{-}=\frac{1}{2\sigma}\left(1-e^{-2\sigma x_0}\right),\qquad\text{[S25]}$$

$$I_{0,x_0}^{+}=\frac{1}{2\sigma}\left(1-e^{-2\sigma x_c}\right)+\frac{e^{-2u_cx_c}}{2(\sigma-u_c)}\left(e^{-2(\sigma-u_c)x_c}-e^{-2(\sigma-u_c)x_0}\right),\qquad\text{[S26]}$$

$$I_{x_0,1}\equiv\int_{x_0}^{1}\frac{\mathrm{d}y}{\psi(y)}=I_{x_0,1}^{+}\Theta(x_0-x_c)+I_{x_0,1}^{-}\Theta(x_c-x_0),\qquad\text{[S27]}$$

$$I_{x_0,1}^{-}=\frac{1}{2\sigma}\left(e^{-2\sigma x_0}-e^{-2\sigma x_c}\right)+\frac{e^{-2u_cx_c}}{2(\sigma-u_c)}\left(e^{-2(\sigma-u_c)x_c}-e^{-2(\sigma-u_c)}\right),\qquad\text{[S28]}$$

$$I_{x_0,1}^{+}=\frac{e^{-2u_cx_c}}{2(\sigma-u_c)}\left(e^{-2(\sigma-u_c)x_0}-e^{-2(\sigma-u_c)}\right),\qquad\text{[S29]}$$

$$J_{x_0,1}\equiv\int_{x_0}^{1}\frac{\mathrm{d}y'}{\psi(y')}\int_0^{y'}\frac{\mathrm{d}z\;\psi(z)}{z(1-z)}=J_{x_0,1}^{+}\Theta(x_0-x_c)+J_{x_0,1}^{-}\Theta(x_c-x_0),\qquad\text{[S30]}$$

$$\begin{aligned}J_{x_0,1}^{-}=&\,(G(x_0,\sigma)-G(x_c,\sigma))-\frac{F(0,\sigma)}{2\sigma}\left(e^{-2\sigma x_0}-e^{-2\sigma x_c}\right)\\&+e^{-2u_cx_c}\;(G(x_c,\sigma-u_c)-G(1,\sigma-u_c))+\frac{e^{-2u_cx_c}}{2(\sigma-u_c)}\\&\times\left(e^{-2(\sigma-u_c)x_c}-e^{-2(\sigma-u_c)}\right)(F(x_c,\sigma)-F(0,\sigma)-F(x_c,\sigma-u_c)),\end{aligned}\qquad\text{[S31]}$$

$$\begin{aligned}J_{x_0,1}^{+}=&\frac{e^{-2u_cx_c}}{2(\sigma-u_c)}\left(e^{-2(\sigma-u_c)x_0}-e^{-2(\sigma-u_c)}\right)\\&\times(F(x_c,\sigma)-F(0,\sigma)-F(x_c,\sigma-u_c))\\&+e^{-2u_cx_c}\;(G(x_0,\sigma-u_c)-G(1,\sigma-u_c)),\end{aligned}\qquad\text{[S32]}$$

$$J_{0,x_0}\equiv\int_0^{x_0}\frac{\mathrm{d}y'}{\psi(y')}\int_0^{y'}\frac{\mathrm{d}z\;\psi(z)}{z(1-z)}=J_{0,x_0}^{+}\Theta(x_0-x_c)+J_{0,x_0}^{-}\Theta(x_c-x_0),\qquad\text{[S33]}$$

$$J_{0,x_0}^{-}=(G(0,\sigma)-G(x_0,\sigma))-\frac{F(0,\sigma)}{2\sigma}\left(1-e^{-2\sigma x_0}\right),\qquad\text{[S34]}$$

$$\begin{aligned}J_{0,x_0}^{+}=&\,(G(0,\sigma)-G(x_c,\sigma))-\frac{F(0,\sigma)}{2\sigma}\left(1-e^{-2\sigma x_c}\right)\\&+e^{-2u_cx_c}(G(x_c,\sigma-u_c)-G(x_0,\sigma-u_c))\\&+\frac{e^{-2u_cx_c}}{2(\sigma-u_c)}\left(e^{-2(\sigma-u_c)x_c}-e^{-2(\sigma-u_c)x_0}\right)\\&\times(F(x_c,\sigma)-F(0,\sigma)-F(x_c,\sigma-u_c)).\end{aligned}\qquad\text{[S35]}$$

The solutions include the following functions:

$$F(x,\sigma)\equiv\mathrm{Ei}(2\sigma x)-e^{2\sigma}\mathrm{Ei}(-2\sigma(1-x)),\qquad\text{[S36]}$$

$$G(x,\sigma)\equiv\frac{e^{-2\sigma x}}{2\sigma}F(x,\sigma)+\frac{1}{2\sigma}\log\left(\frac{1-x}{x}\right),\qquad\text{[S37]}$$

$$\mathrm{Ei}(z)\equiv-\int_{-z}^{\infty}\mathrm{d}t\;\frac{e^{-t}}{t},\qquad\text{[S38]}$$

where we also used the following identities:

$$F(x,\sigma) = -e^{-2\sigma} F(1-x,\,-\sigma) \quad \text{and} \quad G(x,\sigma) = G(1-x,\sigma). \quad \textbf{[S39]}$$

Finally, $\langle T \rangle_{x_0 = x_c}$ can be evaluated numerically and maximized with respect to x_c to find this critical control switch frequency. The result is shown in Fig. S1 and is compared with the corresponding numerical result of the cost-to-go backward iteration for the discrete system.

Finite Mutation Rate. The control objective for the Wright–Fisher model is to avoid the boundaries at $x = 0$ and $x = 1$, which are absorbing in the absence of mutation. For very small values of mutation rate $\mu \ll 1$, these boundaries are still "sticky" in the sense that populations spend on the order of $1/\mu$ generations at a boundary and need $N/\mu\sigma$ generations to escape the drift-dominated boundary region. Interestingly, the value of the control switch frequency x_c, found by numerical integration of the cost-to-go backward iteration, is almost unaffected by a symmetric mutation rate for values of μ up to 10 (Fig. S2). At that point the original control task ceases to be meaningful because typical populations are almost never at the boundaries for more than a single generation. A more interesting situation can be observed when mutation rates are asymmetric: $\mu_{AB} \neq \mu_{BA}$. An extreme situation is constructed in Fig. S2, *Lower*, where initially $x_c \approx 0.9$ without mutation and moves toward $x = 0$ with increasing μ_{BA} while keeping $\mu_{AB} = 0$. This makes the (repulsive) lower boundary at $x = 0$ an increasingly safer place to be.

Minimal Model of Drug Resistance in Cancer

The qualitative aspects of the minimal cancer model introduced in the main text can be analyzed using a system size expansion (1), with the carrying capacity K as a large parameter. It is important to note that in this diffusion limit the details of the microscopic model are not important. For example, the effects of selection or carrying capacity could be included in the death rates, without changing the qualitative aspects of the model. By definition, the stochastic dynamics of the model is encoded in the following update rule for the probability distribution $P(n_s, n_r, t)$ of the number of drug-sensitive and drug-resistant cells

$$P(n_s, n_r, t+1) = \sum_{n_s', n_r'} W(n_s, n_r | n_s', n_r')\; P(n_s', n_r', t), \quad \textbf{[S40]}$$

where the matrix W of transition probabilities factorizes over cell types

$$W(n_s, n_r | n_s', n_r') = W(n_s | n_s')\; W(n_r | n_r'). \quad \textbf{[S41]}$$

Because negative cell number counts are impossible, care needs to be taken for the boundary terms of the transition probability.

$$W(n_i | n_i') = \begin{cases} \text{Skellam}(n_i - n_i';\; B_i(n_s, n_r),\; D_i(n_s, n_r)), & n_i > 0 \\ \text{CDF(Skellam)}(-n_i';\; B_i(n_s, n_r),\; D_i(n_s, n_r)), & n_i = 0 \end{cases} \quad \textbf{[S42]}$$

with the Skellam probability mass function defined below in Eq. **S57** and cell birth and death rates defined in Eq. **5** in the main text (CDF denotes the cumulative density function). The system size (or diffusion) expansion entails the parameter scaling

$$K \to \infty \quad \text{with} \quad \gamma \equiv Kg,\; \alpha \equiv Ka,\; \mu \equiv K\mu_0,\; \phi_{s,r} \equiv Kf_{s,r} \quad \text{const.}, \quad \textbf{[S43]}$$

together with a scaling of time via $\tau = t/N$ (with t measured in generations, i.e. population updates via Eq. **S40**). The relevant relative scales of the model parameters are

$$K \gg \phi > \gamma \gg \alpha,\; \mu \geq 0. \quad \textbf{[S44]}$$

The expansion of birth and death rates in carrying capacity K is as follows:

$$\begin{aligned} B_s(x_s, x_r) &= \frac{(K+\gamma+\alpha)\, K\, x_s}{K + \gamma(x_s + x_r) + \alpha\, x_s} + \mu(x_r - x_s) \\ &= K\, x_s + \gamma\, x_s(1 - x_s - x_r) + \alpha\, x_s\,(1 - x_s) \\ &\quad + \mu(x_r - x_s) + \mathcal{O}(K^{-1}) \\ &\equiv K\, x_s + b_s(x_s, x_r) + \mathcal{O}(K^{-1}), \end{aligned} \quad \textbf{[S45]}$$

$$\begin{aligned} B_r(x_r, x_s) &= K\, x_r + \gamma\, x_r\,(1 - x_s - x_r) - \alpha\, x_s\, x_r + \mu(x_s - x_r) + \mathcal{O}(K^{-1}) \\ &\equiv K\, x_r + b_r(x_s, x_r) + \mathcal{O}(K^{-1}), \end{aligned} \quad \textbf{[S46]}$$

$$D_s(x_s) = K\, x_s + u\;\phi_s\, x_s \equiv K\, x_s + d_s(x_s), \quad \textbf{[S47]}$$

$$D_r(x_r) = K\, x_r + (1-u)\phi_r\, x_r \equiv K\, x_r + d_r(x_r). \quad \textbf{[S48]}$$

The differential growth rate α and the drug-related death rates $\phi_{s,r}$ break the symmetry of the model, such that there is no closed growth law for the total population size $N = n_s + n_r$ alone: even ignoring boundary terms (at $n_s = 0$ and $n_r = 0$) the tumor size would evolve according to

$$\Delta N \sim \text{Skellam}(B_s + B_r, D_s + D_r), \quad \textbf{[S49]}$$

$$\begin{aligned} \langle \Delta N \rangle &= B_s + B_r - D_s - D_r = b_s + b_r - d_s - d_r \\ &= \gamma\, x(1-x) + \alpha\, x_s(1-x) - u\;\phi_s\, x_s - (1-u)\phi_r\, x_r + \mathcal{O}(K^{-1}), \end{aligned} \quad \textbf{[S50]}$$

with $x \equiv x_s + x_r = N/K$. The role of K as carrying capacity (for $u = 0,\; \phi_r = 0$) is now apparent via $\langle \Delta N \rangle(x = 1) = 0$.

The Fokker–Planck equation for this model follows by the same logic and the same procedure from the generation update model Eq. **S40** as in the Wright–Fisher model above. In the variables (x_s, x_r) it is given by [with $P = P(x_s, x_r, \tau)$]

$$\partial_\tau\, P = \left[-\partial_{x_s}(b_s - d_s) - \partial_{x_r}(b_r - d_r) + \left(\partial_{x_s}^2 + \partial_{x_r}^2\right)(x_s + x_r) \right] P, \quad \textbf{[S51]}$$

$$b_s - d_s = (\gamma + \alpha) x_s(1 - x_s) - \gamma\, x_s\, x_r + \mu(x_r - x_s) - u\;\phi_s x_s, \quad \textbf{[S52]}$$

$$b_r - d_r = \gamma\, x_r(1 - x_r) - (\gamma + \alpha)\, x_s\, x_r + \mu(x_s - x_r) - (1-u)\phi_r\, x_r, \quad \textbf{[S53]}$$

with τ measuring time on the scale of K generations. One important observation is that the control strength again appears only linearly in the drift term. This is also true for the corresponding Hamilton–Jacobi–Bellman equation, such that control would be of bang-bang type. For the risk-sensitive control objective in the main text, this is true to leading order in the metastasis rate ν_0 (5).

The form of the birth and death rates above suggests a transformation of variables:

$$(x_s, x_r) \to \left(x \equiv x_s + x_r,\; y \equiv \frac{x_s}{x_s + x_r} \right) \Rightarrow (x_s = x\, y,\; x_r = x(1-y)). \quad \textbf{[S54]}$$

The time evolution of the mean values of these new variables is now given by (1)

$$\begin{aligned}\partial_\tau\langle x\rangle &= \langle b_s + b_r - d_s - d_r\rangle \\ &= \langle(\gamma+\alpha y)x(1-x) - x(u\phi_s y + (1-u)\phi_r(1-y))\rangle,\end{aligned} \quad \text{[S55]}$$

$$\begin{aligned}\partial_\tau\langle y\rangle &= \left\langle \frac{1-y}{x}(b_s-d_s) - \frac{y}{x}(b_r-d_r)\right\rangle \\ &= \langle(\alpha - u\phi_s + (1-u)\phi_r)\, y(1-y) + \mu(1-2y)\rangle.\end{aligned} \quad \text{[S56]}$$

The evolution of the mean relative fraction $\langle y\rangle$ of sensitive cells is equivalent to the evolution of the mean value of the polymorphism frequency within the controlled one-locus two-allele Wright–Fisher model discussed earlier (Eq. **S1**).

Numerical Test of the Cost-To-Go Calculation. An optimal control strategy fulfilling Eq. **8** in the main text can be found numerically by backward iteration of the cost-to-go recurrence Eq. **9** using the exact discrete propagator $W(n_s', n_r'|n_s, n_r)$ defined by the Skellam distribution implicit in Eq. **7**. To test the resulting profile $\overline{u}(x_s, x_r)$, we can evaluate the associated cost function directly using a large ensemble of forward simulations. It should be noted that due to memory and time limitations, the backward iteration can only be performed with a rather small system size on the order of $\max(n_s, n_r) \leq 10^3$. For the forward simulations, only the milder time restriction holds, such that $K \sim \mathcal{O}(10^4)$ is possible. To make the two results comparable, it is necessary to use the same scaled parameters $\alpha = Ka$, etc. In Fig. S5, we compare the probability that metastasis has not yet occurred by time $T = K/\nu$ generations (the control objective to be maximized) as predicted by the cost-to-go calculation with the direct observation of this event in 10^3 forward simulations with $K = 10^4$ in the parameter setting of Fig. 2*A* in the main text.

Uniform Expansion of the Skellam Distribution. The probability mass function of the Skellam distribution with parameters (μ_1, μ_2) is given by

$$n_1 \sim \text{Pois}(\mu_1), n_2 \sim \text{Pois}(\mu_2) \Rightarrow n \equiv n_1 - n_2 \sim \text{Skellam}(\mu_1, \mu_2), n \in \mathbb{Z}$$

$$\text{with} \quad \text{Skellam}(n|\mu_1,\mu_2) = e^{-\mu_1-\mu_2}\left(\frac{\mu_1}{\mu_2}\right)^{n/2} I_{|n|}\left(2\sqrt{\mu_1\mu_2}\right). \quad \text{[S57]}$$

The modified Bessel function $I_n(z)$ could, in principle, be evaluated for fixed z via the following recurrence relation:

$$I_{n-1}(z) - I_{n+1}(z) = \frac{2n}{z} I_n(z). \quad \text{[S58]}$$

However, due to a lack of numerical stability of this recurrence, here we have used the uniform expansion of the Bessel function instead (6),

$$I_\nu(\nu z) \xrightarrow{\nu\to\infty} \frac{e^{\nu\eta}}{\sqrt{2\pi\nu}(1+z^2)^{1/4}}\left(1 + \mathcal{O}\left(\frac{1}{\nu}\right)\right), \quad \text{[S59]}$$

$$\text{with} \quad \eta \equiv \sqrt{1+z^2} + \ln(z) - \ln\left(1+\sqrt{1+z^2}\right). \quad \text{[S60]}$$

This expansion has the additional benefit that we can simply use it for the logarithm of the Skellam distribution,

$$\begin{aligned}\log \text{Skellam}(n) \approx a + b\, n + \|(n,z)\| - \frac{1}{2}\log\|(n,z)\| \\ + |n|\log\frac{z}{|n| + \|(n,z)\|},\end{aligned} \quad \text{[S61]}$$

$$\begin{aligned}\text{with} \quad a &\equiv -(\mu_1+\mu_2) - \frac{1}{2}\log(2\pi), \\ b &\equiv \frac{1}{2}\log\left(\frac{\mu_1}{\mu_2}\right), \\ z &\equiv 2\sqrt{\mu_1\mu_2}, \\ \|(n,z)\| &\equiv \sqrt{n^2+z^2}.\end{aligned}$$

The quality of this approximation is also implicit in the simulation test results shown in Fig. S5.

Including a Control Cost to the Cancer Model. In the main text, the control objective for the cancer model was to avoid a next-stage mutation. Eqs. **8** and **9** describe the situation when control itself is free of cost. One way to introduce an implicit cost of using the drug is by assuming that each application of the drug $(u = 1)$ carries a finite but small probability β_0 per generation that the patient dies as a result of this dose. We consider here only ($u = 0$ or 1), i.e. on–off application of the drug. The equations then generalize to

$$\overline{u}_{0:T}(n_{s0}, n_{r0}) = \underset{u_{0:T}}{\text{argmax}} \exp\left(-\nu_0 \sum_{t=0}^{T}(n_{st}+n_{rt}) - \beta_0 \sum_{t=0}^{T} u_t\right), \quad \text{[S62]}$$

$$J(n_s, n_r, t) = e^{-\nu_0(n_s+n_r)} \max_{u\in\{0,1\}} e^{-\beta_0 u}\langle J(t+1;\, u)\rangle. \quad \text{[S63]}$$

The numerical solution to the recurrence relation can be seen in Fig. S4 below for the parameter sets of main text Fig. 2 *C* and *D* and increasing values of $\beta = K\beta_0$.

Control with Limited Information (*N* only). The optimal control profiles shown in Fig. 2 in the main text are only applicable with perfect information of the tumor composition (n_s, n_r). If only the total population size N can be measured, then there is a different strategy to make a control decision: compute the reduced propagator $W(N_{\tau+\Delta\tau}|N_{\tau-\Delta\tau}, u_{\tau-\Delta\tau}, N_\tau; u_\tau)$ and derive a new control profile that depends on the last two measurements and the last applied control. This is clearly an approximation, such that the resulting control protocol cannot be considered optimal in the mathematical sense. In deriving the reduced propagator, we use the shorthand notation $N' = N_{\tau+\Delta\tau}$, $N = N_\tau$, $M = N_{\tau-\Delta\tau}$, $u = u_\tau$, $v = u_{\tau-\Delta\tau}$, and $n = n_s$.

$$\begin{aligned}W(N'|N,M,v;u) = \sum_{n'=0}^{N'}\sum_{n=0}^{N}\sum_{m=0}^{M} W(n', N'-n'|n, N-n; u)\times\ldots \\ \cdots \times P(n, N-n|m, M-m, N, v) \\ \times P(m, M-m|M, N, v).\end{aligned} \quad \text{[S64]}$$

The first term on the right-hand side is the microscopic propagator, expressed as the product of the two Skellam distributions for n_s and n_r. The second term is the probability to go from $(m, M-m)$ to $(n, N-n)$ under control v, given that the final population size is N,

$$P(n,N-n|m,M-m,N,v)=\frac{W(n,N-n|m,M-m;v)}{\sum_{k=0}^{N}W(k,N-k|m,M-m;v)}.$$

[S65]

The last term is the probability that the system was at $(m, M-m)$, given that a transition took place from M to N under control v,

$$P(m,M-m|M,N,v)=\frac{\sum_{n=0}^{N}W(n,N-n|m,M-m;v)}{\sum_{k=0}^{M}\sum_{n=0}^{N}W(n,N-n|k,M-k;v)}.$$

[S66]

All these conditional probabilities can be approximated using the logarithmic expansion of the Skellam distribution above. For the parameter setting of Fig. 2*C* in the main text, the resulting control profile is shown in Fig. S6.

1. Van Kampen NG (2007) *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam), 3rd Ed, Vol 1.
2. Gardiner CW (2009) *Stochastic Methods, Springer Series in Synergetics* (Springer, Berlin), 4th Ed, Vol 13.
3. Ewens WJ (2004) *Mathematical Population Genetics* (Springer, New York).
4. Kappen HJ (2005) Path integrals and symmetry breaking for optimal control theory. *J Stat Mech* 2005:P11011.
5. Broek B, Wiegerinck W, Kappen H (2012) Risk sensitive path integral control. arXiv: 1203.3523.
6. Olver F (1974) *Asymptotics and Special Functions* (Academic, New York).

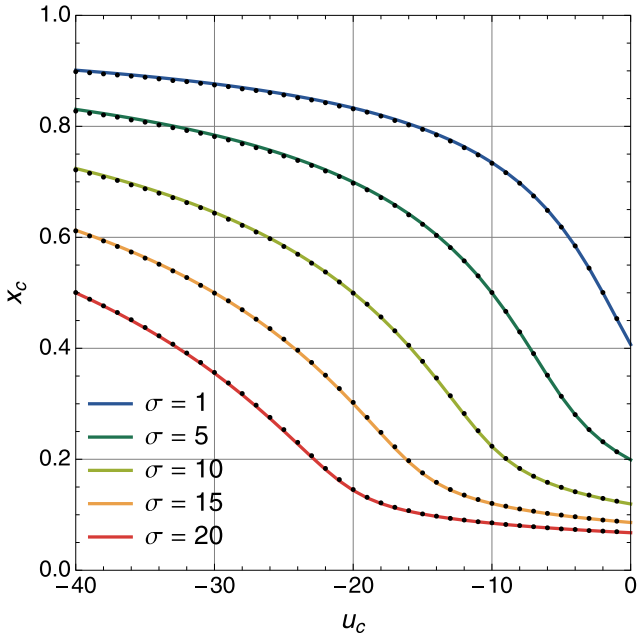

Fig. S1. Control-switch frequency $x_c(u_c,\sigma)$ found by maximizing the analytical expression for the mean survival time (solid lines) and by backward iteration of Eq. **2** in the main text. Note that $x_c(-2\sigma,\sigma)=0.5$.

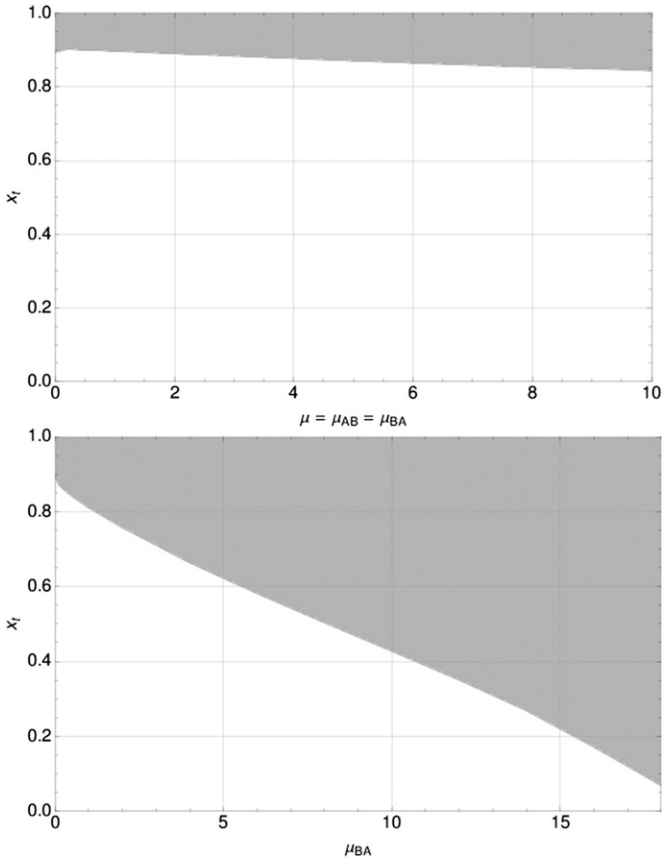

Fig. S2. Control-switch frequency x_c found by backward iteration of Eq. **2** in the main text for finite mutation rates. For both panels, $\sigma = 10$ and $\sigma - u_c = -100$. (*Upper*) The symmetric mutation rate does not greatly change the initial value of $x_c \approx 0.9$ ($u = -u_c$ in the gray area and $u = 0$ in the white area). (*Lower*) $\mu_{AB} = 0$ such that the upper boundary is still absorbing while the lower boundary becomes safer for increasing values of μ_{BA}.

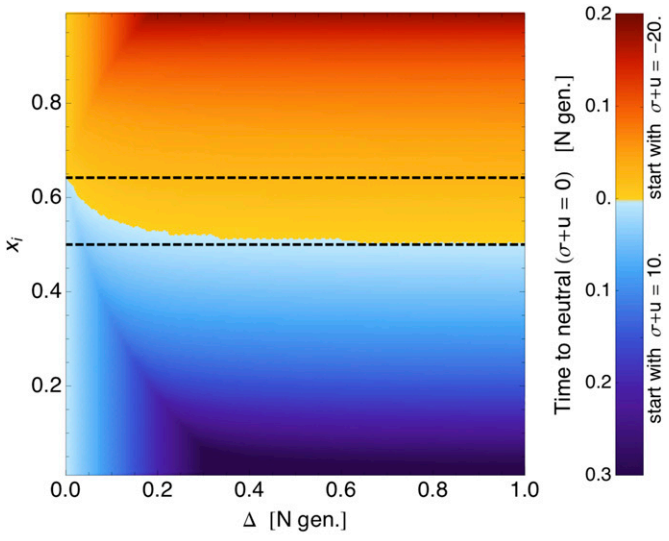

Fig. S3. For finite time Δ between consecutive measurements, the preemptive control aims for a safe position x_{safe} away from the boundaries (boundary between blue and orange) by switching to a neutral regime $(u = -\sigma)$ after a certain waiting time (see color bar on the right). At x_{safe}, the waiting time to neutral is zero, i.e. the system is immediately set to neutral. As Δ becomes bigger, x_{safe} moves from $x_c \approx 0.64$ to 0.5 and the control strategy shifts from playing-to-win to playing-not-to-lose.

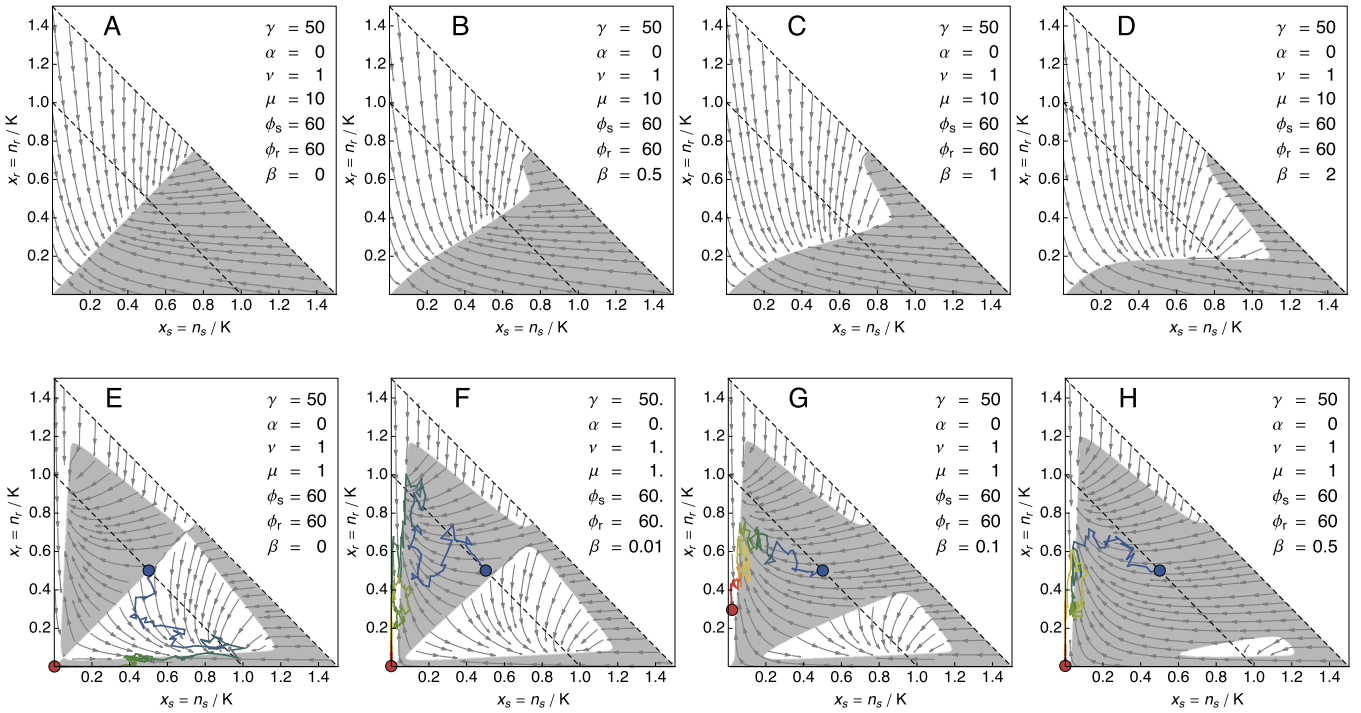

Fig. S4. Influence of a control cost on the perfect information control profiles as seen in main text Fig. 2. (*A–D*) The finite control cost reduces the area where the drug is applied ($u = 1$, gray area). In the lower row (*E–H*), the increasing control cost actually increases the area of drug application. However, the typical time that the drug is applied is reduced because controlled trajectories tend to spend more time at the edges.

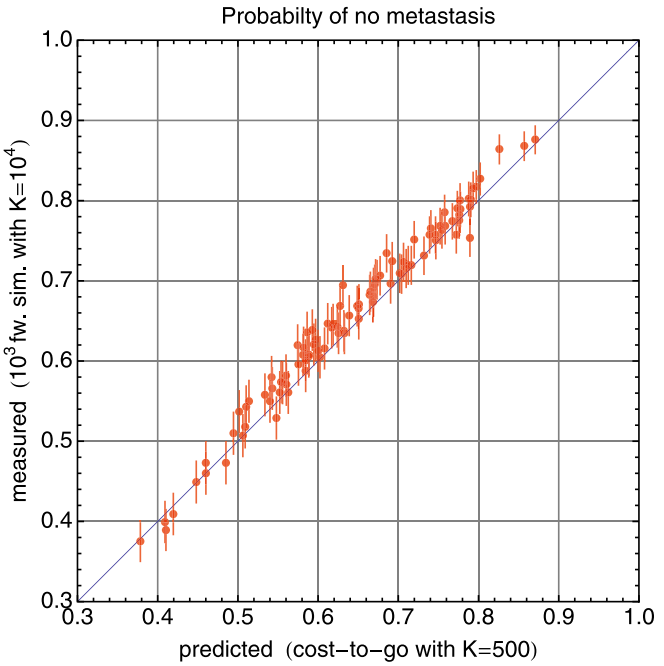

Fig. S5. Comparison of the predicted probability that metastasis has not yet occurred by time $T = K/\nu$ generations in a cancer cell population optimally controlled according to the profile (and parameters) shown in main text Fig. 2*A* to the measured fraction of 10^4 forward simulations with that property. The prediction follows from the cost-to-go dynamic programming calculation (see Eq. **9** in the main text) performed numerically with $K = 500$ and $N \leq 750$. The forward simulations were carried out with $K = 10^4$, $N \leq 1.5K$.

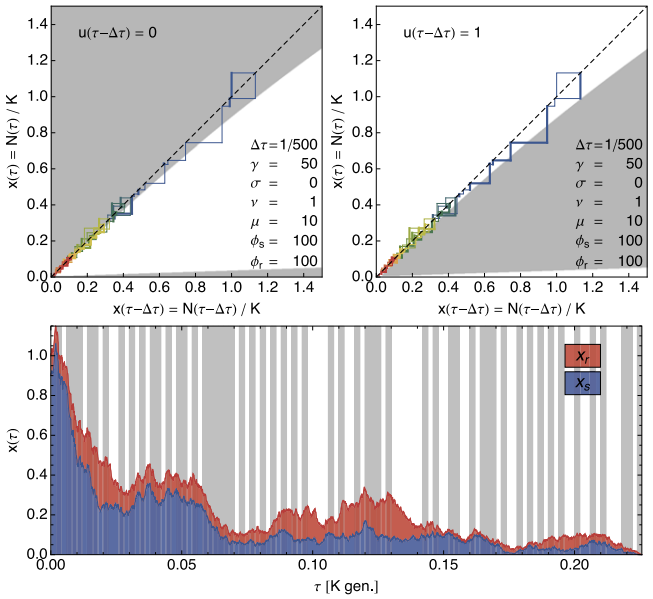

Fig. S6. Control of a tumor via its total size. In the parameter setting of Fig. 2*C*, a majority rule would be optimal with perfect information. However, when only the total population size $N = n_s + n_r$ can be measured, the needed information is not directly available. The control profile above tries to estimate the inner composition of the tumor indirectly from the immediate response $N(\tau - \Delta\tau) \to N(\tau)$ to the presence $[u(\tau - \Delta\tau) = 1$, gray areas] or absence of the drug $[u(\tau - \Delta\tau) = 0$, white areas]. A controlled trajectory is shown on top of the control profile, where thick lines indicate a response big enough to continue the current drug regimen. (*Upper*) The same trajectory as a function of time is shown as a stacked area plot (*Lower*). The color of the background corresponds to the current control setting.