PNAS RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY
IMMUNOLOGY AND INFLAMMATION

An inference model gives insights into innate immune adaptation and repertoire diversity

Yawei Qin[a], Emily M. Mace[b], and John P. Barton[a,c,1]

Edited by Marco Colonna, Washington University in St Louis School of Medicine, St. Louis, MO; received April 12, 2023; accepted August 8, 2023

The innate immune system is the body's first line of defense against infection. Natural killer (NK) cells, a vital part of the innate immune system, help to control infection and eliminate cancer. Studies have identified a vast array of receptors that NK cells use to discriminate between healthy and unhealthy cells. However, at present, it is difficult to explain how NK cells will respond to novel stimuli in different environments. In addition, the expression of different receptors on individual NK cells is highly stochastic, but the reason for these variegated expression patterns is unclear. Here, we studied the recognition of unhealthy target cells as an inference problem, where NK cells must distinguish between healthy targets with normal variability in ligand expression and ones that are clear "outliers." Our mathematical model fits well with experimental data, including NK cells' adaptation to changing environments and responses to different target cells. Furthermore, we find that stochastic, "sparse" receptor expression profiles are best able to detect a variety of possible threats, in agreement with experimental studies of the NK cell repertoire. While our study was specifically motivated by NK cells, our model is general and could also apply more broadly to explain principles of target recognition for other immune cell types.

innate immunity | statistical inference | biophysics | natural killer cells

Significance

Innate immune cells such as natural killer (NK) cells can detect and respond to unhealthy target cells with aberrant ligand expression. How do they balance self-tolerance with the ability to respond to threats? Here, we describe a model in which immune cells adapt to the distribution of ligands on healthy target cells, which allows them to identify and respond to abnormal targets. Our model recapitulates NK cell behaviors, quantitatively fits experimental data, and explains the utility of experimentally observed sparse receptor expression profiles in the NK cell repertoire. This model provides a basis to predict the immune responses to novel stimuli and could apply more broadly to other cell types.

Natural killer (NK) cells are an important component of the innate immune system, responsible for recognizing and destroying cells that are foreign, infected by intracellular pathogens, or cancerous (1–7). NK cells were originally noted for their ability to kill target cells that express low or undetectable levels of major histocompatibility (MHC) class I (8). Subsequent investigations demonstrated that NK cells integrate signals through a broad array of activating and inhibitory receptors, many of which bind to self-ligands (4, 9–11).

NK cell activation is a complex and context-dependent process. While NK cells from normal hosts will typically kill target cells with low MHC class I expression, NK cells from mice and humans that constitutively express low levels of MHC class I are self-tolerant (12–16). These observations point toward an adaptive process, termed education, through which the threshold for activation is controlled by the local environment (17, 18). Qualitative models of education have been developed, positing that stronger inhibitory signaling is associated with greater NK cell functionality (19–21). However, even this picture is complicated by findings that NK cells that completely lack inhibitory receptors can play an important role in controlling infection (22) and fighting cancer (23).

Here, we propose a simple framework to understand how NK cells can achieve self-tolerance while maintaining the ability to respond to unhealthy targets. Ultimately, NK cells must use input from their receptors to discriminate between normal and unhealthy target cells. We therefore frame the recognition of unhealthy target cells as an inference problem akin to anomaly detection in computer science. We then develop a simple model to solve this inference problem, which is informed by prior experimental results. In our model, immune cells adaptively learn the typical distribution of ligands on healthy cells through repeated encounters, allowing them to respond to significant deviations from typical ligand expression in rare, unhealthy targets. We show that our model is consistent with known NK cell behaviors and quantitatively fits experimental data.

We also extend our model to explore the consequences of multiple pairs of receptors and ligands, which are known to be important for NK cells (9) and other immune cells (24). Intriguingly, recent studies have revealed dramatic heterogeneity in the complement of receptors that individual NK cells express (25, 26). The function of this heterogeneity, however, has remained unclear. We show that populations of immune cells with "sparse" patterns of receptor expression, like those observed in experimental settings, are better able to respond to a variety of different targets than ones with "dense" receptor expression profiles. Here, we emphasize that sparsity refers to the heterogeneous pattern

Author affiliations: [a]Department of Physics and Astronomy, University of California, Riverside, CA 92521; [b]Department of Pediatrics, Columbia University Irving Medical Center, New York, NY 10032; and [c]Department of Computational and Systems Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15260

Author contributions: Y.Q. and J.P.B. designed research; Y.Q. and J.P.B. performed research; Y.Q., E.M.M., and J.P.B. analyzed data; and Y.Q., E.M.M., and J.P.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: jpbarton@pitt.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2305859120/-/DCSupplemental.

Published September 11, 2023.

of receptor expression across individual NK cells, where particular types of receptors are expressed on only a fraction of cells, rather than the density of receptors on the cell surface. Sparse receptor expression is especially important for separating a signal of aberrant ligand expression due to infection or malignant transformation from noise due to normal variation in ligand concentrations. This suggests that heterogeneity in the NK cell repertoire may improve the immune system's ability to respond to multiple threats.

An Inference Model for Target Cell Recognition

Immune cells respond to signals from their receptors that can be either activating or inhibitory. The strength of the signals that they receive depends on both the concentration of receptors on the immune cell surface and the concentration of the corresponding ligands on the target cell surface. For simplicity, we represent the net signal that an individual immune cell receives from a target cell with a single variable x. Larger values of x represent more activating signals, while smaller values represent more inhibitory signals.

To prevent autoimmunity, immune cell responses must be tuned to avoid activation against normal, healthy cells. This could be achieved by the immune cell by tracking the distribution of signals $P_r(x)$ that it receives during target cell encounters (Fig. 1). Under normal conditions, the great majority of target cells should be healthy. Signals that are substantially more activating than those from typical cells, such that $P_r(x)$ is very small, are thus likely to originate from outlier cells that may be infected, stressed, or transformed. Biologically, information needed to represent the signal distribution could be encoded by the level of intracellular proteins involved in signaling cascades or by adjusting the distribution or spatial organization of surface receptors on the immune cell membrane.

We characterize the signal distribution that an immune cell receives from targets in its local environment with a Gaussian function, $P(x|\mu, \lambda)$. The true mean μ_t and precision $\lambda_t = 1/\sigma_t^2$ are unknown and must be estimated through multiple encounters with target cells. Estimating the variance σ_t^2 in signal values, as well as the mean, is crucial to differentiate true outlier target cells from ones with normal variation in surface ligand expression. In Bayesian inference, one begins with a prior distribution $P_{\text{prior}}(\mu, \lambda)$ which constitutes an initial guess for the unknown parameters. This distribution is then updated each time data (i.e., a signal) is received following Bayes' rule (*SI Appendix*, section A).

While we treat the signal x received by an immune cell in an abstract way, logarithmic distributions are very common in nature, and x may be thought of as proportional to the logarithm of receptor/ligand concentrations. As one example, distributions of immune receptors and internal signaling molecules in T cells (27) have been observed to follow a lognormal distribution. Logarithmic sensing of signals is also observed in many systems, including bacterial chemotaxis (28), developmental signaling (29), and responses to stimulation in both the innate (30) and adaptive immune systems (27).

In general, the true signal mean and variance will not be constant in time. For example, the distribution of ligands expressed on target cells may vary as an immune cell migrates from one tissue to another. To accommodate time-varying signals, we employ a modified Bayesian update rule where the strength of the prior distribution is fixed. We begin with a normal-gamma prior for the signal mean μ and precision λ,

$$P(\mu, \lambda|m, \kappa, \alpha, \beta) = \frac{\beta^\alpha \sqrt{\kappa}}{\Gamma(\alpha)\sqrt{2\pi}} \lambda^{\alpha-1/2} e^{-\beta\lambda-\kappa\lambda\frac{(\mu-m)^2}{2}}. \quad [1]$$

which is the conjugate prior for a Gaussian distribution with unknown mean and variance. This is a choice of mathematical convenience that allows us to easily write down analytical expressions for how parameters are updated as new signals are received, but it is not essential for our results. Here, $\Gamma(\alpha)$ represents the gamma distribution. The parameters κ and α are proportional to the number of measurements used to estimate the mean and variance, which we hold fixed. Following Eq. **1**, the mean value of μ is m, and the mean value of λ is α/β. When the immune cell binds to a new target cell and receives a signal x, these estimates shift,

$$m \to \frac{\kappa m + x}{\kappa + 1}, \quad \beta \to \frac{\alpha - 1}{\alpha - \frac{1}{2}} \left[\beta + \frac{\kappa}{\kappa + 1} \frac{(x - m)^2}{2}\right]. \quad [2]$$

This differs from the standard Bayesian update in that κ and α are held fixed, and the updated value of β is shrunk by a factor of $(\alpha - 1)/\left(\alpha - \frac{1}{2}\right)$ to compensate (*SI Appendix*, section A). We have also verified that this model is comparable to a Bayesian model for an explicitly time-varying signal distribution (*SI Appendix*, section E).

Our inference framework introduces a memory length for adaptation, encoded by the parameters κ and α. The larger κ

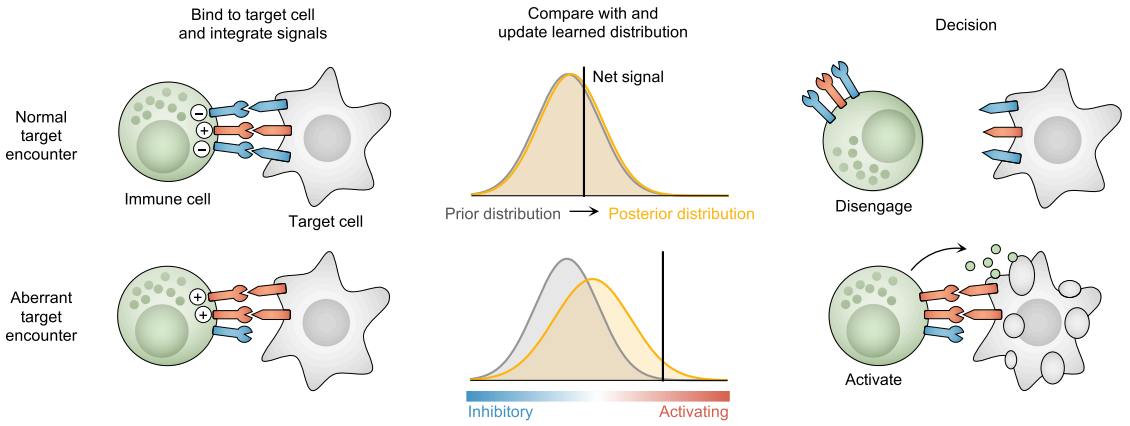

Fig. 1. Model overview. Immune cells receive both activating and inhibitory signals from target cells that they encounter. Net signals are used to update an internal estimate of the signal distribution, reflecting the balance of activating and inhibitory ligands on target cell surfaces in the current environment. Signals from target cells that are far more activating than typical ones stimulate an immune response.

and α are, the less the estimated mean and variance will shift when a new signal is received. The update rule described in Eq. **2** is equivalent to placing an exponentially decaying weight on measurements of the signal distribution, emphasizing more recent signals over older ones (*SI Appendix*, section A). Larger values of κ and α result in a slower decay, and thus a longer memory. We emphasize that this notion of memory is distinct from the concept of memory in adaptive immunity. Memory length presents a tradeoff: When the memory is long, it is possible to adapt to a specific distribution precisely, but adaptation to new environments is slow.

Integrating over the unknown mean and precision in Eq. **1**, the internal estimate of the signal distribution $P_r(x)$ takes the form of a shifted, scaled t-distribution (*SI Appendix*, section A). During each target cell encounter, the immune cell activates if it receives a signal x that is substantially more activating than typical signals from the estimated distribution. Specifically, we assume that activation occurs when

$$P_r(y > x) = \int_x^\infty dy \int_{-\infty}^\infty d\mu \int_0^\infty d\lambda \sqrt{\frac{\lambda}{2\pi}} e^{-\frac{\lambda}{2}(y-\mu)^2} P(\mu, \lambda) < \theta, \quad [3]$$

for some threshold value θ. The threshold θ must be small to avoid autoimmunity. If an immune cell were to learn the exact signal distribution in an environment where all target cells are healthy, then θ would be the probability that the immune cell activates against a healthy cell. We will treat θ as constant, but in principle, θ could be modulated adaptively by the immune system. For example, modulation by cytokines could provide heightened surveillance during infection.

Below, we show that this simple model recapitulates many features of NK cell behavior. This includes both qualitative observations and quantitative fits to three experimental datasets. Extension to a model with multiple receptors/ligands also reveals the potential advantage of a sparse and stochastic distribution of receptor expression on individual immune cells, which has been observed in experiments.

Results

Immune Adaptation in a Static Environment. We first checked the ability of our model to recover the true parameters of a test signal distribution. We compared the estimated values of μ and σ ($\hat{\mu} = \langle \mu \rangle = m$, $\hat{\sigma} = \sqrt{\langle 1/\lambda \rangle} = \sqrt{\beta/(\alpha - 1)}$, where $\langle \cdot \rangle$ denotes an average over Eq. **1**) to those of the true signal distribution. Fig. 2 shows that immune cells adapt to the distribution of signals in the environment, even when the initial parameters of the prior are far from the true ones. The number of target cell encounters required to approach the true parameters depends on the initial distance from them, and on the memory length. The shorter the memory length, the faster the convergence (*SI Appendix*, Fig. S1).

Heterogeneous Responses Induced by Finite Memory. Finite memory lengths also mean that adaptation to the true signal mean and variance is not perfect. Shorter memory lengths result in greater "noise" in the inferred parameters of the signal distribution. Together with the stochastic nature of target cell encounters, this imperfect adaptation also implies that there will be a range of immune cell responses to particular target cells, even among immune cells with identical receptors. We systematically explored this heterogeneity by characterizing the distribution of estimated signal parameters m and β for a population of identical immune cells with finite memory that inhabit the same environment (*SI Appendix*, section B and Figs. S2 and S3).

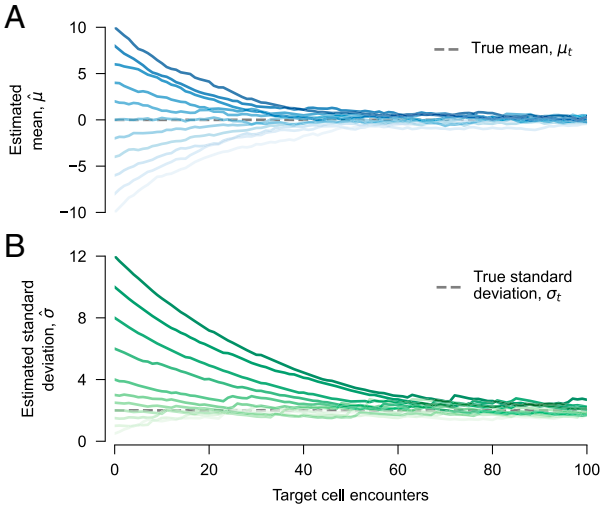

Fig. 2. Immune cells adapt to a static environment. The mean $\hat{\mu}$ and SD $\hat{\sigma}$ of the internal representation of the signal distribution converge toward the true mean ($\mu_t = 0$) and SD ($\sigma_t = 2$) of the signal distribution in the environment. (*A*) Convergence to μ_t from various initial values of m. The initial value of $\beta = (\alpha - 1)$ is the same in all cases. (*B*) Convergence to σ_t from various initial values of β. The initial value of $m = 0$ is the same in all cases. For both sections, the memory length is set by $\kappa = 20$ and $\alpha = 10$.

Consistent with experimental findings, we observe that individual immune cells can have varying responses to the same target cells. Investigations of the patterns of target cell killing have observed widely varying responses for individual NK cells: Some kill many targets efficiently, while others are inactive (31–36). This finding holds not only for diverse primary cells (31–33) but also for NK cell lines (34–36), where cells in the population would be predicted to have homogeneous expression and density of receptors.

Immune Adaptation in Dynamic Environments. NK cells show a remarkable ability to adapt to different environments. As noted above, although MHC class I is a powerful inhibitory ligand, NK cells in hosts with low levels of MHC class I expression are self-tolerant. Yet, they are also hyporesponsive to MHC class I–deficient target cells that would usually be killed by NK cells from normal hosts (12, 14). A pair of experiments showed that these behaviors are dynamic, rather than being fixed during development. When mature NK cells were transferred from MHC class I–deficient mice into mice with normal MHC class I expression, they regained their ability to kill MHC-deficient targets (37, 38). Conversely, NK cells from normal mice that were transferred into the MHC class I–deficient environment became hyposensitive (37). Importantly, this shift in behavior occurs without the need for cell division or changes in the composition of receptors on the NK cell surface (37). Similarly, NK cells can be temporarily desensitized through long-duration exposure to NKG2D ligands (39–41) or other activating stimuli (42–44).

We ran a series of simulations to mimic the transient exposure of immune cells to different environments or levels of stimulus. In our simulations, immune cells began in a "normal" environment. To quantify their ability to activate against targets that express high levels of activating ligands and/or low levels of inhibitory ligands, we computed the probability that an individual immune cell activates in response to an aberrant target cell that presents much more activating signals than the normal cell population.

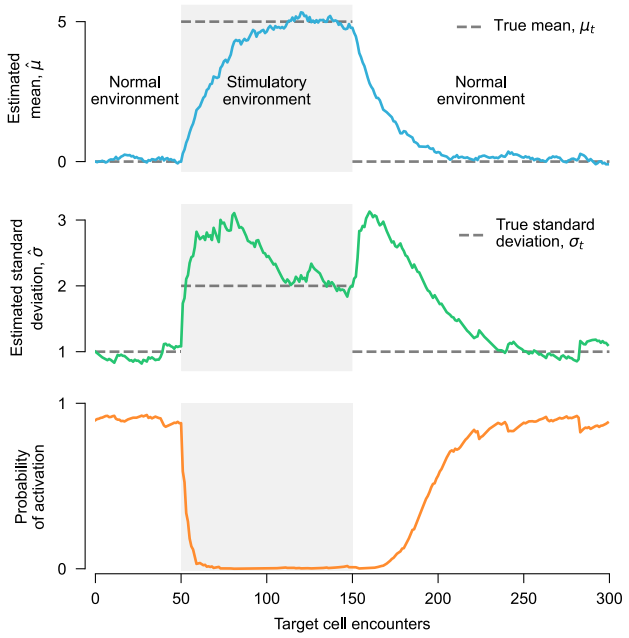

Fig. 3. Immune cells adapt to changing environments, mimicking experimentally observed development of hyposensitivity and recovery. Adaptation of the estimated signal mean $\hat{\mu}$, SD $\hat{\sigma}$, and probability of activation against an aberrant target as an immune cell is transferred between different environments. In the initial, "normal" environment ($\mu_t = 0$, σ_t=1), the level of activating stimulus is low, and the immune cell is primed to respond to aberrant targets. After transfer to a new, more stimulating environment (shaded region, $\mu_t = 5$, $\sigma_t = 2$), the immune cell adapts to the new signal distribution and progressively loses the ability to respond to aberrant targets. When the immune cell is returned to the "normal" environment, responsiveness is gradually restored. To compute the probability of activation, we used a Gaussian signal distribution for aberrant targets that matches the environment in the shaded region. Initial parameters of the estimated signal distribution are $m = 0$ and $\beta = 9$ (leading to $\hat{\mu} = 0$, $\hat{\sigma} = 1$), with memory parameters $\kappa = 20$ and $\alpha = 10$, and threshold $\theta = 0.01$.

The probability of activation against these outlier targets is initially high (Fig. 3).

To simulate a change in environment, after 50 encounters with normal targets, we switched the distribution of signals to match the signal distribution from "aberrant" targets described above. As the immune cell adapts to this new environment, its probability of activation by aberrant target cells decreases, ultimately settling near zero. However, this hyposensitivity is not permanent. After a total of 150 target cell encounters, we return the immune cell to the normal environment. As the immune cell encounters healthy cells, it regains its capacity to activate against aberrant targets.

This trajectory mimics the development of hyposensitivity and restoration of normal function described in experiments (37–44). Similar dynamics are observed consistently in our simulations (*SI Appendix*, Fig. S4), though the speed of adaptation depends on memory length, as discussed above, and the difference in environments. This is also observed in a modified, thresholded signaling model that mimics a saturation effect for strong activating or inhibitory stimuli (*SI Appendix*, section A and Figs. S4 and S5).

Quantitative Comparisons with Experiments. To quantitatively test our model, we assessed its ability to recover the results of three experimental studies (38, 45, 46). Because adaptation to stimulus is a central feature of our model, we focused on experiments that measured, directly or indirectly, the outcome of multiple rounds of exposure to target cells for NK cells. We also identified experiments that measured quantitative outcomes that would allow for direct comparison with our model. For comparing our model to these experiments, we always fixed the mean signal strength for normal cells to $\mu = 0$, assumed a unit signal variance for all target cells, and fixed the memory parameter $\kappa = 2\alpha$.

We first studied data from two sets of experiments where NK cells were exposed to multiple target cells. In the first experiment (45), NK cells encountered Daudi cells (a malignant B cell line) that were either opsonized with Rituximab, an antibody that can activate NK cells through CD16, or transfected to express MICA, a ligand for the activating receptor NKG2D. NK cells interacted with two target cells of each type in varying orders, and it was recorded whether the NK cell killed the first target only, the second target only, both targets, or neither of them. To test our ability to predict the outcome of novel sequences of interactions, we used only Rituximab–Rituximab and MICA–MICA data to fit model parameters using approximate Bayesian computation (ABC) (47) (*SI Appendix*, section C). Specifically, we fit the memory length α (fixing $\kappa = 2\alpha$) and the mean signal strengths μ_R and μ_M for Rituximab and MICA targets, respectively, to minimize the sum of squared errors between the model and data. Fig. 4*A* shows an excellent quantitative fit to the data (sum of squared errors equal to 0.08), including Rituximab–MICA and MICA–Rituximab encounters that were not used for training.

Beyond the general fit to data, our model makes two notable predictions. First, we predict that when an NK cell encounters

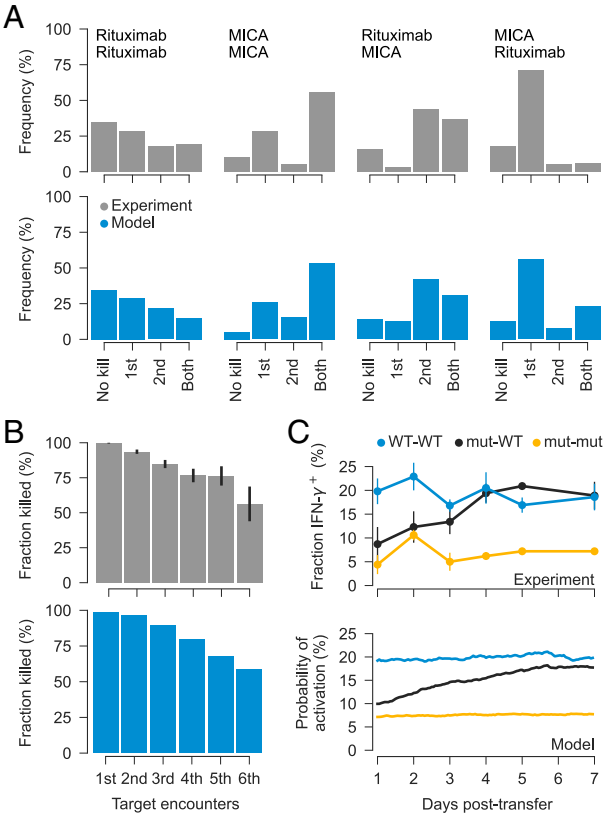

Fig. 4. Our model quantitatively fits experimental data. (*A*) Experimental data (45) and model predictions for sequential encounters with two types of target cells labeled Rituximab and MICA, respectively. (*B*) Experimental data (46) and model fit for the fraction of target cells killed after a certain number of serial interactions. (*C*) Experimental data (38) and model fit for NK cell responsiveness in a normal environment (WT) versus MHC class I–deficient environment (mut). NK cells transferred from the MHC class I–deficient environment to the normal one at day zero are labeled (mut-WT). See (*SI Appendix*, section C) for details on experimental datasets and model fitting.

the same type of target cell two times in a row, the probability that it kills the first target cell but not the second will be higher than the probability that it kills the second target cell but not the first. Similarly, when an NK cell encounters two different types of target cells, its probability of killing both will be higher if it encounters the less stimulatory target cell first (in this experiment, the one labeled Rituximab). These phenomena are clearly observed in the data (*SI Appendix*, section C and Fig. 4*A*), but they cannot be reproduced by a simple model that assumes that the probability that an NK cell kills a particular type of target cell is constant over time (*SI Appendix*, Fig. S7).

Next, we compared our model to experiments (46) in which IL-2-activated NK cells encountered HeLa-CD48 target cells. These target cells are HeLa cells transfected to express CD48, a ligand for the activating receptor 2B4. As above, we used ABC to fit the memory length α and mean signal strength μ_H for HeLa-CD48 target cells, selecting for parameters that minimized the squared difference between simulated and experimental response frequencies normalized by the experimental uncertainty in response frequencies due to finite sampling. We used this normalization to prevent the model fit from being overly biased by response frequencies after a large number of target cell encounters, which were computed from fewer interactions. Our model readily recovers the steady decline in the fraction of NK cells that kill their targets as NK cells undergo multiple encounters with HeLa-CD48 target cells (*SI Appendix*, section C and Fig. 4*B*).

Finally, we studied an experiment (38) in which NK cells from normal mice (WT), MHC class I–deficient mice (mut), and NK cells that had been transferred from MHC class I–deficient mice to normal ones were compared. NK cell production of interferon gamma (IFN-γ) after anti-NK1.1 stimulation, a readout for NK cell activity, was then measured at multiple time points. NK cells from WT mice were more responsive than NK cells from mutant mice. Responses from NK cells that were transferred from mutant mice to WT gradually approached those of NK cells from WT mice that were not transferred. Here, we fit the mean signal strength in MHC class I–deficient mice μ_D and anti-NK1.1 stimulation μ_A, using a similar memory length as inferred in the past two experiments, $\alpha = 20$, and assuming an average of one target cell interaction per hour. As in the previous cases, our model quantitatively recovers these dynamics (Fig. 4*C*).

Remarkably, across these experimental datasets, the memory lengths that best fit the data are nearly identical: $\alpha = \kappa/2 \sim 20$ (*SI Appendix*, section C). This convergence is striking because the experiments involve the engagement of different NK cell receptors and different metrics for NK cell activity. The values for κ and α inferred from experimental data suggest an effective NK cell memory length of ~25 target cell encounters, calculated by computing the "half-life" for exponentially decaying signal measurements (*SI Appendix*, section A).

Multiple Receptors/Ligands and the Immune Repertoire. NK cells use a wide array of germline-encoded activating and inhibitory receptors to recognize their targets (4, 9–11). However, not all of these receptors are expressed by every NK cell (48). Recent single-cell experiments have further revealed that NK cell receptor expression is "sparse," with each receptor only expressed on a small fraction (average ~ 20%) of individual cells (25, 26). This result may seem surprising. In principle, one might expect that expressing more receptors would provide an NK cell with more information about the target cells that it surveys, allowing it to better identify outliers. What advantage could sparse receptor expression patterns provide to the immune system?

To explore this question, we extended our model to consider contributions from multiple receptors and ligands (*SI Appendix*, section D). For simplicity, we assumed that receptor expression by individual cells was binary, that is, each receptor i is either expressed ($e_i = 1$) or not expressed ($e_i = 0$). The results we describe below using binary receptor expression are qualitatively identical to a more complex model with continuous variation in receptor expression (*SI Appendix*, section D and Figs. S8 and S9). We write the net signal received by an NK cell as the sum of the signals from each of its receptors,

$$x = \sum_{i=1}^{n_r} a_i e_i x_i. \qquad [4]$$

Each individual signal x_i scales in proportion with the concentration of the ligand that binds to receptor i on the target cell surface, and n_r is the number of receptor/ligand pairs. The parameter a_i encodes activation/inhibition, with $a_i = 1$ for activating receptors and $a_i = -1$ for inhibitory receptors. To simulate diversity in the immune repertoire, we generated a large population of immune cells with random receptor expression patterns following one of two rules: sparse (probability of expressing a receptor $P = 0.2$, consistent with experiments) and dense ($P = 0.8$).

We then tested the ability of each population of immune cells to recognize aberrant target cells with perturbed ligand expression. To comprehensively explore how differences in the immune cell repertoire affect target cell recognition, we considered multiple scenarios. Perturbations could be either sparse (concentrated in a few ligands) or dense (spread randomly across all ligands). We considered two additional subtypes of perturbations: ones that are purely activating (i.e., through the upregulation of activating ligands and/or downregulation of inhibitory ligands) and "hidden" ones, where, for example, upregulation of activating ligands is balanced by the upregulation of inhibitory ligands. Hidden perturbations mimic cases where aberrant cells may partially evade innate immunity, such as when some viruses disrupt the expression of activating ligands on the infected cell surface (49).

Fig. 5 shows that a subpopulation of immune cells with sparse receptor expression is able to efficiently recognize and respond to all the different types of aberrant targets that we considered. Immune cells that express many receptors are unable to effectively respond to aberrant cells with changes in ligand expression that are concentrated in just a few ligands, or ones where activating ligands are hidden by increased inhibitory ligand expression.

To better understand the factors that underlie successful target cell recognition, we quantified the alignment between the distribution of receptors expressed by immune cells and the ligands on target cells. As described above, we can consider the pattern of receptor expression for each immune cell as a vector, $\mathbf{e} = \{e_1, e_2, \ldots, e_{n_r}\}$. We can also quantify the levels of ligand expression on target cells in the same way, $\boldsymbol{\ell} = \{\ell_1, \ell_2, \ldots, \ell_{n_r}\}$. Let $\boldsymbol{\ell}_h$ and $\boldsymbol{\ell}_a$ represent the average levels of ligand expression on healthy and aberrant cells, respectively. We then quantify the alignment between receptor expression for a particular immune cell and the perturbation that defines an aberrant cell population as

$$\frac{\mathbf{e} \cdot \Delta\boldsymbol{\ell}}{\sqrt{(\mathbf{e} \cdot \mathbf{e}) \times (\Delta\boldsymbol{\ell} \cdot \Delta\boldsymbol{\ell})}}, \qquad [5]$$

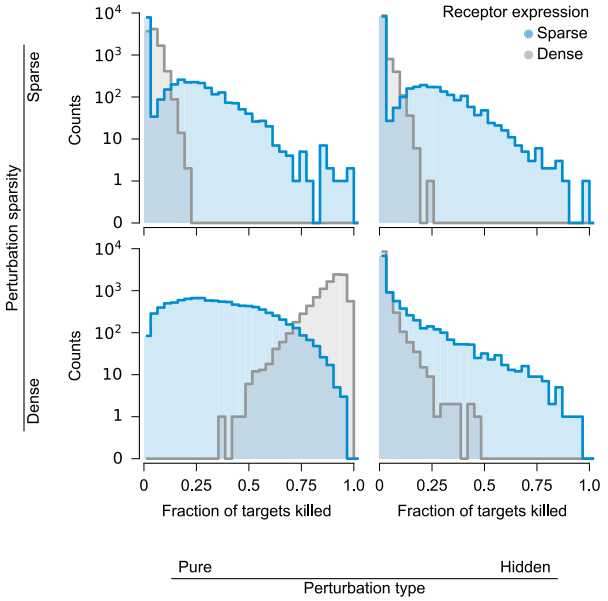

Fig. 5. Immune cells with sparse receptor expression are better able to detect a wide variety of aberrant target cells. For all types of aberrant cells that we considered, there exist some immune cells with sparse receptor expression that are able to kill a large fraction of aberrant cells that they encounter. In contrast, immune cells that express many receptors are unable to efficiently kill aberrant target cells with "hidden" perturbations, where some changes in ligand expression promote activation and some promote inhibition. Immune cells with dense receptor expression are also unable to efficiently eliminate aberrant targets that only have unusual levels of expression for one or two ligands.

where $\Delta\boldsymbol{\ell} = \mathbf{a} \times (\boldsymbol{\ell}_a - \boldsymbol{\ell}_b)$, $\times$ denotes element-wise or scalar multiplication, and $\cdot$ represents a dot product. As above, entries in the vector $\mathbf{a}$ have $a_i = 1$ if receptor i is activating and $a_i = -1$ if it is inhibitory.

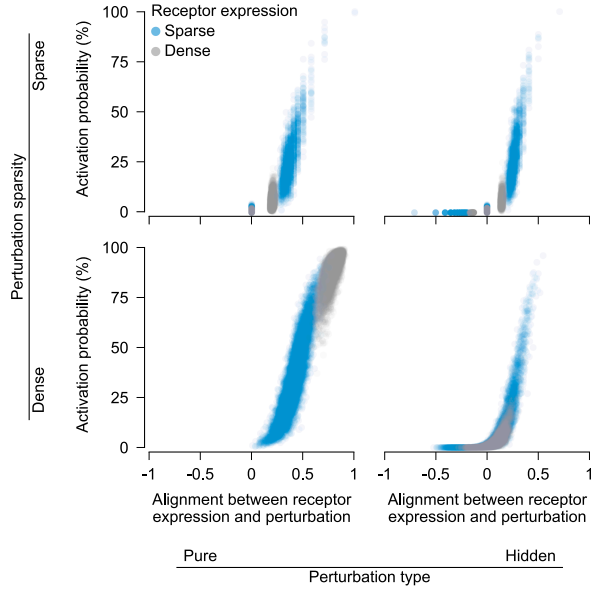

Fig. 6. Immune cells best recognize aberrant targets when they express the specific combination of receptors that recognize ligands with perturbed expression. Immune cells with sparse receptor expression are more likely to align with changes in ligand expression for a variety of aberrant target cells, allowing them to kill these targets more efficiently. To quantify the alignment between immune cell receptors and perturbations, we treated the perturbations in ligand expression in aberrant cells and receptor expression patterns in immune cells as vectors and computed their normalized inner product.

This analysis suggested that populations of immune cells with sparse receptor expression patterns are better able to recognize diverse targets because they maximize the ratio of signal to noise. Immune cells must tolerate normal variation in ligand expression on healthy cells, which grows additively with the number of receptors that an immune cell expresses. When receptor expression is sparse, some immune cells will express receptors that correspond precisely to the ligands that are affected in aberrant cells (resulting in high alignment, as quantified above), allowing them to reliably recognize and eliminate these cells (Fig. 6). In contrast, immune cells with dense receptor expression receive many signals that are irrelevant for target detection, making it more challenging to distinguish aberrant targets from normal variation.

Discussion

Here, we developed a simple, quantitative model for how NK cells discriminate between healthy and unhealthy target cells. Our model is grounded in the idea of self/nonself discrimination as a kind of outlier detection problem. In our model, which qualitatively recovers observed NK cell behaviors and fits quantitative data from a variety of experiments, NK cell responsiveness is adaptively shaped by the targets that they encounter. Analysis of experimental data pointed to a consistent NK cell "memory length" of roughly 25 target cell encounters. Our model also suggests that the sparse and variegated patterns of receptor expression on individual NK cells have a purpose, allowing individual NK cells with particular sets of receptors to detect subtle perturbations in ligand expression on infected or transformed cells while maintaining robust self-tolerance.

In this work, we mostly considered homogeneous populations of cells. However, a more realistic scenario would include a mixture of different cell types, including mostly normal cells together with a smaller fraction of unhealthy ones. In such cases, our model predicts that immune cells remain responsive to aberrant target cells as long as they comprise a small fraction of all cells (*SI Appendix*, Fig. S10). The spatial organization of different cell types could also affect target cell recognition, a topic we plan to explore in future work.

Quantitative analyses of immunity have often focused on the adaptive immune system and antigen-specific recognition of foreign material to distinguish self from nonself (50–53). The "algorithm" for self/nonself discrimination that we describe operates in a very different way. Rather than learning to detect specific pathogens, our model immune cells learn the properties of nearby healthy cells, which allows them to respond to aberrant cells that may be infected, stressed, or transformed. Learning in our model is "unsupervised" in the sense that immune cells do not have external information about whether the targets that they encounter are healthy or not. Nonetheless, it operates reliably following the simple assumption that the great majority of targets that immune cells encounter are likely to be normal, healthy cells.

Recent work has also applied ideas from Bayesian inference to immunity, focusing in particular on rules for optimizing the adaptive immune repertoire (53, 54). There, models were constructed to optimally allocate immune cells with different antigen specificities to combat a shifting environment of pathogens. Our work is similar in that we also consider adaptation to varying environments. However, we focus on the adaptation of individual immune cells. The parameters κ and α set a natural time scale over which memory of the environment is retained, but we do not presuppose an optimal memory length. Through detailed analyses, we found that our model performs very similarly to

alternative inference models that explicitly encode information about a time-varying environment (*SI Appendix*, section E and Figs. S11 and S12). Other intriguing studies have developed analogies between machine learning and immunity (55, 56), including a model of negative selection in T cells, where encounters with self-peptides are central (56). More generally, the problem of estimating time-varying signal distributions has some similarities with estimation using Kalman filters (57, 58). An important difference in the present case is that the signal variance must also be estimated, and the ways that the signal mean and variance change are unknown.

Similarly, our model can be compared with the discontinuity theory of immunity (59), which posits that immune cells in general respond to sharp changes in the environment. One of the main differences between our model and the discontinuity theory is that we explicitly consider the variance of signals in the environment. Beyond the comparisons with experiment that we have performed, there is additional evidence that variance in ligand expression is important. Recent work showed that MHC class I–deficient hematopoietic cells are spared in mice that also have hematopoietic cells with normal levels of MHC class I expression but only if the MHC class I–deficient cells comprise a substantial fraction of all cells (60).

Though we have focused on NK cells, the model we developed may also apply more broadly to other cell types. For example, macrophages have many similarities with NK cells. Many activating receptors for macrophages respond to self-ligands that indicate cellular damage, and macrophage activity is inhibited by ubiquitously expressed surface proteins such as CD47 (24), similar to NK cell inhibition by MHC class I. Macrophages from CD47-deficient hosts are also tolerant of cells with low levels of CD47 expression, which would usually be phagocytosed (61). Extended exposure to lipopolysaccharide (LPS) in macrophages results in blunted responses to additional stimulation by LPS, a phenomenon known as endotoxin tolerance (62, 63). When LPS is withdrawn, macrophages gradually recover normal function (62). There are also some conceptual similarities between our model and the "tunable threshold" model, originally applied to T cell signaling (64, 65). Recent work demonstrated that T cells adapt to basal levels of T cell receptor (TCR) signaling and that cells with stronger basal signaling were less responsive (66). Importantly, this work also demonstrated that even T cells with identical TCRs exhibit heterogeneous responses to stimulus (66), which is one of the main predictions of our model.

Materials and Methods

Details of this section are contained in *SI Appendix*. Specifically, *SI Appendix*, section A provides details about our inference model and its derivation. *SI Appendix*, section B describes our investigation of the heterogeneity in responses across a population of immune cells. Comparisons with experimental data are described in *SI Appendix*, section C. In *SI Appendix*, section D, we describe the extension of the basic model to include multiple receptors and ligands. Finally, *SI Appendix*, section E provides a detailed comparison of our inference model, in which the potential change in the environment that an immune cell encounters over time is implicit, with one that explicitly accounts for time-varying signals.

Data, Materials, and Software Availability. Data and code used in our analysis are available at the GitHub repository https://github.com/bartonlab/paper-innate-immune-adaptation (67). This repository also contains Jupyter notebooks that can be run to reproduce the results presented here.

ACKNOWLEDGMENTS. Computations were performed using the computer clusters and data storage resources of the University of California, Riverside High Performance Computing Center, which were funded by grants from NSF (MRI-1429826) and NIH (1S10OD016290-01A1). This work is supported in part by R01AI137073 to E.M.M.

1. S. I. Khakoo *et al.*, HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. *Science* **305**, 872–874 (2004).
2. M. P. Martin *et al.*, Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. *Nat. Genet.* **39**, 733 (2007).
3. G. Alter *et al.*, Differential natural killer cell-mediated inhibition of HIV-1 replication based on distinct KIR/HLA subtypes. *J. Exp. Med.* **204**, 3027–3036 (2007).
4. E. Vivier, E. Tomasello, M. Baratin, T. Walzer, S. Ugolini, Functions of natural killer cells. *Nat. Immunol.* **9**, 503 (2008).
5. I. Waldhauer, A. Steinle, NK cells and cancer immunosurveillance. *Oncogene* **27**, 5932 (2008).
6. J. Pahl, A. Cerwenka, Tricking the balance: NK cells in anti-cancer immunity. *Immunobiology* **222**, 11–20 (2017).
7. L. Chiossone, P. Y. Dumas, M. Vienne, E. Vivier, Natural killer cells and other innate lymphoid cells in cancer. *Nat. Rev. Immunol.* **18**, 671–688 (2018).
8. K. Kärre, H. G. Ljunggren, G. Piontek, R. Kiessling, Selective rejection of H-2-deficient lymphoma variants suggests alternative immune defence strategy. *Nature* **319**, 675 (1986).
9. L. L. Lanier, NK cell receptors. *Annu. Rev. Immunol.* **16**, 359–393 (1998).
10. E. O. Long, H. Sik Kim, D. Liu, M. E. Peterson, S. Rajagopalan, Controlling natural killer cell responses: Integration of signals for activation and inhibition. *Annu. Rev. Immunol.* **31**, 227–258 (2013).
11. L. Martinet, M. J. Smyth, Balancing natural killer cell activation through paired receptors. *Nat. Rev. Immunol.* **15**, 243 (2015).
12. N. S. Liao, M. Bix, M. Zijlstra, R. Jaenisch, D. Raulet, MHC class I deficiency: Susceptibility to natural killer (NK) cells and impaired NK activity. *Science* **253**, 199–202 (1991).
13. P. Hoglund *et al.*, Recognition of β2-microglobulin-negative (β2m-) t-cell blasts by natural killer cells from normal but not from β2m-mice: Nonresponsiveness controlled by β2m-bone marrow in chimeric mice. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10332–10336 (1991).
14. M. Bix *et al.*, Rejection of class I MHC-deficient haemopoietic cells by irradiated MHC-matched mice. *Nature* **349**, 329 (1991).
15. H. de La Salle *et al.*, Homozygous human tap peptide transporter mutation in HLA class I deficiency. *Science* **265**, 237–241 (1994).
16. J. Zimmer *et al.*, Activity and phenotype of natural killer cells in peptide transporter (tap)-deficient patients (type i bare lymphocyte syndrome). *J. Exp. Med.* **187**, 117–122 (1998).
17. S. Kim *et al.*, Licensing of natural killer cells by host major histocompatibility complex class I molecules. *Nature* **436**, 709 (2005).
18. N. Anfossi *et al.*, Human NK cell education by inhibitory receptors for MHC class I. *Immunity* **25**, 331–342 (2006).
19. W. M. Yokoyama, S. Kim, How do natural killer cells find self to achieve tolerance? *Immunity* **24**, 249–257 (2006).
20. D. H. Raulet, R. E. Vance, Self-tolerance of natural killer cells. *Nat. Rev. Immunol.* **6**, 520 (2006).
21. P. Brodin, K. Kärre, P. Höglund, NK cell education: Not an on-off switch but a tunable rheostat. *Trends Immunol.* **30**, 143–149 (2009).
22. M. T. Orr, W. J. Murphy, L. L. Lanier, "Unlicensed" natural killer cells dominate the response to cytomegalovirus infection. *Nat. Immunol.* **11**, 321–327 (2010).
23. N. Tarek *et al.*, Unlicensed NK cells target neuroblastoma following anti-GD2 antibody treatment. *J. Clin. Invest.* **122**, 3260–3270 (2012).
24. P. R. Taylor *et al.*, Macrophage receptors and immune recognition. *Annu. Rev. Immunol.* **23**, 901–944 (2005).
25. A. Horowitz *et al.*, Genetic and environmental determinants of human NK cell diversity revealed by mass cytometry. *Sci. Transl. Med.* **5**, 208ra145-208ra145 (2013).
26. D. M. Strauss-Albee *et al.*, Human NK cell repertoire diversity reflects immune experience and correlates with viral susceptibility. *Sci. Transl. Med.* **7**, 297ra115-297ra115 (2015).
27. O. Feinerman, J. Veiga, J. R. Dorfman, R. N. Germain, G. Altan-Bonnet, Variability and robustness in T cell activation from regulated heterogeneity in protein levels. *Science* **321**, 1081–1084 (2008).
28. V. Sourjik, N. S. Wingreen, Responding to chemical gradients: Bacterial chemotaxis. *Curr. Opin. Cell Biol.* **24**, 262–268 (2012).
29. L. Goentoro, M. W. Kirschner, Evidence that fold-change, and not absolute level, of β-catenin dictates Wnt signaling. *Mol. Cell* **36**, 872–884 (2009).
30. R. E. Lee, S. R. Walker, K. Savery, D. A. Frank, S. Gaudet, Fold change of nuclear NF-κb determines TNF-induced transcription in single cells. *Mol. Cell* **53**, 867–879 (2014).
31. B. Vanherberghen *et al.*, Classification of human natural killer cells based on migration behavior and cytotoxic response. *Blood* **121**, 1326–1334 (2013).
32. A. E. Christakou *et al.*, Live cell imaging in a micro-array of acoustic traps facilitates quantification of natural killer cell heterogeneity. *Integr. Biol.* **5**, 712–719 (2013).
33. K. Guldevall *et al.*, Microchip screening platform for single cell assessment of NK cell cytotoxicity. *Front. Immunol.* **7**, 119 (2016).
34. P. J. Choi, T. J. Mitchison, Imaging burst kinetics and spatial coordination during serial killing by single natural killer cells. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6488–6493 (2013).
35. Y. Xu, S. Zhou, Y. W. Lam, S. W. Pang, Dynamics of natural killer cells cytotoxicity in microwell arrays with connecting channels. *Front. Immunol.* **8**, 998 (2017).
36. L. A. Gwalani, J. S. Orange, Single degranulations in NK cells can mediate target cell killing. *J. Immunol.* **200**, 3231–3243 (2018).

37. N. T. Joncker, N. Shifrin, F. Delebecque, D. H. Raulet, Mature natural killer cells reset their responsiveness when exposed to an altered MHC environment. *J. Exp. Med.* **207**, 2065–2072 (2010).
38. J. M. Elliott, J. A. Wahle, W. M. Yokoyama, MHC class I-deficient natural killer cells acquire a licensed phenotype after transfer into an MHC class I-sufficient environment. *J. Exp. Med.* **207**, 2073–2079 (2010).
39. D. E. Oppenheim *et al*., Sustained localized expression of ligand for the activating NKG2D receptor impairs natural cytotoxicity in vivo and reduces tumor immunosurveillance. *Nat. Immunol.* **6**, 928 (2005).
40. J. D. Coudert, L. Scarpellino, F. Gros, E. Vivier, W. Held, Sustained NKG2D engagement induces cross-tolerance of multiple distinct NK cell activation pathways. *Blood* **111**, 3571–3578 (2008).
41. M. Champsaur, L. L. Lanier, Effect of NKG2D ligand expression on host immune responses. *Immunol. Rev.* **235**, 267–285 (2010).
42. T. C. George, J. R. Ortaldo, S. Lemieux, V. Kumar, M. Bennett, Tolerance and alloreactivity of the LY49D subset of murine NK cells. *J. Immunol.* **163**, 1859–1867 (1999).
43. J. C. Sun, L. L. Lanier, Tolerance of NK cells encountering their viral ligand during development. *J. Exp. Med.* **205**, 1819–1828 (2008).
44. S. K. Tripathy *et al*., Continuous engagement of a self-specific activation receptor induces NK cell tolerance. *J. Exp. Med.* **205**, 1829–1841 (2008).
45. K. Srpan *et al*., Shedding of CD16 disassembles the NK cell immune synapse and boosts serial engagement of target cells. *J. Cell Biol.* **217**, 3267–3283 (2018).
46. I. Prager *et al*., NK cells switch from granzyme B to death receptor-mediated cytotoxicity during serial killing. *J. Exp. Med.* **216**, 2113–2127 (2019).
47. J. Lintusaari *et al*., Elfi: Engine for likelihood-free inference. *J. Mach. Learn. Res.* **19**, 1–7 (2018).
48. D. H. Raulet, R. E. Vance, C. W. McMahon, Regulation of the natural killer cell receptor repertoire. *Annu. Rev. Immunol.* **19**, 291–330 (2001).
49. Y. Ma, X. Li, E. Kuang, Viral evasion of natural killer cell activation. *Viruses* **8**, 95 (2016).
50. J. D. Farmer, N. H. Packard, A. S. Perelson, The immune system, adaptation, and machine learning. *Physica D* **22**, 187–204 (1986).
51. S. Forrest, A. S. Perelson, L. Allen, R. Cherukuri, "Self-nonself discrimination in a computer" in *Proceedings of 1994 IEEE Computer Society Symposium on Research in Security and Privacy* (IEEE, 1994), pp. 202–212.
52. A. S. Perelson, G. Weisbuch, Immunology for physicists. *Rev. Mod. Phys.* **69**, 1219 (1997).
53. A. Mayer, V. Balasubramanian, A. M. Walczak, T. Mora, How a well-adapting immune system remembers. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 8815–8823 (2019).
54. A. Mayer, V. Balasubramanian, T. Mora, A. M. Walczak, How a well-adapted immune system is organized. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5950–5955 (2015).
55. T. J. Rademaker, E. Bengio, P. François, Attack and defense in cellular decision-making: Lessons from machine learning. *Phys. Rev. X* **9**, 031012 (2019).
56. I. Wortel, C. Keşmir, R. J. de Boer, J. N. Mandl, J. Textor, Is T cell negative selection a learning algorithm? *Cells* **9**, 690 (2020).
57. G. Welch, G. Bishop, *An Introduction to the Kalman filter* (University of North Carolina Chapel Hill, Chapel Hill, NC, 1995), pp. 127–132.
58. R. E. Kalman, A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82**, 35–45 (1960).
59. T. Pradeu, S. Jaeger, E. Vivier, The speed of change: Towards a discontinuity theory of immunity? *Nat. Rev. Immunol.* **13**, 764–769 (2013).
60. M. D. Bern *et al*., Inducible down-regulation of MHC class I results in natural killer cell tolerance. *J. Exp. Med.* **216**, 99–116 (2019).
61. H. Wang *et al*., Lack of CD47 on nonhematopoietic cells induces split macrophage tolerance to CD47null cells. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13744–13749 (2007).
62. M. A. West, W. Heagy, Endotoxin tolerance: A review. *Crit. Care Med.* **30**, S64–S73 (2002).
63. S. L. Foster, D. C. Hargreaves, R. Medzhitov, Gene-specific control of inflammation by TLR-induced chromatin modifications. *Nature* **447**, 972 (2007).
64. Z. Grossman, W. E. Paul, Adaptive cellular interactions in the immune system: The tunable activation threshold and the significance of subthreshold responses. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10365–10369 (1992).
65. Z. Grossman, W. E. Paul, Dynamic tuning of lymphocytes: Physiological basis, mechanisms, and function. *Annu. Rev. Immunol.* **33**, 677–713 (2015).
66. W. M. Zinzow-Kramer, A. Weiss, B. B. Au-Yeung, Adaptation by naïve CD4+ T cells to self-antigen-dependent TCR signaling induces functional heterogeneity and tolerance *Proc. Natl. Acad. Sci. U.S.A.* **116**, 15160–15169 (2019).
67. Y. Qin, E. M. Mace, J. P. Barton, bartonlab/paper-innate-immune-adaptation. GitHub. https://github.com/bartonlab/paper-innate-immune-adaptation. Deposited 23 August 2023.