# Star Methods

# Method Details

# A. Comparison to other models

We next briefly discuss our interpretation of some existing models for self/nonself dynamic discrimination in immune systems, and compare them to the IFFL model.

## Tunable Activation Threshold (TAT)

This model was suggested by Grossman and Paul (Grossman and Paul 1992) in 1992, motivated by the realization that "self/nonself discrimination may be much more complex than the simple failure of competent lymphocytes to recognize self-antigens". The authors argued that for a stimulus to cause cell activation, the excitation level must exceed an activation threshold, and when engaged in persistent sub-threshold interactions, cells are protected against chance activation. In the TAT model, an *activation threshold* for an immune cell is dynamically modulated by an environment-dependent recent excitation history. This history is summarized by an *excitation index*, which we will denote as $x(t)$, which computes a sort of weighted average of the cell's past excitation levels. Given temporal excitation events, which we denote by $u(t)$, it is assumed that the cell undergoes perturbations that depend on the difference between $u(t)$ and the memory variable $x(t)$. The key assumption is that such a perturbation, which we write as $y(t) := u(t) - x(t)$, must exceed a fixed critical value, which we denote by $\theta$, in order to cause activation. In other words, it must be the case that $u(t) - x(t) > \theta$, or equivalently, $u(t) > x(t) + \theta$ (this is how we interpret the statement in (Grossman and Paul 1992) that "the activation threshold equals the excitation index plus that critical value") for activation to occur. Cells maintained at a high level of excitation $x(t)$ therefore are relatively insensitive to activation, thus being in some sense anergic. The authors deduce from their model that "upon gradual increase in the levels of excitation…a cell is not likely to be activated…it will become progressively anergic" which is intuitively equivalent to our remark about the lack of continued excitation under slow increases in antigen presentation. With our notations, the model suggested in (Grossman and Paul 1992) is:

$$\dot{x} = \alpha u(u - x)$$

for some constant $\alpha$, and the output would be $y = u - x$. (No explicit population-based nor signaling mechanism was given.) Notice that we then can derive a differential equation for $y$:

$$\dot{y} = u(\dot{u}/u - \alpha y)$$

and this means, roughly, that $y$ should approach $\dot{u}/u$, the logarithmic derivative of the input $u$, so that a "log sensing" property is satisfied by the output. Moreover, when the input is constant, the output converges to the same value (zero), independently of the actual value of the input, so we have perfect adaptation. Moreover, we expect $y$ to be small

(and thus not exceeding the threshold $\theta$) unless $u$ changes fast in the sense that its logarithmic derivative is large. For example, for $u(t)$ increasing linearly, $\dot{u}$ would be a constant, so $\dot{u}/u = 0$ and therefore $y(t) \to 0$ as $t \to \infty$. On the other hand, for an exponentially increasing $u(t)$, $y(t)$ converges to a value proportional to the exponential rate. These properties are analogous to those satisfied by our model.

## Discontinuity theory of immunity

This model was suggested by Pradeu, Jaeger, and Vivier (Pradeu, Jaeger, and Vivier 2013) in 2013 as a "unifying theory of immunity". Their key hypothesis is that effector immune responses are induced by an "antigenic discontinuity" by which they mean a "sudden" modification of molecular motifs with which immune cells interact. The authors present evidence that natural killer (NK) cells and macrophages are activated by transient modifications, but adapt (ceasing to be responsive) to long-lasting modifications in their environment, and then propose to extend this principle to other components of the immune system, such as B cells and T cells. They also argue that although tumors give rise to effector immune responses, "a persistent tumour antigen diminishes the efficacy of the antitumor response". In summary, their criterion of immunogenicity is the phenomenological antigenic discontinuity and not the nature of the antigen, including both "discontinuities" arising from self motifs such as tumors as well as from non-self motifs such as bacterial or viral infections. As examples of mechanisms for desensitization they mention receptor internalization, degradation or inactivation of signaling proteins. A concrete example of the latter is the dephosphorylation triggered by immunoreceptor tyrosine-based inhibition motif (ITIM)-containing receptors antagonizing kinases triggered by immunoreceptor tyrosine-based activation motif (ITAM)-containing receptors. The authors also mention Treg population dynamics. Using our notations, the model in (Pradeu, Jaeger, and Vivier 2013) starts by computing a running average of the absolute value of consecutive differences in inputs presented at discrete times on a sliding window $K$ time units long:

$$\Delta u(t) := \frac{1}{K}\sum_{i=1}^{K} |u(t-i+1) - u(t-i)|$$

and then taking as the output $y = \Theta(\Delta u)$, where $\Theta$ is a sigmoidal saturating function. The authors employ

$$\Theta(x) = \frac{\alpha}{1 + e^{-\mu(x-\theta)}}$$

but one could equally well (and perhaps easier to justify mechanistically) employ a Hill-type function $\Theta(x) = \frac{Vx^n}{K^n + x^n}$. In continuous time, and assuming that the input is differentiable, we could interpret

$$\Delta u(t) \approx \left\| \dot{u}\,|_{[t-K,t]} \right\|_1 = \int_{t-K}^{t} |\dot{u}(s)|\, ds\,,$$

where the right hand term is the total variation of the input on this sliding window. Note the absolute value, which means that, in this model, activation is symmetrically dependent on increases or decreases of the excitation: decreases may help with "missing self" recognition, in which the expression of a "self" marker suddenly decreases, thus triggering a response. As with our model, slow variations in the input will lead to small $y(t)$, with the threshold function $\Theta$ resulting in an ultrasensitive, almost binary, response (provided that $\mu$ or $n$ are large, in the two suggested functions $\Theta$).

## Growth threshold conjecture

This model was suggested by Arias, Herrero, Cuesta, Acosta and Fernández-Arias (Arias et al. 2015) in 2015 as "a theoretical framework for understanding T-cell tolerance" based on the hypothesis that "T cells tolerate cells whose proliferation rates remain below a permitted threshold". As in the other works, the authors postulate that T cells tolerate cognate antigens (irrespectively of their pathogenicity) as long as their rate of production is low enough, while those antigens that are associated with pathogenic toxins or structural proteins of either infectious agents or aggressive tumor cells are highly proliferative, and therefore will be targeted as foes by T cells. In summary, once again the postulate is that a strong immune response will be mounted against fast-growing populations while slow-growing ones will be tolerated. The model in (Arias et al. 2015) is not one of change detection as such, but it is a closed-loop system that includes both detection and a killing effect on pathogens. To compare with our previous models, let us again denote the pathogen population size (or a density in a particular environment) by $u(t)$ and the effector cell population by $y(t)$. The authors give for $y$ a second order equation $\ddot{y} = -\delta y + \alpha u$, modeled on a spring-mass system that balances a "restoring to equilibrium force" to its activation by pathogens. We prefer to write the system as a set of first order ODE's. Thus, we let $x := \dot{y}$, and write:

$$\begin{aligned} \dot{u} &= (\lambda - \kappa y)u \\ \dot{x} &= \alpha u - \delta y \\ \dot{y} &= x\,. \end{aligned}$$

The $u$ equation has an exponential growth term balanced by a kill rate that depends on the effector population. The effector population integrates the amount of $x$ (which we might interpret as an intermediate type of cell); the growth of $x$ is driven by pathogens, with a negative feedback from $y$ (in essence an integral feedback on $x$), but there is no obvious biological mechanism for this model. Observe that when there is no pathogen, this results in a harmonic oscillator for $x$ and $y$, with sustained oscillations and even negative values. In any event, the authors computationally obtain a bifurcation-like diagram in the $(\lambda, \kappa)$ plane, dividing this plane into two regions, labeled "tolerance" (of infection, hence, failure of the immune system) and "intolerance". These regions show how to trade off the growth rate $\lambda$ of the pathogen versus the parameter $\kappa$, which represents a combination of affinity and clearance rate, and various conclusions regarding evasion strategies and the role of fever and even Treg cells are qualitatively derived from there.

## B. Theory for system without autocatalysis

We collect here mathematical results for the system when $h$ is linear and $f$ only contains linear degradation or inactivation terms. This system is easier to study theoretically than the system with feedback, and provides much intuition about the general case, besides it being a local approximation in suitable regimes. The equations are then as follows (we write $u$ as the last variable now, because we will separately study the first two equations):

[sys:production.expo.no_h]

$$\begin{aligned} \dot{x} &= -\delta_x x + \beta u \\ \dot{y} &= \mu \frac{u}{x} - \delta_y y \\ \dot{u} &= (\lambda - \kappa y) u \end{aligned}$$

The constants $\delta_x, \beta, \mu, \delta_y, \kappa$ are positive, but $\lambda$ is allowed to be negative, for completeness, although the interesting case is $\lambda \geq 0$. The scalar functions of time $x = x(t)$, $y = y(t)$, and $u = u(t)$ take positive values. It is easy to verify that, for any positive initial conditions, solutions remain positive for all times.

We will separately study the first two equations ([sys:production.expo.no_h]ab), viewing $u = u(t)$ as an external input to the IFFL described by ([sys:production.expo.no_h]ab), and viewing $y = y(t)$ as an output or response of the system. Later, we "close the loop".

**Remark.** In the system ([sys:production.expo.no_h]abc), and in particular in the system ([sys:production.expo.no_h]ab), one may assume without loss of generality that $\delta_x = b = c = 1$. This is because we may eliminate these parameters by rescaling variables. Indeed, substituting

$$x = \frac{\beta}{\delta_x} x^*, \; y = \frac{\mu}{\beta} y^*, \; t = \frac{1}{\delta_x} t^*, \; \delta_y^* = \frac{\delta_y}{\delta_x}, \; \lambda^* = \frac{\lambda}{\delta_x}, \; \kappa^* = \frac{\mu\kappa}{\delta_x \beta},$$

into system ([sys:production.expo.no_h]abc), one obtains:

[sys:production.expo.fixabc1]

$$\begin{aligned} \frac{dx^*}{dt^*} &= -x^* + u \\ \frac{dy^*}{dt^*} &= \frac{u}{x^*} - \delta_y^* y^* \\ \frac{du}{dt^*} &= (\lambda - \kappa y^*) u \end{aligned}$$

### IFFL's responses to various classes of inputs

Let us consider the system ([sys:production.expo.no_h]ab), a differentiable function $u = u(t)$ viewed as an external input or forcing function, and any (positive) solution $(x(t), y(t))$ corresponding to this input. We are interested first in understanding how the growth rate of the input affects the asymptotic values of the output variable $y$.

We denote the derivative of $\ln u(t)$ with respect to $t$ as follows:

$$v(t) := \frac{\dot{u}(t)}{u(t)}$$

and its limsup and liminf as $t \to \infty$

$$\underline{\mu} = \liminf_{t\to\infty} v(t)\,, \quad \overline{\mu} = \limsup_{t\to\infty} v(t)\,.$$

We assume that $v$ is bounded, and thus both of these numbers are finite. We also introduce the following function:

$$p(t) := \frac{u(t)}{x(t)}.$$

Since

$$\dot{p} = \dot{u}/x - u\dot{x}/x^2 = (u/x)[\dot{u}/u - \dot{x}/x] = (u/x)[\dot{u}/u - (-\delta_x x + \beta u)/x]$$
$$= (u/x)[\dot{u}/u + \delta_x - \beta u/x]\,,$$

we have that $p$ satisfies the following ODE with input $v$:

$$\dot{p} = p(\delta_x + v - \beta p)\,.$$

**Lemma.** Let $u$ be a differentiable input to system ([sys:production.expo.no_h]ab) with $\delta_x = \beta = \mu = 1$. With the above notations,

$$\max\{0, 1+\underline{\mu}\} \;\le\; \liminf_{t\to\infty} p(t) \;\le\; \limsup_{t\to\infty} p(t) \;\le\; \max\{0, 1+\overline{\mu}\}$$

*Proof.* Since $\delta_x = \beta = \mu = 1$,

$$\dot{p} = p(1 + v - p)\,.$$

To prove the upper bound, we consider two cases, $1+\overline{\mu} < 0$ and $1+\overline{\mu} \ge 0$. In the first case, let $\varepsilon := -(1+\overline{\mu}) > 0$; the definition of $\overline{\mu}$ gives that, for some $T \ge 0$, $1 + v(t) < -\varepsilon/2$ for all $t \ge T$. It follows that $\dot{p} \le p(-\varepsilon/2 - p)$ for all $t \ge T$. Thus, $\dot{p} < 0$ whenever $p > 0$, from which it follows that $\limsup_{t\to\infty} p(t) = \lim_{t\to\infty} p(t) = 0$. Suppose now that $1+\overline{\mu} \ge 0$. Pick any $\varepsilon > 0$ and a $T = T(\varepsilon) \ge 0$ such that $v(t) \le \overline{\mu} + \varepsilon$ for all $t \ge T$. For such $t$, $\dot{p} = p(1 + v - p) \le p(1 + \overline{\mu} + \varepsilon - p)$. This implies that $\dot{p} < 0$ whenever $p(t) > 1 + \overline{\mu} + \varepsilon$, which implies that $\limsup_{t\to\infty} p(t) \le 1 + \overline{\mu} + \varepsilon$. Letting $\varepsilon \to 0$, we conclude that $\limsup_{t\to\infty} p(t) \le 1 + \overline{\mu}$. We next prove the lower bound. Pick any $\varepsilon > 0$ and a $T = T(\varepsilon) \ge 0$ such that $v(t) \ge \underline{\mu} - \varepsilon$ for all $t \ge T$. Thus $\dot{p} = p(1 + v - p) \ge p(1 + \underline{\mu} - \varepsilon - p)$ for all $t \ge T$. This implies that $\dot{p} > 0$ whenever $p(t) < 1 + \underline{\mu} - \varepsilon$ (recall that $p(t) > 0$ for all $t$, since by assumption $u(t) > 0$ and $x(t) > 0$ for all $t$). Therefore $\liminf_{t\to\infty} p(t) \ge 1 + \underline{\mu} - \varepsilon$, and letting $\varepsilon \to 0$ we have $\liminf_{t\to\infty} p(t) \ge 1 + \underline{\mu}$. Since $p(t) \ge 0$ for all $t$, we also have

$\liminf_{t\to\infty} p(t) \geq \max\{0,1+\mu\}$. This completes the proof. In particular, if $v(t) \to \nu$ as $t \to \infty$ then $\mu = \overline{\mu} = \nu$, so we have as follows.

**Corollary.** If $v(t) \to \nu$ as $t \to \infty$ then $\lim_{t\to\infty} p(t) = \max\{0,1+\nu\}$.

For the original system ([sys:production.expo.no_h]ab), we have as follows.

**Proposition.** Consider a solution of ([sys:production.expo.no_h]ab), with a differentiable $u(t) > 0$ as input and $x(t) > 0, y(t) > 0$. Assuming that $\nu = \dot{u}/u$ is bounded, we have:

$$\frac{\mu}{\beta\delta_y}\max\left\{0, \delta_x + \mu\right\} \;\leq\; \liminf_{t\to\infty} y(t) \;\leq\; \limsup_{t\to\infty} y(t) \;\leq\; \frac{\mu}{\beta\delta_y}\max\{0, \delta_x + \overline{\mu}\}$$

*Proof.* We first assume that $\delta_x = \beta = \mu = 1$. Let $\underline{p} := \liminf_{t\to\infty} p(t)$ and $\overline{p} := \limsup_{t\to\infty} p(t)$. Equation ([sys:production.expo.no_h]b) can be written as $\dot{y} = p - \delta_y y$. This is a linear system forced by the input $p = p(t)$. Pick any $\varepsilon > 0$. Then there is some $T = T(\varepsilon)$ such that $\underline{p} - \varepsilon < p(t) < \overline{p} + \varepsilon$ for all $t \geq T$. For such $t$, $\dot{y}(t) > 0$ whenever $y(t) < (1/\delta_y)(\underline{p} - \varepsilon)$ and $\dot{y}(t) < 0$ whenever $y(t) > (1/\delta_y)(\overline{p} + \varepsilon)$. It follows that $(1/\delta_y)(\underline{p} - \varepsilon) \leq y(t) \leq (1/\delta_y)(\overline{p} + \varepsilon)$ for all $t \geq T$. Letting $\varepsilon \to 0$ we conclude that

$$\underline{p}/\delta_y \;\leq\; \liminf_{t\to\infty} y(t) \;\leq\; \limsup_{t\to\infty} y(t) \;\leq\; \overline{p}/\delta_y$$

and the inequalities follow when $\delta_x = \beta = \mu = 1$. To deal with general parameters, we recall that ([rescaling:production]ab) are obtained with $x = \frac{\beta}{\delta_x}x^*$, $y = \frac{\mu}{\beta}y^*$, $t = \frac{1}{\delta_x}t^*$, and $\delta_y^* = \frac{\delta_y}{\delta_x}$. Note that $t^* \to \infty$ if and only if $t \to \infty$. Thus holds for $p^* = u/x^* = (\beta/\delta_x)p$, $y^*$, and $\delta_y^\star$ in place of $p$, $y$, and $\delta_y$. Similarly, holds for $p^* = u/x^*$ and

$$\underline{\mu}^* = \liminf_{t\to\infty} v^*(t^*)\,, \quad \overline{\mu} = \limsup_{t\to\infty} v^*(t^*)\,,$$

where $v^* = \frac{du/dt^*}{u} = (1/\delta_x)v$, so $\underline{\mu}^* = (1/\delta_x)\underline{\mu}$ and $\overline{\mu}^* = (1/\delta_x)\overline{\mu}$. Therefore,

$$\begin{aligned}\liminf_{t\to\infty} y(t) \;&=\; \liminf_{t^*\to\infty} \frac{\mu}{\beta} y^*(t^*) \;\geq\; \frac{\mu}{\beta}\frac{\underline{p}^*}{\delta_y^\star} \;=\; \frac{\mu}{\beta}\frac{\underline{p}^*}{\delta_y/\delta_x} \;=\; \frac{\delta_x\mu}{\beta\delta_y}\underline{p}^* \;=\; \frac{\delta_x\mu}{\beta\delta_y}\max\left\{0,1+\underline{\mu}^*\right\} \\ &= \frac{\mu}{\beta\delta_y}\max\left\{\delta_x + \underline{\mu}\right\}.\end{aligned}$$

A similar remark applies to limsup, and the result follows.

**Corollary.** If $v(t) \to \nu$ as $t \to \infty$ then $\lim_{t\to\infty} y(t) = \frac{\mu}{\beta\delta_y}\max\{0, \delta_x + \nu\}$. Three particular cases are:

- When $u(t)$ has sub-exponential growth, meaning that $d\ln u/dt \leq 0$, then $\limsup_{t\to\infty} y(t) \leq \frac{\delta_x \mu}{\beta \delta_y}$.
- In particular, if $u(t) = K_0 + C_0 t$ is linear, then $\nu = 0$ and thus $\lim_{t\to\infty} y(t) = \frac{\delta_x \mu}{\beta \delta_y}$.
- If $u(t) = C_0 e^{\nu t}$ is exponential, then $\lim_{t\to\infty} y(t) = \frac{\mu}{\beta \delta_y} \max\{0, \delta_x + \nu\}$.

In conclusion, when $u$ is constant, or even with linear growth, the value of the output $y(t)$ converges to a constant, which does not depend on the actual constant value, or even the growth rate, of the input. For constant inputs, this is called the "perfect adaptation" property. If, instead, $u$ grows exponentially, then $y(t)$ converges to a steady state value that is a linear function of the logarithmic growth rate.

## IFFL's as feedback controllers

As we remarked, in the case of exponential inputs $u(t) = e^{\nu t}$, $\lim_{t\to\infty} y(t) = \overline{y} = \frac{c}{\beta \delta_y} \max\{0, \delta_x + \nu\}$. Now suppose that, in turn, $u(t)$ satisfies equation ([sys:production.expo.no_h]c), which means that $v(t) = \lambda - \kappa y(t)$, and therefore $\nu = \lim_{t\to\infty} v(t) = \lambda - \kappa \overline{y}$. This gives an implicit equation for the rate $\nu$:

$$\nu = \lambda - \kappa \overline{y} = \lambda - \frac{\mu \kappa}{\beta \delta_y} \max\{0, \delta_x + \nu\}\,.$$

We now solve this equation.

Denote

$$F(\lambda) \;=\; \frac{\lambda \beta \delta_y - \mu \kappa \delta_x}{\beta \delta_y + \mu \kappa}\,.$$

Suppose first that $\lambda \leq \delta_x$. Then, since $\delta_x + F(\lambda) = (\delta_x + \lambda)\theta$ (where $\theta = \beta\delta_y/(\beta\delta_y + \mu\kappa)$), $\mu = F(\lambda)$ satisfies $\delta_x + \mu \geq 0$ and also, rewriting $\mu = F(\lambda)$, $\mu$ is the unique solution of with $\delta_x + \mu \geq 0$. There are no solutions with $\delta_x + \mu < 0$, because such a solution would have to satisfy $\mu = \lambda$, but $\delta_x + \lambda \geq 0$. Suppose instead that $\lambda > \delta_x$. Then $\mu = \lambda$ is the unique solution of with $\delta_x + \mu < 0$. There are no solutions with $\delta_x + \mu \geq 0$, because such a solution would have to satisfy $\mu = F(\lambda)$ and therefore have $\delta_x + \mu = \delta_x + F(\lambda) = (\delta_x + \lambda)\theta < 0$, a contradiction. In summary, when $\lambda \geq -\delta_x$, the unique solution of is $\mu = F(\lambda)$, and when $\lambda < -\delta_x$ it is $\mu = \lambda$.

Note that when

$$\mu \delta_x \kappa \;>\; \beta \delta_y \lambda$$

(which happens automatically when $\lambda < 0$) the formula $\nu = F(\lambda)$ gives that $\nu < 0$, that is, $u(t) \to 0$ as $t \to +\infty$. Conversely, if $\mu\delta_x\kappa < \beta\delta_y\lambda$, then $\nu > 0$ and so $u(t) \to \infty$ as $t \to +\infty$. Qualitatively, this makes sense: a large feedback gain $\kappa$, or a small growth rate $\lambda$ in the absence of feedback, leads to the asymptotic vanishing of the $u$ variable.

In addition, from the formula $\overline{y} = \frac{\mu}{\beta\delta_y}\max\{0, \delta_x + v\}$ we conclude the following piecewise linear formula for the dependence of the limit of the output on the parameter $\lambda$ that gives the growth rate of exponentially growing $u$ when there is no feedback:

$$\overline{y} \;=\; \begin{cases} 0 & \text{if } \lambda < -\delta_x \\ \dfrac{\mu(\delta_x + \lambda)}{\beta\delta_y + \mu\kappa} & \text{if } \lambda \geq -\delta_x \end{cases}$$

These considerations provide helpful intuition about the closed-loop system, but they do not prove that is necessary and sufficient for stability, nor do they show the validity of for the closed-loop system. The reason that the argument is incomplete is that there is no *a priori* reason for $u(t)$ to have the exponential form $u(t) = C_0 e^{vt}$. We next provide a rigorous argument.

## Analysis of the closed-loop system

**Theorem.** Suppose that $(x(t), y(t), u(t)$ is a (positive) solution of ([sys:production.expo.no_h]abc), and define

$$v(t) := \; \dot{u}(t)/u(t) = \lambda - \kappa y(t)\,,$$

$$p(t) := \; u(t)/x(t)\,,$$

$\overline{y}$ by formula , which we repeat here:

$$\overline{y} \;=\; \begin{cases} 0 & \text{if } \delta_x + \lambda < 0 \\ \dfrac{\mu(\delta_x + \lambda)}{\beta\delta_y + \mu\kappa} & \text{if } \delta_x + \lambda \geq 0 \end{cases}$$

$\overline{p} = (\delta_y/\mu)\overline{y}$, and

$$\overline{v} \;=\; \begin{cases} \lambda & \text{if } \delta_x + \lambda < 0 \\ \lambda - \kappa\dfrac{\mu(\delta_x + \lambda)}{\beta\delta_y + \mu\kappa} & \text{if } \delta_x + \lambda \geq 0\,. \end{cases}$$

Then:

$$\begin{aligned} \lim_{t\to\infty} y(t) &= \overline{y} \\ \lim_{t\to\infty} p(t) &= \overline{p} \\ \lim_{t\to\infty} v(t) &= \overline{v}\,. \end{aligned}$$

and

$$\lim_{t\to\infty} u(t) \;=\; \begin{cases} 0 & \text{if } \delta_x\mu\kappa > \beta\delta_y\lambda \\ \infty & \delta_x\mu\kappa < \beta\delta_y\lambda \end{cases}$$

*Proof.* Substituting $v(t) = \lambda - \kappa y(t)$ into , we have the surprising and very useful fact that there is a closed system of just two differential equations for $p$ and $y$:

[sys:py]

$$\begin{aligned} \dot{p} &= p(\delta_x + \lambda - \kappa y - \beta p) \\ \dot{y} &= \mu p - \delta_y y\,. \end{aligned}$$

(This system could be viewed as a non-standard predator-prey of system, where $y$ behaves as a predator and $p$ as a prey.) In all of the real plane, there are two equilibria of this system, one at $p = y = 0$ and the other at $p = \frac{\delta_y(\delta_x+\lambda)}{\beta\delta_y+c\kappa}$, $y = \frac{c(\delta_x+\lambda)}{\beta\delta_y+c\kappa}$. The second equilibrium point is in the interior of first quadrant if and only if $\delta_x + \lambda > 0$.

We start by evaluating the Jacobian matrix of the linearized system. This is:

$$J = \begin{pmatrix} \delta_x + \lambda - \kappa y - 2\beta p & -p\kappa \\ \mu & -\delta_y \end{pmatrix}$$

which, when evaluated at $p = y = 0$, has determinant $-\delta_y(\delta_x + \lambda)$ and trace $\delta_x + \lambda - \delta_y$, and when evaluated at $(\overline{p}, \overline{y})$ has trace

$$\frac{-\beta\delta_y(\delta_x + \lambda)}{c\kappa + \beta\delta_y} - \delta_y$$

and determinant $\delta_y(\delta_x + \lambda)$. Thus, when $\delta_x + \lambda > 0$, the trace is negative and the determinant is positive, so the equilibrium $(\overline{p}, \overline{y})$ is stable, and (0,0) is a saddle because the determinant of the Jacobian is negative at that point. When instead $\delta_x + \lambda \leq 0$, the only equilibrium with non-negative coordinates is (0,0), and the determinant of the Jacobian is positive there, while the trace is negative, so this equilibrium is stable.

We note that, in general, if have shown that there is a limit $v(t) \to \overline{v}$ as $t \to \infty$ then $u(t) \to 0$ as $t \to \infty$ if $\overline{v} < 0$ and $u(t) \to \infty$ as $t \to \infty$ if $\overline{v} > 0$ Indeed, in the first case there is some $T \geq 0$ so that for $t \geq T$, $v = \dot{u}/u < \overline{v}/2$, meaning that $d(e^{-\overline{v}t/2}u(t))/dt \leq 0$, and hence $e^{-\overline{v}t/2}u(t) \leq e^{-\overline{v}T/2}u(T)$, so $u(t) \leq e^{\overline{v}(t-T)/2}u(T) \to 0$ (since $\overline{v} < 0$). Similarly, in the second case we use that there is some $T \geq 0$ so that for $t \geq T$, $v = \dot{u}/u > \overline{v}/2$, meaning that $d(e^{-\overline{v}t/2}u(t))/dt \geq 0$, and hence $e^{-\overline{v}t/2}u(t) \geq e^{-\overline{v}T/2}u(T)$, so $u(t) \geq e^{\overline{v}(t-T)/2}u(T) \to \infty$ (since $\overline{v} > 0$).

Consider first the case $\delta_x + \lambda \leq 0$. Then $\dot{p} = p(\delta_x + \lambda - \kappa y - \beta p) \leq p(-\kappa y - \beta p) < 0$ for all $p > 0$, and therefore $p(t) \to \overline{p} = 0$ as $t \to \infty$. We may now view the linear system $\dot{y} = \mu p - \delta_y y$ as a one-dimensional system with input $p(t) \to 0$, which implies that also $y(t) \to \overline{y} = 0$. In turn, this implies that $v = \lambda - \kappa y \to \overline{v} = \lambda < 0$. By the general fact proved earlier about limits for $u(t)$, we know that $u(t) \to 0$ as $t \to \infty$. This completes the proof when $\delta_x + \lambda \leq 0$.

So we assume from now on that $\delta_x + \lambda > 0$. We will show that, in this case, all solutions with $p(t) > 0$ and $y(t) > 0$ globally converge to the unique equilibrium $(\overline{p}, \overline{y})$. Once that this is proved, it will follow that $v(t) \to \overline{v} = \lambda - \kappa\overline{y}$. Now, this value of $\overline{v}$, for $\overline{y}$ picked as in

(case $\delta_x + \lambda \geq 0$), coincides with $\nu = F(\lambda) = \frac{\lambda\beta\delta_y - \mu\kappa\delta_x}{\beta\delta_y + \mu\kappa}$. So $\overline{v} < 0$ if $\mu\kappa\delta_x > \lambda\beta\delta_y$ and $\overline{v} > 0$ if $\lambda\beta\delta_y > \mu\kappa\delta_x$, and this provides the limit statement for $u(t)$, completing the proof.

We next show global convergence. A sketch of nullclines (see Figure S1 for a numerical example) makes convergence clear, and helps guide the proof. Consider any $P \geq (\delta_x + \lambda)/\beta$ and any $Y \geq \mu P/\delta_y$ and the rectangle $[0, P] \times [0, Y]$.

On the sides of this rectangle, the following properties hold:

- On the set $\{0\} \times (0, Y)$, $\dot{p} \geq 0$, because $\dot{p} = 0$.
- On the set $\{P\} \times (0, Y)$, $\dot{p} \leq 0$, because $\dot{p} = p(\delta_x + \lambda - \beta P) \leq 0$, by the choice of $P$.
- On the set $(0, P) \times \{0\}$, $\dot{y} \geq 0$, because $\dot{y} = \mu p > 0$.
- On the set $(0, P) \times \{Y\}$, $\dot{y} \leq 0$. because $\dot{y} = \mu p - \delta_y Y \leq \mu P - \delta_y Y \leq 0$ by the choice of $Y$.
- At the corner point (0,0), $\dot{p} \geq 0, \dot{y} \geq 0$, because $\dot{p} = \dot{y} = 0$.
- At the corner point $(0, Y)$, $\dot{p} \geq 0, \dot{y} \leq 0$, because $\dot{p} = 0, \dot{y} = -\delta_y Y < 0$.
- At the corner point $(P, 0)$, $\dot{p} \leq 0, \dot{y} \geq 0$, because $\dot{p} = p(\delta_x + \lambda - \beta P) \leq 0, \dot{y} = \mu P > 0$.
- At the corner point $(P, Y)$, $\dot{p} \leq 0, \dot{y} \leq 0$, because $\dot{p} = p(\delta_x + \lambda - \beta P - \kappa Y) < p(\delta_x + \lambda - \beta P) \leq 0, \dot{y} = \mu P - \kappa y \geq 0$.

These properties imply that the vector field points inside the set at every boundary point and therefore it is forward-invariant, meaning that every trajectory that starts in this set remains there for all positive times (Clarke et al. 1998). The rest of the proof of stability uses the Poincaré-Bendixson Theorem together with the Dulac-Bendixson criterion. Note that, for any initial condition $\xi = (p(0), y(0))$ one can always pick a large enough value of $P$ and $Y$ so that $(p(0), y(0)) \in [0, P] \times [0, Y]$. The invariance property guarantees that the omega limit set $\omega^+(\xi)$ is a nonempty compact connected set, and the Poincaré-Bendixson Theorem insures that such a set is one of the following: (a) the equilibrium (0,0), (b) a periodic orbit in the interior of the square, or (c) the equilibrium $(\overline{p}, \overline{y})$ (Hirsch and Smale 1974). Note that a homoclinic orbit around (0,0) cannot exist, because the unstable manifold of this equilibrium is the entire $y$ axis. For the same reason, if $\xi$ has positive coordinates, $\omega^+(\xi) \neq (0,0)$. Therefore, all that we need to do is rule out periodic orbits. Consider the function $\varphi(p, y) = 1/p$. The divergence of the vector field

$$\begin{pmatrix} \frac{1}{p}(p(\delta_x + \lambda - \kappa y - \beta p)) \\ \frac{1}{p}(\mu p - \delta_y y) \end{pmatrix} = \begin{pmatrix} \delta_x + \lambda - \kappa y - \beta p \\ \mu - \delta_y y/p \end{pmatrix}$$

is

$$\frac{\partial \delta_x + \lambda - \kappa y - \beta p}{\partial p} + \frac{\partial \mu - \delta_y y}{\partial y} = -\beta - \delta_y / p \,,$$

which has a constant sign (negative). The Dulac-Bendixson criterion (Hirsch and Smale 1974) then guarantees that no periodic orbits can exist, and the proof is complete.

## C. Perfect adaptation and scale-invariance

A system is said to be *perfectly adapting* provided that its response returns asymptotically to a pre-stimulus value under constant stimulation. This property is typically exhibited by sensory systems processing light, chemical, and other signals, and it has been extensively investigated both experimentally and mathematically (Alon 2006; Keener and Sneyd 2009). In particular, when subjecting a perfectly adapting system to a step-wise input signal, as shown in Figure S2A , the output of the system settles, after a transient response, to a basal value which does not depend on the magnitude of the stimulus. The response amplitude and timing, on the other hand, typically depends on the input magnitude. This notion can be refined as follows. Suppose that every step has the same relative or "fold" change, $u_{i+1}/u_i =$ constant, as shown in the figure. For *scale-invariant* systems, the responses to such steps have the exact same shape, amplitude, and duration.

The alternative term "fold change detection" is sometimes used for this property, to emphasize the fact that such systems can only react differently if the fold changes are not the same. To put it in another way, such systems can give different responses if difference $\log u_{i+1} - \log u_i$ is nonzero (log sensing) as opposed to $u_{i+1} - u_i$. The precise mathematical definition of scale-invariance involves arbitrary input signals: responses to arbitrary scaled inputs as in Figure S2B, and not only piecewise constant ones, should be the same, provided that the internal state starts from a preadapted value. We refer the reader to (Shoval, Alon, and Sontag 2011) for technical details.

Scale invariance or fold change detection (FCD) is a strengthening of the Weber-Fechner "log sensing" property, which is sometimes defined as the requirement that the maximum amplitude of responses to two scaled inputs should be the same, but not necessarily their exact shape or even timing. Recent interest in the FCD property was largely triggered by the papers (Goentoro and Kirschner 2009; Cohen-Saidon et al. 2009), in which fold-change detection behavior was experimentally observed in a Wnt signaling pathway and an EGF pathway, respectively; these are highly conserved eukaryotic signaling pathways that play roles in embryonic patterning, stem cell homeostasis, cell division, and other central processes. Later, the paper (Shoval et al. 2010) predicted scale invariant behavior in *E. coli* chemotaxis, a prediction which was subsequently experimentally verified (Lazova et al. 2011). Similar results are available for other bacterial species, for example *R. sphaeroides*, for which theoretical predictions made in (Hamadeh, Ingalls, and Sontag 2013) were experimentally confirmed in (Wadhams and Armitage 2004). A mathematical study of scale invariance, together with a necessary and sufficient characterization in terms of solutions of a partial differential equation, can be found in (Shoval, Alon, and Sontag 2011). It has been recently shown that all scale invariant systems compute a certain type of differentiation operator, such as logarithmic derivatives (Lang and Sontag 2016).

One example of a scale invariant system is the IFFL that underlies our model, which we repeat here for ease of reference:

$$\begin{aligned}\dot{x} &= -\delta_x x + \beta u \\ \dot{y} &= \mu\frac{u}{x} - \delta_y y\end{aligned}$$

where $\beta, \mu, \delta_x, \delta_y$, are some positive constants and $u(t)$ is viewed as an external stimulus. For any given input function $u(t)$ and initial values $x(0)$ and $y(0)$, the solution of this system can be found by first solving the scalar linear ordinary differential equation for $x(t)$, and then plugging this result together with $u(t)$ into the $y$ equation, which is also a linear ODE. For a constant input $u(t) \equiv u_0 > 0$, there is a globally asymptotically stable steady state, given by

$$x = \frac{\beta u_0}{\delta_x}, \quad y = \frac{\delta_x \mu}{\beta \delta_y}.$$

At steady state, the output $y$ is independent of the particular value of the constant input $u_0$, meaning that the system is perfectly adapting. Suppose next that $(x(t), y(t))$ is any solution of the system corresponding to an input $u(t)$, now not necessarily a step function. It is then immediate to verify that $(px(t), y(t))$ is a solution corresponding to the input $pu(t)$, $t \geq 0$, for any nonzero constant scaling factor $p$:

$$\begin{aligned}\dot{x} &= -\delta_x x + \beta u \\ \dot{y} &= \mu\frac{u}{x} - \delta_y y\end{aligned}$$

implies

$$\begin{aligned}(\dot{px}) &= -\delta_x(px) + \beta(pu) \\ \dot{y} &= \mu\frac{pu}{px} - \delta_y y = \mu\frac{u}{x} - \delta_y y\,.\end{aligned}$$

Thus, this system responds with the same output signal $y(t)$ to two inputs which differ only in scale, provided that the initial state $x(t)$ had already adapted to the input at time $t < 0$. In other words, given a step input that jumps from $u(t) = u_0$ for $t < 0$ at time $t = 0$ and an initial state at time $t = 0$ that has been pre-adapted to the input $u(t)$ for $t < 0$, $x(0) = \beta u_0/\delta_x$, the solution is the same as if, instead, the input would have been $pu(t)$ for $t > 0$, but starting from the respective pre-adapted state $p\beta u_0/\delta_x$. This means that our IFFL subsystem is scale-invariant.

It would be very interesting to test experimentally the response to scaled versions of antigen presentation, to verify if such scale invariance holds, even in an approximate fashion.

## D. Details on the model used for simulations

In this section, we explain the terms in the differential equations used in simulations, including the parameters used. Of course, our model is only a cartoon of a hugely

complicated system of interlocking processes. Moreover, even if the model were mechanistic, which it is not, numbers would depend on the specific tumor or infection tissue being modeled. Thus, these algebraic forms and numbers are offered only as a plausible scenario.

As explained in the main text, $u$ represents an immune challenge, specifically a tumor in this case, while $x$ and $y$ might represent populations of activated and specific T suppressor (CD4 $^+$ CD25 $^+$ Treg) and cytotoxic T cells (CD8 $^+$ cells) respectively. We use as a guide in our modeling the paper by Kirschner and Panetta (Kirschner and Panetta 1998), which has become a classic reference for tumor-immune interactions in the presence of cytokines (no regulatory T cells in that model), together with the more recent paper by Khailaie et al. (Khailaie et al. 2013) which described a model of immune activation in the presence of both chemokines and also regulatory T cells (no tumor dynamics in that model).

Treg cells play a central role in cytotoxic T cell regulation. The various Treg mechanisms can be arranged into four groups centered around four basic modes of action (Vignali, A. A., and L. W. 2008): (1) inhibitory cytokines, including IL-10, IL-35) and TGF-$\beta$, (2) cytolysis through granzyme-A- and granzyme-B-dependent and perforin-dependent killing mechanisms, (3) metabolic disruption through CD25-dependent cytokine-deprivation-mediated apoptosis, cAMP-mediated inhibition, and adenosine–purinergic adenosine receptor (A2A)-mediated immunosuppression, and (4) targeting dendritic cells through mechanisms that modulate DC maturation and/or function.

## Cell number units

Since we use parameters from both (Kirschner and Panetta 1998) and (Khailaie et al. 2013), it is thus important to clarify the units used in these sources.

Kirschner and Panetta's paper gives "volume" as the unit for cell populations. Since many of these parameters were in turn obtained from the foundational paper by Kuznetsov et al. (Kuznetsov et al. 1994), which provided one of the first differential equation models for interactions between tumors and the immune system, one can compare the two papers, to map their unit to cell numbers. For this purpose, we can compare the value of the carrying capacity of tumors ("$B$" in the simulations that we provided) in both papers. In (Kirschner and Panetta 1998) $B = 10^{-9}$, and in (Kuznetsov et al. 1994) $B = 2 \times 10^{-9}$. Ignoring the factor of 2, this means that "volume" = number of cells. This is confirmed by comparing the Michaelis-Menten constant for IL-2 activation $g_1$ ($g$ in the second paper): $2 \times 10^{-7}$ volume units and $2.019 \times 10^{-7}$ T cells respectively. Therefore, we will be interpreting cell units in (Kirschner and Panetta 1998) as numbers of cells. In our simulations, we use $B = 10^{-3}$, because we prefer to switch to units of $10^6$ cells. In Khailaie et al.'s paper (personal communication from first author), "cell" means nondimensional units, cells/$C_0$, where $C_0$ is an unspecified reference quantity of cells. Now, Figure 5 in (Khailaie et al. 2013) shows stable branches of equilibria under antigen stimulation in ranges of 2 to 30 nondimensionalized T cells, while in their companion experimental paper (Milanez-Almeida et al. 2015), the same authors provide estimates of T cells in various tissues in mice in the range $10^6$ to $8 \times 10^6$. Thus approximately $C_0 = 10^6$ cells is consistent with the

analysis in (Khailaie et al. 2013), and so we will interpret the numbers in that reference in units of $10^6$ cells.

## The autocatalytic term $\frac{Vy^2}{K+y}$

This term is intended to model a cytokine-mediated positive feedback loop on effector T cells. Cytokines are molecules that act as immunomodulating agents and mediate communication among immune systems components and their environment. Their concentrations can increase up to 1,000-fold during inflammatory conditions. Examples of cytokines include interleukins such as IL-2 and IL-6, interferons, and TNF. The role of cytokines in anti-tumor responses, and in particular IL-2, has been the subject of much study (Dranoff 2004) and of mathematical modeling since at least the work of Kirschner and Panetta (Kirschner and Panetta 1998), who proposed a simple differential equation model that includes variables for tumor load, effector immune cells, and cytokines. In their model, activated T cells produce cytokines, specifically IL-2, which in turn enhance lymphocyte activation, growth and differentiation, in particular of the cytotoxic T cell (CTL) population. The effect is through a positive feedback that is both autocrine, that is, acting on the cells that produce it, and paracrine, acting on nearby cells. This role of IL-2 in enhancing T-cell proliferation and differentiation is one reason that IL-2 was originally named "T-cell growth factor," although by now many other immunoregulatory functions of IL-2 are known.

The term that represents the effect of the cytokine (IL-2) on $y$ in (Kirschner and Panetta 1998) is $p_1yz/(g_1+z)$, where the cytokine $z$ satisfies the differential equation $z' = p_2uy/(g_3+u) - \mu_3 z$. This equation models IL-2 secretion by activated effector T cells, with a Michaelis-Menten kinetics to account for self-limiting production of IL-2, together with a degradation rate. To obtain $z$ as a function of $y$, we assume that this variable is at equilibrium; on the saturation regime of antigen load $u$ we obtain $z = (p_2/\mu_3)y$. Now substituting this expression into the differential equation for $y$, we have the autocatalytic term

$$\frac{(p_1p_2/\mu_3)y^2}{g_1 + (p_2/\mu_3)y} = \frac{Vy^2}{K+y}$$

where $V = p_1$ and $K = \mu_3 g_1/p_2$. If we start, instead, from (Khailaie et al. 2013), the corresponding term in the differential equation for $\dot{y}$ is $ayz$, where $z$ now satisfies a different equation, $\dot{z} = dy - eyz - fz$ and the term $eyz$ represents IL-2 consumption rate by T cells. Nonetheless, under the same equilibrium assumptions we obtain $z = dy/(f+ey)$, which when substituted into $ayz$ gives

$$\frac{(ad/e)y^2}{(f/e)+y} = \frac{Vy^2}{K+y}$$

where $V = ad/e$ and $K = f/e$. In other words, we derived the same functional form as when starting from (Kirschner and Panetta 1998).

Plausible parameters values can be obtained from (Kirschner and Panetta 1998) or from (Khailaie et al. 2013). The parameters used in (Kirschner and Panetta 1998) were $p_1 = 0.1245$, $p_2 = 5$, $g_1 = 2 \times 10^7$, and $\mu_3$ was arbitrarily picked as 10 from the range 8.31 to 33.27 using a half-life for IL-2 of 30 to 120 minutes given in (Rosenberg and Lotze 1986). Plugging these into the formulas given above, we obtain $V = 0.1245$ and $K = 10^6 K_0$, where $K_0$ ranges from 33 to 133. As discussed earlier, we are reading the units in the paper (Kirschner and Panetta 1998) as individual cell counts. When translating to our units of $10^6$ cells, we obtain that $K$ in their model ranges between 33 and 133. (The argument is: if we rescale variables letting $\eta = y/10^6$, then the corresponding term in $\dot{\eta}$ is $10^{-6} V (10^6 \eta)^2 / (10^6 K_0 + 10^6 \eta) = V\eta^2/(K_0 + \eta)$, which means that $K = K_0$ when writing the equation in terms of $\eta$.) Using parameters from (Khailaie et al. 2013) gives similar results. As discussed earlier, we are reading the units in that paper as $10^6$ cells. These parameters are picked in (Khailaie et al. 2013) as follows: $a = 0.4$, $d = 0.01$, $e = 0.01$, $f = 1$. Plugging these into the formulas given above, these lead to $V = 0.4$ and $K = 100$. In summary, one paper gives $K$ between 33 and 133 and $V = 0.1245$, and the other paper uses $K = 100$ and $V = 0.4$. We therefore take $K = 100$ and for $V$ pick an average, $V = 0.25$ of the two values. Note that the units of $K$ are $10^6$ cells, and the units of $V$ are day $^{-1}$.

## The fratricide term $-\varepsilon y^2$

Following the T cell model in (Khailaie et al. 2013), we include the term $-\varepsilon y^2$ for cell-contact-dependent activation-induced cell death in activated T cells, a process known as "fratricide". Activated T cells express the receptor FasR, also known as apoptosis antigen 1 (APO-1 or APT), cluster of differentiation 95 (CD95) or tumor necrosis factor receptor superfamily member 6 (TNFRSF6), as well as the ligand for this molecule, FasL; fratricide can result from direct cell contact or from cleavage of FasL ("death ligand"), and the ligation of FasR by soluble FasL results in apoptotic cell death, mediated by caspase activation (Flaherty 2011). It is believed that the exposure to tumor antigens in T cells might mediate fratricide (Leisegang et al. 2010). Callard, Stark, and Yates (Callard, Stark, and Yates 2003) modeled the fratricide mechanism by a nonlinear death term $-\varepsilon y^2$ and speculate that Fas-mediated apoptosis results in a density-dependent death rate for T cell homeostasis that does not require competition for resources nor quorum-sensing mechanisms for density estimation. From (Khailaie et al. 2013), we pick $\varepsilon = 10^{-5}$, in units of day $^{-1}$ ($10^6$ cells) $^{-1}$.

## The decay terms $-\delta_x x$ and $\delta_y y$

These represent linear degradation of activated T and Treg cells. The values $\delta_x x = \delta_y y = 0.1$ are from (Khailaie et al. 2013). Units of both are day $^{-1}$.

## The term $\beta u$

Stimulation of regulatory cells is a very complex process that involves a wide variety of antigen presenting cells and other mediators. TRegs are exported from the thymus and recirculate through secondary lymphoid tissues as "central" TReg cells, and get activated through T cell receptor (TCR) ligation, CD28 co-stimulation and/or interleukin-2 (IL-2), which induce upregulation of expression of interferon regulatory factor 4 (IRF4), which

then orchestrates their differentiation into "effector" TReg cells (Liston and Gray 2014). We make the simplest possible assumption: the rate of activation is proportional to the immune challenge such as a tumor population, that is, we postulate a term $\beta u$ in $\dot{x}$. It is virtually impossible to give a numerical value for the parameter $\beta$, since this value depends on the nature of the immune challenge, spatial relations between antigen presenting cells and T cells, and so forth. Khailaie et al. (Khailaie et al. 2013) simply use a term $+k(t)$ to represent this stimulation (where $k(t)$ is the product of antigen stimulation "$\beta$" and the supply $N$ of naive T cells or resting Treg cells, and introducing an unspecified multiplier to model possibly different effects on T cells compared to Tregs). This additive input is naturally modeled by $\beta = 1$, and we take that simplest possible value. Units are day $^{-1}$.

## The term $\mu u/x$

There are various ways to justify this term. We picked a mathematical form for the effect of the immune challenge $u$ and regulatory elements $x$ on effector cells $y$ that is the simplest possible to model activation by $u$ and repression by $x$. Let us discuss why this choice is reasonable phenomenologically. The term "regulatory T cell" (Treg) actually encompasses several subclasses of cells that help in peripheral tolerance, preventing autoimmune diseases, and down-modulating immune responses. These cells they affect many other immune components, from B cells to helper cells (Th1, Th2, Th17) and cytotoxic T cells, through both direct and indirect interactions. These interactions form an extremely complicated and poorly understood network that includes inhibitory molecules such as CTLA4 and messaging by cytokines (TGF-$\beta$, IL-10, IL-35, and others) which result in the suppression of helper cell differentiation and in indirect down-regulation of MHC and costimulatory molecules on antigen-presenting cells, thereby reducing T cell activation. The repression of T cell activation through TCR-MHC is one way to see the negative effect of $x$ on $y$. Another is the indirect effect through inhibitory cytokines such as IL-10, TGF-$\beta$, and IL-35 that can suppress T cell activation. The simplest mass-action kinetics model would assume independent effects: activation by $u$ and repression by $x$, leading to a term of the form $h_1(u)h_2(x)$ driving $y$ activation. For the effect of $u$, let us take $h_1(u) = c_1 u$, for some constant $c_1$. If we assume that $x$ cells (or messenger molecules) repress through binding to a certain type of receptor, and $R$ represents the number (or fraction, or concentration, depending on units) of free receptors, then at equilibrium we would have $kRx = R_0$, where $R_0$ quantifies occupied receptors, and from a conservation $R + R_0 = R_T$ assuming a constant total number of receptors, we would have that $R = R_T/(1 + kx)$ is the number of free (unbound) receptors, so unless $k \ll 1$ we may take $h_2(x) = c_2/x$ for some constant $c_2$. These arguments will result in the algebraic form $h(x, u) = Mu/x$. A different justification is as follows. Let us assume that there is an intermediate variable $z$, which might represent for example a population of helper T cells (Th cells or CD4$^+$ T cells) which helps activate the cytotoxic T population $y$ and is itself activated by the immune challenge $u$ and inactivated by the regulatory variable $x$. The simplest equation would be $\dot{z} = -\mu_0 xz + \beta u$, where we are assuming that helper cells are also being activated in a manner proportional to the magnitude of the immune challenge, and $\mu_0 x$ represents the $x$-dependent degradation of $z$. We assume that $\dot{y}$ has a term $z$ corresponding to activation by helper cells. Assuming that this equation is at equilibrium, we may substitute $z = (\beta/\mu_0)u/x$ into the $\dot{y}$ equation, giving a term $\mu u/x$, where $\mu = \beta/\mu_0$. (If helper and T cell

activations are at similar timescales and the equilibrium assumption is not made, one add may the $z$ differential equation explicitly. We prefer to keep the model simpler, but see Supplement Section F for simulations using that model.) Khailaie et al. (Khailaie et al. 2013), include in T cell dynamics a similar mass-action degradation or inactivation term, using a rate constant 0.1. Following this, we pick the value $\mu_0 = 0.1$, so that, together with $\beta = 1$ we have $\mu = 10$. As $u$ and $x$ are both in units of $10^6$ cells, $\mu$ has units day $^{-1}$.

### The terms $\lambda u$ and $-\kappa y u$

The term $\lambda u$ is a standard exponential growth term. We view $\lambda$ as a varying parameter, which quantifies the initial exponential growth of the immune challenge.

The killing term $-\kappa u y$ in the $\dot{u}$ equation represents a simple mass-action suppression of the immune challenge, such as cytotoxic T cells killing tumor cells. The constant $\kappa$ depends on many factors, such as the type of tumor, size and geometry of tumor microenvironment, accessibility of tumor cells to vasculature, and so forth. In the original paper by Kuznetsov et al. (Kuznetsov et al. 1994), one finds $\kappa = 1.101 \times 10^{-7}$ in units of day $^{-1}$ cells $^{-1}$ which when normalized to units of $10^6$ cells would give the value $\kappa = 1.101 \times 10^{-1}$, This value seems to be too large for most cancers. For example, based on fits to experimental data, the recent paper (Wang, Klinke, and Wang 2015) obtains a number which is many orders of magnitude smaller. That paper analyzes the killing by cytotoxic CD8+ T cells of MHCI $^{+}$ tumor cells in a B16 mouse metastatic melanoma model, and determines a killing term for such cells of the following form (with different notations here): $-[c/(\varepsilon + U)]Yu$, where $Y$ is the concentration of effector CD8 $^{+}$ T cells in the tumor microenvironment, using units of cells/mm $^{3}$, $c$ is a constant that quantifies MHCI positive tumor death rate due to T effectors, and has the value $2.49 \times 10^{-13}$ in units $mm^3$ day $^{-1}$, $U$ is the total number of tumor cells, $\varepsilon$ is a "small number" to account for other cells, and $u$ is the number of major histocompatibility complex class I positive tumor cells. Since $[c/(\varepsilon + U)]Y$ has units day $^{-1}$, if we convert to $y$ in units of $10^6$ cells, we obtain $cY = \kappa y$ where $k = 2.49 \times 10^{-7}/(\varepsilon + U)$ has units ($10^6$ cells) $^{-1}$ day $^{-1}$. Depending on the number of cells $U$ in the tumor, this number $\kappa$ could be very small, and it is certainly less than $2.49 \times 10^{-7}$. To take another example, Kirschner and Panetta (Kirschner and Panetta 1998) employ a Michaelis-Menten killing term $-auy/(g_2 + y)$, with $a = 1$ and $g_2 = 10^5$. Given these wide ranges, we pick $\kappa = 10^{-5}$ for our simulations. Units are ($10^6$ cells) $^{-1}$ day $^{-1}$. (A two-zone behavior of tumor elimination can also be found with $\kappa = 10^{-4}$, $\kappa = 10^{-3}$, $\kappa = 10^{-2}$, and $\kappa = 10^{-1}$, but shifting the range of $\lambda$'s at which different behaviors arise.)

## Sensitivity to parameters in the function $f$

We recall the definition of the function $f$:

$$f(y) = \frac{Vy^2}{K + y} - \varepsilon y^2 - \delta_y y\,.$$

The main requirement for the theoretical analysis in the main text is that $f$ have a cubic form as illustrated in Figure 11, so that then the nullcline analysis in Figure 2 applies. In other words, $f$ should have one zero at $\eta_1 = 0$ and two positive zeros $\eta_2, \eta_3$ so that

$f(y) < 0$ for $\eta_1 < y < \eta_2$, $f(y) > 0$ for $\eta_2 < y < \eta_3$, and $f(y) < 0$ for $\eta_3 < y$. (Observe that signs gets reversed in the nullclines in Figure 2, because of the negative sign in the formula $p = h^{-1}(-f(y))$.) Writing $-f(y) = \frac{y}{K+y} g(y)$, where

$$g(y) = \varepsilon y^2 + (\delta_y - V + K\varepsilon)y + K\delta_y$$

and using that $\frac{y}{K+y}$ is positive for $y > 0$ and zero at $y = 0$, the requirements on $f$ translate into the requirement that the parabola $g(y)$ have two positive zeros $\eta_2, \eta_3$ (and be negative in between them), which is equivalent to:

$$(\delta_y - V + K\varepsilon)^2 > 4\varepsilon\delta_y K \quad \text{and} \quad \delta_y - V + K\varepsilon < 0$$

For our parameters, $V = 0.25$, $K = 100$, $\delta_y = 0.1$, $\varepsilon = 10^{-5}$, we have $\delta_y - V + K\varepsilon \approx -0.1490$, $(\delta_y - V + K\varepsilon)^2 \approx 0.0222$, and $4\varepsilon\delta_y K = 4 \times 10^{-4}$, so that these conditions are satisfied. These requirements imply that the maximal autocatalytic strength $V$ should be large, and the degradation constant $\delta_y$ and the fratricide constant $\varepsilon$ should be small.

## E. Nullclines for model and parameters used in text

Figure S3 shows the nullclines for this system for various increasing values of $\lambda$, as well as some typical solution trajectories, showing their convergence to values under, over, under, and finally again over the threshold which determines tolerance or rejection of the immune challenge. This is perfectly consistent with our theoretical predictions.

## F. A model with an intermediate population

We consider here that a slightly different model, in which $u$ and $x$ affect the effector variable $y$ only indirectly, through production and repression respectively of a "helper cell" population. Supplemental Figure 9 plots simulation results (all parameters exactly the same as in earlier model), showing that this model leads to similar results as those for the simpler model.

$$\begin{aligned}
\dot{u} &= [\lambda(1 - Bu) - \kappa y]u \\
\dot{x} &= -\delta_x x + \beta u \\
\dot{y} &= z + \frac{Vy^2}{K + y} - \varepsilon y^2 - \delta_y y \\
\dot{z} &= u - (1/\mu)\, xz
\end{aligned}$$

## G. More details on exponential rate detection and two-zone experimental results

In our model, an embedded IFFL acts as an estimator the rate of exponential increase of the immune challenge. I briefly mentioned the work of (Johansen et al. 2008). Let me discuss here some more relations to that work. The authors state that "antigenic stimulation increasing exponentially over days was a stronger stimulus for CD8 T cells and antiviral immunity than a single dose or multiple dosing with daily equal doses" and concluded that "at a clonal level, T cells are capable of decoding the kinetics of antigen exposure." They found that IL-2 activation at constant dosage of antigen is almost zero, at linearly increasing dose is higher, and at exponential doses is highest, and concluded (Figure 7, caption) that "exponential in vitro stimulation of CD8 T cells enhances IL-2 production and cytotoxicity." These experimental observations are all roughly consistent with activation of the autocatalytic loop in our model under higher exponential rates. In 2008, Kündig and collaborators, based on this work, obtained a patent (Kundig et al. 2008) for "A method for enhancing T cell response" based on the principle that immunogenicity is enhanced by "exponentially increasing antigenic stimulation of class I MHC CD8+ T cell response …in a manner independent of the dose of the antigen."

Another conclusion of the analysis is the existence of intermediate regions of challenge (e.g., tumor) growth in which the challenge will be eliminated by the immune system, with challenges in lower as well as in larger regions not being eliminated. This existence of disjoint regions of tumor elimination depending on rate of growth is strongly reminiscent of two related phenomena, "sneaking through" and "two-zone" tumor tolerance, which have been much discussed since the mid-1960s. The idea of tumors "sneaking through" from immune control originated with the findings in (Klein 1966) of intermediate regions in which tumors can be eliminated. Further, Gatenby, Basten, and Creswick in the early 1980s (Gatenby, Basten, and Creswick 1981) argued that this four-region phenomenon specifically depends on T-cell repression (just as in our model through the regulatory $x$ variable), and framed this role of suppressor T cells on regulating tumor immune response in the more general idea of low zone tolerance (tolerance to antigens under repeated exposure to small antigen doses). This work was, in turn, motivated by seminal work by Kolsch and coworkers (Haubeck and Kolsch 1982), who injected exponentially increasing numbers of irradiated syngeneic ADJ-PC-5 plasmacytoma cells into BALB/c mice, starting with 2 cells at day 1, 4 at day two, and doubling subsequent doses for 15 days until about $10^5$ were received, and proposed the induction of T suppressor cells (what one now calls Treg cells) as an early event in tumorigenesis that regulated CTL activity. To test their ideas, Gatenby et al. carried out experiments that show sneaking-through behavior as well as the failure of this behavior when "suppressor T cells" are eliminated, see Figure 4A and Figure 4B respectively. Murine sarcoma Meth A was administered in varying doses to BALB/c mice, and incidence of tumors was measured in each group of 12-42 mice, at two weeks after the last mouse died from tumor. Similar results on sneaking-through had been reported by Kolsch and Mengersen in previous work in which mastocytoma BM3 injected cells were injected into BALB/c mice, see Figure 4C. Care must be taken when interpreting

these experimental numbers in terms of a model. The numbers reported are for "tumor incidence," meaning percentages of mice in which tumors were detected by some predetermined point. If we assume that survival (until mouse sacrifice, or indirect death from the tumor) depends probabilistically on tumor size, then we could think of tumor incidence as a proxy for size. Another difference is that, in these works, the different regions correspond to the magnitude of an initial tumor inocula in animal subjects, rather than growth rates. Nonetheless, there is a surprisingly strong resemblance between our plots and the experimental ones. This picture is at least consistent with a larger initial rate of increase in exposure leading to tumor suppression, as in our model.