Ligand Discrimination in Immune Cells: Signal Processing Insights into Immune Dysfunction in ER+ Breast Cancer

Adina Matache[1+], Joao Rodrigues Lima Jr[2+], Maxim Kuznetsov[1], Konstancja Urbaniak[1], Sergio Branciamore[1], Andrei S. Rodin[1], Peter P. Lee[2*], and Russell C. Rockne[1*]

[1]Department of Computational and Quantitative Medicine

[2]Department of Immuno-Oncology

Beckman Research Institute of City of Hope, Duarte, CA, 91010 USA.

+co-first authors amatache@coh.org, jrodrigueslimaj@coh.org

*co-corresponding authors rrockne@coh.org, plee@coh.org

Abstract: Prior studies have shown that approximately 40% of estrogen receptor positive (ER+) breast cancer (BC) patients harbor immune signaling defects in their blood at diagnosis, and the presence of these defects predicts overall survival. Therefore, it is of interest to quantitatively characterize and measure signaling errors in immune signaling systems in these patients. Here we propose a novel approach combining communication theory and signal processing concepts to model ligand discrimination in immune cells in the peripheral blood. We use the model to measure the specificity of ligand discrimination in the presence of molecular noise by estimating the probability of error, which is the probability of making a wrong ligand identification. We apply our model to the JAK/STAT signaling pathway using high dimensional spectral flow cytometry measurements of transcription factors, including phosphorylated STATs and SMADs, in immune cells stimulated with several cytokines (IFNγ, IL-2, IL-6, IL-4, and IL-10) from 19 ER+ breast cancer patients and 32 healthy controls. In addition, we apply our model to 10 healthy donor samples treated with a clinically approved JAK1/2 inhibitor. Our results show reduced ligand identification accuracy and higher levels of molecular noise in BC patients as compared to healthy controls, which may indicate altered immune signaling and the potential for immune cell dysfunction in these patients. Moreover, the inhibition of JAK1/2 produces ligand

misidentification and molecular noise rates similar to, or even greater than, those observed in breast cancer. These results suggest a means to improve the use of signaling kinase inhibitor therapies by identifying patients with favorable ligand discrimination specificity profiles in their immune cells.

One Sentence Summary: We use a communication model to measure ligand discrimination errors in immune cells and molecular noise in cytokine signaling in ER+ breast cancer patients as compared to healthy controls.

Keywords: JAK STAT, immune signaling, ligand discrimination, CD4+, cell communication, information theory, ER+ breast cancer, peripheral blood

INTRODUCTION

Precise and efficient communication between immune cells is essential for maintaining homeostasis and for mounting effective immune responses to pathogens. Cytokine-mediated signaling is a fundamental mechanism through which cells exchange information, that can be conceptualized as a communication network where cells act as both senders and receivers of signals. Here we apply concepts of information transmission, reception, and processing from digital communication theory to quantitatively study the function—and dysfunction—of immune signaling networks in healthy homeostasis and in breast cancer. Previous studies have shown that approximately 40% of estrogen receptor positive (ER+) breast cancer (BC) patients harbor immune signaling defects in their blood at diagnosis, and the presence of these defects predicts overall survival (*1–3*). It is therefore of interest to quantitively characterize and measure signaling errors and molecular noise in immune signaling systems.

Cell signaling is composed of several precisely regulated steps: cells release signaling molecules, such as cytokines, which diffuse between cells to interact with receptors on target cells. Receptor engagement initiates a cascade of intracellular events, including phosphorylation of proteins, that act in concert to determine cellular function or behavior. These processes mirror key principles of communication theory, including signal generation, message encoding, transmission, and reception, which are essential components for understanding how information is conveyed and interpreted between cells, particularly within the immune system. Cells encode signals from ligands through STAT phosphorylation. For example, IL-10 stimulation results in STAT3 phosphorylation, IL-2 induces STAT5 phosphorylation, and STAT6 is strongly phosphorylated by IL-6 (*4–7*). However, we find that each cytokine induces a pattern of phosphorylation across multiple STATs/SMADs and these patterns encode information that enable ligand discrimination (*8*).

Integrating digital communication theory into the study of cell signaling provides a mathematically rigorous and structured approach to studying cell signal processing. By examining the probability of signaling error and signal-to-noise ratio, we can gain insights into the efficiency and accuracy of cellular communication in immune cells from healthy donors versus patients with breast cancer. This interdisciplinary perspective not only enhances our understanding of signal transduction and cellular response mechanisms but also has significant implications for advancing research in immunology, and cell biology, and for therapeutic interventions which may interfere with, or may be designed to correct, errors in cell signaling. Here we study how communication theory principles can enhance our understanding of cell signaling alterations in peripheral blood immune cells from ER+ breast cancer, as compared to healthy donors.

Communication model

A growing number of studies have used information-theoretic approaches to analyze complex single-cell data and to understand reliable communication in cell populations (*9*–*14*), and to compile a compendium of responses to cytokine stimuli (*15*). Shannon's information theory (*16*), originally developed for digital communication systems, established fundamental limits on communication and information transmission. In Shannon's formulation, the basic problem of reliable transmission of information is stated in statistical terms, using probabilistic models for information sources and communication channels. At a high level, all communication systems have an information source (input) which sends a message, encoded or mapped into a unique signal by a transmitter. The signal is then transported from the transmitter to the receiver through a channel. Regardless of the physical medium used for transmission of the information, the main feature of the channel is that the transmitted signal is corrupted in a random manner by a variety

of mechanisms, such as additive noise. The receiver decodes the corrupted signal into a decoded message, which is consumed in turn by an information sink (output). Under the principles of information theory, reliable transmission of information is possible if the information rate from the source is less than the channel capacity, which is the maximal mutual information (MI) between the input and the output of the channel.

This communication model is general enough to be applied to systems other than digital communication systems, such as a cell communication system (Figure 1). An intuitive application to cellular signaling is that the source is a cell that secretes a cell-signaling protein or cytokine (the message), recognized by a receptor (the transmitter). The receptor initiates a signaling pathway (the channel), which results in the activation of a transcription factor (the receiver). The transcription factor in turn induces gene expression (the output). When cells are stimulated with different cytokines, each cell may respond differently due to molecular noise, variations in ligand-receptor affinity, and other factors. Here we focus on the JAK/STAT intracellular signaling axis (*5*), with each phosphorylation event of a STAT (pSTAT) or SMAD (pSMAD) molecule modeled as a random variable with an empirical distribution. Under a Gaussian distribution assumption, the problem of signal identification is similar to the problem of signal detection in digital communications, where the noise is assumed to be an additive, zero-mean Gaussian noise (*17*). The main difference is that in digital communication systems, the set of possible transmitted signals is known, while in cellular communication systems, the signals carried by cytokines are unknown. Therefore, we must first define the cytokine-induced signals from the responses determined by the phosphorylation status of the STATs. With a zero-mean Gaussian noise assumption, the signal is given by the mean of the pSTAT responses; therefore, the Gaussian assumption is conveniently apropos. Then, the basic communication model for immune signaling in the JAK/STAT signaling pathway consists of signal generation, noisy

channel transmission, and signal detection. The signal generation model maps each cytokine-induced signal to a multi-dimensional vector with elements equal to the means of the transcription factor responses. This vector defines a unique pattern of response for each cytokine treatment, which we call a *codeword*. The noisy channel adds a zero-mean Gaussian noise to the signaled codeword, and the detector performs optimal signal detection and identification.

Unlike previous studies that applied information-theoretic concepts such as channel capacity or mutual information to determine how much information is transmitted through signaling pathways, we adopt the perspective of detection theory to determine the specificity of signaling networks. As such, we view cytokine discrimination as the following inference problem to be solved by the cell: *given the number of intracellular readout molecules, determine if an extracellular ligand is present, and decide on its type.*

RESULTS

Profiling signaling response

To characterize the specificity of immune signaling responses in estrogen receptor positive (ER+) breast cancer patients, we analyzed peripheral blood mononuclear cells (PBMCs) from 51 subjects, 32 of which were healthy donors (HD), and 19 of which were newly diagnosed with ER+ breast cancer (BC) (Table S1, S2). We systematically treated the PBMCs with 5 different cytokines alone or in combination with a JAK1/2 inhibitor ruxolitinib to study signaling responses. The cytokines included IL-2, IL-4, IL-6, IL-10, and IFNγ. After stimulation for 15 minutes, cells were fixed and stained with a panel of 27 different intra- and extracellular markers, analyzed with spectral flow cytometry, and gated into distinct cell populations. Transcription factor phosphorylation was quantified by pSTAT1, pSTAT3, pSTAT4, pSTAT5, pSTAT6, and pSMAD2/3. Cell surface markers were used to identify CD4+ T cell subsets

including central memory (TCM) CD4+ CD45RA-CD27+, effector memory (TEM) CD4+ CD45RA-CD27-, and naïve T-cells CD4+ CD45RA+CD27+ (similarly for CD8+ T cells), CD20+ B cells, naïve B cells CD19+CD27-, classical monocytes CD14+CD16-, and CD3-CD16+ NK cells. Only cell populations with at least 200 cells were included in the analysis. The gating strategy to identify cell populations is provided in supplementary materials (Figures S1, S2).

Modeling ligand encoding and discrimination

Given the probabilistic nature of signal detection, we formulated the input signal (ligand) identification problem as a multi-hypothesis testing problem, where the receiver decides which of $M = 6$ signals is the input by choosing one of the following hypotheses H_i, H_0: baseline (no cytokine stimulation), H_1: stimulation with IL-4, H_2: stimulation with IL-2, H_3: stimulation with IL-10, H_4: stimulation with IL-6, or H_5: stimulation with IFNγ. The signal response space is the set of all responses (pSTAT1, pSTAT3, pSTAT4, pSTAT5, pSTAT6, pSMAD2/3) to all signals (Figure 2).

In order to compute the probability of correct ligand identification, we compute the probability of choosing the hypothesis H_i given the true hypothesis H_j, (Q_{ij}). Then, the overall probability of error, P_e, is given by $P_e = \sum_{j=0}^{M-1} P(H_j) \sum_{i=0, i\neq j}^{M-1} Q_{ij}$ where $\mathrm{P}(H_j)$ is the a priori probability of the hypothesis H_j. The optimal signal detector that minimizes the probability of error is called the maximum likelihood (ML) detector (*17*). Details are provided in the Materials and Methods.

In addition to the overall probability of error, we used the signal-to-noise ratio (SNR) to characterize the specificity of the signaling response. In digital communication, SNR is a measure of fidelity of signal transmission and detection by receivers. There is a direct relationship between SNR and the probability of error in signal identification. Assuming a

communication system with M possible discrete signals transmitted under an additive white Gaussian noise channel, the measured response to a discrete signal m is given by: $x = m + z$ where z is a zero-mean Gaussian random variable with variance σ^2. The SNR is then given by $SNR = \frac{E[m^2]}{\sigma^2}$, where $E[m^2] = \frac{1}{M}\sum_{i=0}^{M-1} m_i^2$ if the signals are equiprobable.

Examination of ligand identification error rates in healthy donors and ER+ breast cancer patients

We applied our communication model to calculate the probability of error and the signal-to-noise ratio, to characterize the specificity of ligand identification for each sample and cell type and compared BC to HD (Figure 3). This analysis reveals how immune cell populations differ in both the signal fidelity (SNR) and probability of signal identification error between BC and HD. Naïve CD4+ T cells and classical monocytes showed the largest error rates, with classical monocytes having lower SNR as compared to CD4+ T cells. Naïve CD8+ T cells showed lower SNR and higher error rates in BC as compared to HD. In contrast, classical monocytes show the same range of SNR for BC and HD but differ by an order of magnitude in the probability of error. Other cell types, such as naïve B cells, show overlapping SNR and error rates, with some healthy donor naïve B cells showing error rates exceeding those of breast cancer patients. NK cells showed the widest range of both SNR and error rates for both HD and BC. We observed increased error rates and reduced SNR for BC as compared to HD for nearly all immune cell subtypes (Figure 3B). SNR and P_e graphs for CD4+, CD8+ T, and B cell subsets are provided in supplementary material (Figure S3-S5).

Immune signaling error profiles

In order to compare the profile of signaling error rates across all cell types in ER+ breast cancer patients, we plotted the SNR and P_e values together on common SNR-P_e axes (Figure 4A). When compared directly, the cells grouped into 4 distinct regions, consisting of combinations of low and high SNR and P_e, with classical monocytes showing low signal and high error, CD4+ T cells and central and effector memory subsets showing high signal and high error, B cells (naïve and memory) showing high signal and low error, and CD8+ T cells (including TCM and TEM) showing low signal and low error. To quantitatively compare these profiles for both BC and HD samples, we performed hierarchical clustering on the negative log transformed P_e values (Figure 4B). This analysis reveals a clear separation between BC and HD in terms of signaling specificity (columns), with a higher overall probability of signal misidentification in BC compared to HD, as well as similarity of cell types (rows), with B, CD4+, and CD8+ cells clustering together. This analysis reveals patterns of signal detection errors across immune cell types that are similar across HD and BC samples.

In addition to the probability of error we examined error rates for pairs of ligands, which gives rise to a confusion matrix. This analysis reveals that IL-2 and IL-4 pair-wise error rates dominate the overall probability of error in naïve CD4+ T cells. In other words, IL-2 and IL-4 are more likely to be confused one for another compared to the other cytokine pairs. The analysis also indicates a trend in elevated pair-wise error rates between IL-2 and IL-4 in CD4+ TCM cells and NK cells (Figures S6-S11). No statistically significant correlations were found between signaling error rates or SNR with patient age, tumor stage, or receptor expression.

Signaling detection alterations induced by JAK inhibition

In order to investigate the clinical relevance of ligand identification error rates and signal fidelity, we treated 10 of the healthy donor samples with a clinically approved JAK1/2 inhibitor,

ruxolitinib. The JAK1/2 inhibitor increased the identification error rate and decreased the signal fidelity (increased noise) in all immune cell subtypes as compared to controls, and in most cases increased the error rate to levels comparable to BC samples with the notable exception of Naïve CD4+ T cells, which showed a drastic decrease in signal fidelity but no change in error rate (Figure 5A). In several cases, the JAK1/2 inhibitor increased the error rate beyond that observed in BC samples, with the largest increase observed in naïve CD8+ T cells.

To compare the immune cell types to each other, we plotted the ruxolitinib samples together on common SNR-P_e axes (Figure 5B). This analysis revealed the stark differences in effect of the ruxolitinib by cell type, with classical monocytes and CD8+ T cells (including naïve, TEM, and TCM subsets) having the largest error rates and smallest SNR. In contrast, CD4+ T cells (including naïve, TEM, and TCM subsets) showed the smallest error rate and maintained a relatively large SNR.

DISCUSSION

We have presented a communication model to quantify signal fidelity and error rates based on digital communication theory and shown that immune cells in the peripheral blood of ER+ breast cancer patients exhibit increased ligand identification error rates and reduced signal fidelity as compared to healthy controls. We observed patterns of ligand detection error rates across immune cells, suggesting that a higher probability of ligand misidentification in BC patients may be an indication of altered immune signaling, potentially leading to immune dysfunction in BC patients. We observed that the inhibition of JAK1/2 also significantly alters both error rates and SNR in healthy donors to a level comparable to those observed in breast cancer patients (*7*). This leads us to hypothesize that BC patients that have lower signal detection error rates could have

better prognoses through improved immune system functioning. In contrast, signaling kinase inhibiting therapies such as JAK inhibitors may alter immune signaling and function in ways beyond simple suppression.

Although prior studies have applied concepts from information and communication theory to characterize cell signaling (*9*, *12*, *13*, *18–23*), this work is the first to provide a measurement of ligand detection probability as a function of the signal-to-noise ratio by integrating the response of multiple transcription factors to multiple stimuli, thus enabling a more holistic view of signaling responses across multiple cell types.

Several prior experimental and theoretical studies have shown that pSTAT abundance and signal transduction peak around 15 minutes following cytokine stimulation (*24–28*). We therefore hypothesized that signal detection accuracy should be maximized, and consequently, the probability of error minimized, at or around this timepoint. We note that signaling errors may be detected and subsequently corrected after STAT phosphorylation events and prior to protein production or other functional events such as proliferation, exhaustion, or microenvironmental factors. As a potential explanation for the signaling errors we observed, we explored the potential connection between receptor expression and SNR and P_e through correlation analysis, as several prior studies have identified receptor dimerization to be a key mechanism of ligand discrimination(*8*, *29–34*). We found no consistent patterns across cell types or receptors for either healthy donors or BC samples that would explain differences observed in either SNR or P_e (Figure S12-S15).

A limitation of this work is the focus on intracellular signaling events, without analysis of their downstream consequences, such as proliferation or T cell exhaustion. As a foundational study, our primary aim was to establish the validity and applicability of this communication modeling framework to the well-characterized first steps of the cell signaling process in a controlled

experimental setting. Because the detection error rate is calculated by integrating the signaling response of 5 cytokine stimulations individually, we interpret the error rate as the probability that a cell incorrectly identifies an isolated signal/ligand and does not consider the case when multiple ligands are presented simultaneously, as is likely the case *in vivo*. Moreover, it is difficult, perhaps technically impossible at the time of this study, to measure the response of a cell to multiple simultaneous signals in an *in vivo* setting. As such, investigation into the functional outcomes of signaling errors, the establishment of a critical error rate, and the application of these methods to an *in vivo* setting remain outstanding questions and form the foundation for future research.

One significant finding of our analysis is the contrast in signaling characteristics between CD4+ and CD8+ T cells. CD4+ T cells, including naïve, effector, and central memory subsets, stand out as having the highest error rates in BC, despite also showing large SNR, and yet, CD4+ T cells appear to be the least sensitive to JAK1/2 inhibition, with a reduced SNR but no change in the detection error rate. An intriguing aspect of this analysis is the pairwise error analysis which reveals the elevated potential for IL-2 versus IL-4 signal misidentification in CD4+ T cells in breast cancer patients. In comparison, CD8+ T cells have a low SNR and low error rate, and yet are extraordinarily sensitive to JAK1/2 inhibition, with orders of magnitude increase in error rate and the most drastic decrease in SNR. These dynamics suggest complex signaling interplay between T cell subsets that may provide avenues for the identification of patients with favorable immune signaling profiles or for therapeutic targeting.

MATERIALS AND METHODS

Human peripheral blood samples

Peripheral blood samples were obtained from breast cancer patients treated at City of Hope National Medical Center in Duarte, California, in compliance with protocols approved by the Institutional Review Board (IRB 21368 and 19186). The study cohort consisted of individuals with newly diagnosed breast cancer, all of whom were estrogen receptor-positive (ER+), progesterone receptor-positive (PR+) and HER2/neu receptor-negative (HER2-), grade Ia-IIb, with a mean age of 57.3 years at diagnosis (range 35-76 years). All samples were obtained prior to treatment. TNM tumor staging and percent positive ki67 staining were collected for each patient. A summary of patient characteristics is provided in Supplemental Table S1. Blood was collected in EDTA-treated tubes. Peripheral blood mononuclear cells were subsequently isolated using Ficoll-Paque density gradient centrifugation (Cytiva, Marlborough, MA, USA), following the manufacturer's protocol. The isolated PBMCs were cryopreserved in a solution containing 10% dimethyl sulfoxide (DMSO) and fetal bovine serum (FBS). Age-matched healthy control samples were acquired from the City of Hope Blood Donor Center.

Cell culture

Cryopreserved PBMCs were carefully thawed and incubated overnight (16 hours) in RPMI 1640 medium, enriched with 10% fetal bovine serum and 1% penicillin-streptomycin-glutamine (PSG), under controlled conditions (37°C, 5% CO2). Cell counts and viability assessments were conducted using a hemocytometer and trypan blue exclusion method (Sigma-Aldrich). Subsequently, the cells were cultured in a 96 deep-well plate at densities ranging from 0.5 to 1 × 10^6 cells/ml in fresh RPMI 1640 medium (Thermo Fisher Scientific Inc., MA, USA).

Cytokine stimulation

Following a resting period, PBMCs were cultured either untreated, stimulated with cytokine alone, or in combination with 0.1 mmol/L ruxolitinib (Cayman Chemical, Ann Arbor, MI, USA), a selective Janus kinase (JAK) 1/2 inhibitor. Cytokines included IFNγ (50 ng/ml), IL-10 (50 ng/ml), IL-2 (50 ng/ml), IL-6 (50 ng/ml), or IL-4 (50 ng/ml) (PeproTech, Rocky Hill, NJ, USA) at 37°C for 15 minutes. Following stimulation, cells were fixed with 1.5% paraformaldehyde (PFA) for 10 minutes at room temperature to preserve cellular structures and signaling intermediates. Fixed cells were then washed with phosphate-buffered saline (PBS) and permeabilized using ice-cold 100% methanol. Methanol-treated cells were stored at -80°C until further analysis. Before antibody staining, the fixed and permeabilized cells were washed three times with staining buffer (PBS supplemented with 1% fetal bovine serum).

Phospho flow cytometry

Phospho flow cytometry was performed using the following antibodies: STAT4-AF647 (clone 38/p-Stat4), CD14-APC-Cy7 (clone HCD14), CD20-AF700 (clone H1), STAT6-V450 (clone 18/pStat6), PD-L1-BV510 (clone 29E.2A3), CD3-BV570 (clone UCHT1), PD1-BV605 (clone EH12.1), CD33-BV750 (clone p67.6), CD27-BV786 (clone L128), CD45RA-BUV395 (clone HI100), CD4-BUV563 (clone SK3), CD16-BUV737 (clone 3G8), CD8-BUV805 (clone SK1), STAT3-AF488 (clone 4/p-Stat3), STAT1-Percp-Cy5.5 (clone 4a), SMAD2/3-PE (clone O72-670), Foxp3-PE-CF594 (clone 259D/C7), and STAT5-PE-Cy7 (clone 47). Antibody dilutions were prepared according to the manufacturer's instructions and optimized through preliminary experiments to achieve optimal staining. The incubation was conducted for 45 minutes at room temperature. All antibodies were sourced from BioLegend (San Diego, CA, USA) or BD Biosciences (Franklin Lakes, NJ, USA).

Data acquisition and gating strategy

Stained cells were analyzed using a Cytek Aurora flow cytometer, equipped with lasers at 355 nm, 405 nm, 488 nm, 561 nm, and 640 nm. Compensation settings were established using single-stain controls along with a negative control. Data acquisition was conducted at a rate of 1000 events per second, with between 50,000 and 100,000 events collected per sample. Gating strategies for cell population identification are provided in Supplementary Material Figures S1, S2.

Data processing

Because we observed that pSTAT or pSMAD measurements did not always follow a Gaussian distribution, with their observed distribution tending closer to a log-normal distribution, all pSTAT/pSMAD measurements were first transformed with a log-like transformation (eq. 1). Then the transformed measurements were fit to a Gaussian mixture (GM) with two components. Each component is a multivariate normal distribution, where each variate corresponds to one of the 6 signaling molecules (5 pSTATs and pSMAD2/3). The log-transformed measurements together with knowledge of the GM parameters were fed to the optimal detector in a Monte-Carlo style simulation to obtain the probability of correct signal detection, or equivalently, the average probability of error in each analyzed signaling system. Here, one signaling system refers to the signaling pathway corresponding to one HD or BC patient and one cell type.

All pSTAT/pSMAD response data were log transformed as:

$$x_i(k) = \log(d_i(k) - \alpha_i(k) + 1)\,, 0 \le i \le M-1,\ \ 1 \le k \le K \qquad (1)$$

where $d_i(k)$ is the measured response of the kth pSTAT/pSMAD to the ith treatment, $M = 6$ is the number of treatments (5 cytokines and untreated case), and $\alpha_i(k)$ is a parameter found by taking the minimum over all $d_i(k)$ measurements belonging to a given cell type and donor

category (HD or BC). A global offset $\alpha_i(k) - 1$ was introduced to ensure the argument of the log function is greater or equal to one, since $d_i(k) - \alpha_i(k) \geq 0$.

Measurement data was organized in comma separated value (.csv) files, one file per subject, cytokine treatment, and cell type, provided in supplemental materials. Given the large amount of data, the measurements were consolidated in structure arrays, one per HD/BC subject, cell type, and cytokine treatment. Each field in the structure array is a matrix with $K = 6$ columns (one column for each pSTAT/pSMAD) and N_i rows, where N_i is the number of gated cells for the *i*th cytokine treatment.

Gaussian mixture modeling

GM modeling assumes that the log-transformed responses have a multi-variate Gaussian mixture distribution with two components. That is, the probability density function (PDF) of $\boldsymbol{x_i} = [x_i(1), x_i(2), \ldots x_i(K)]$ is given by:

$$p(\boldsymbol{x_i}) = \rho_i \aleph\left(\boldsymbol{m}_{1,i}, \boldsymbol{\Sigma}_{1,i}\right) + (1 - \rho_i)\aleph\left(\boldsymbol{m}_{2,i}, \boldsymbol{\Sigma}_{2,i}\right), \quad (2)$$

where $\aleph\left(\boldsymbol{m}_{1,i}, \boldsymbol{\Sigma}_{1,i}\right)$ and $\aleph\left(\boldsymbol{m}_{2,i}, \boldsymbol{\Sigma}_{2,i}\right)$ are multi-variate Gaussians with means $\boldsymbol{m}_{1,i}$ and $\boldsymbol{m}_{2,i}$, and covariance matrices $\boldsymbol{\Sigma}_{1,i}$ and $\boldsymbol{\Sigma}_{2,i}$ respectively, and $0 \leq \rho_i \leq 1$ is the mixing parameter. The means, covariance matrices, and mixing parameters were obtained from the log-transformed measurements using the MATLAB function "*fitgmdist*," which implements the Expectation-Maximization algorithm.

Probability of error computation

In order to compute the probability of error, let $Q_{ij} = P(H_i|H_j)$ be the probability of choosing the hypothesis H_i given the true hypothesis H_j. Then, the overall probability of error, P_e, is given by $P_e = \sum_{j=0}^{M-1} P(H_j) \sum_{i=0, i\neq j}^{M-1} Q_{ij}$ where $\mathrm{P}(H_j)$ is the a priori probability of the hypothesis H_j. The optimal signal detector that minimizes the probability of error under the Bayes strategy is one that maximizes the probability of observed data under the hypothesis H_i. More precisely, suppose that $\boldsymbol{x}$ is the multi-variate observation and $p(\boldsymbol{x}|H_i)$ are the conditional probability density functions (PDFs) of $\boldsymbol{x}$ under H_i. Moreover, assume equiprobable hypotheses, i.e., $P(H_i) = 1/M$, which is a reasonable assumption in the absence of prior information on the frequencies of H_i. Then, the optimal detector decides H_i if $p(\boldsymbol{x}|H_i)$ is maximized. That is, it decides H_i if $p(\boldsymbol{x}|H_i) > p(\boldsymbol{x}|H_j)$, for all $j \neq i$. This detector is called the maximum likelihood detector (*17*).

To evaluate the average error probability for one sample and one cell type, all the log-transformed pSTAT/pSMAD responses for all hypotheses (M matrices $\boldsymbol{X}_i$ of size $N_i \times K, 0 \leq i \leq M-1$) were fed into a simulator that implements the ML detector and accumulates the number of detection errors over all the simulated trials ($N = \sum_{i=0}^{M-1} N_i$). In this simulator implementation, one trial corresponds to a single cell response for each hypothesis, i.e. a row in the $\boldsymbol{X}_i$ matrices. The simulation loops over each hypothesis $H_j, 0 \leq j \leq M-1$ and each trial $n, 1 \leq n \leq N_j$, and computes the conditional probability of the hypothesis H_j (the measured response $\boldsymbol{X}_j(n)$) given the hypothesis $H_i, 0 \leq i \leq M-1$. With the GM assumption, this conditional probability is given by:

$$p(H_j|H_i) = \rho_i p(\boldsymbol{X}_j(n)|\boldsymbol{m}_{1,i}) + (1-\rho_i)(\boldsymbol{X}_j(n)|\boldsymbol{m}_{2,i})$$

where

$$p(\boldsymbol{X}_j(n)|\boldsymbol{m}_{1,i}) = \frac{1}{\sqrt{(2\pi)^M \det(\boldsymbol{\Sigma}_{1,i})}} exp\left\{-\frac{1}{2}(\boldsymbol{X}_j(n) - \boldsymbol{m}_{1,i})\boldsymbol{\Sigma}_{1,i}^{-1}(\boldsymbol{X}_j(n) - \boldsymbol{m}_{1,i})^T\right\}$$

$$p(\boldsymbol{X}_j(n)|\boldsymbol{m}_{2,i}) = \frac{1}{\sqrt{(2\pi)^M \det(\boldsymbol{\Sigma}_{2,i})}} exp\left\{-\frac{1}{2}(\boldsymbol{X}_j(n) - \boldsymbol{m}_{2,i})\boldsymbol{\Sigma}_{2,j}^{-1}(\boldsymbol{X}_j(n) - \boldsymbol{m}_{2,i})^T\right\}$$

$\boldsymbol{m}_{1,i}$ and $\boldsymbol{m}_{2,i}$ are the signals corresponding to the hypothesis H_i and $\boldsymbol{\Sigma}_{1,i}$ and $\boldsymbol{\Sigma}_{2,i}$ are the noise covariance matrices corresponding to the hypothesis H_i.

Then, the ML detector selects the hypothesis H_{i^*} such that the conditional probability $p(H_j|H_i)$ is maximized:

$$i^* = argmax\left\{p(\boldsymbol{X}_j(n)|H_i), 0 \leq i \leq M-1\right\}.$$

If $i^* \neq i$, we declare a detection error and a counter in a $M \times M$ error matrix $\boldsymbol{Q}$ is incremented at the index (i^*, i). Otherwise, a correct decision is made and the counter at the index (i, i) is incremented in the error matrix. Finally, the conditional probability Q_{ij}, the probability of choosing the hypothesis H_i given the true hypothesis H_j is estimated from $Q_{ij} = \frac{\boldsymbol{Q}(i,j)}{N}$, and the overall probability of error P_e is given by $P_e = 1 - P_c = 1 - \sum_{i=0}^{M-1} P(H_i)Q_{ii}$. Note that the confusion matrix is the matrix having the element indexed by *i*th row and *j*th column equal to the conditional probabilities Q_{ij}.

Signal-to-noise ratio computation

SNR is a dimensionless ratio of signal power P_S to noise power P_N. We extend the SNR definition to the GM mixture model as follows. We assume a communication system with M possible signals which is described by the following relationship:

$$x = \begin{cases} m_1 + z_1 & \text{with probability } \rho \\ m_2 + z_2 & \text{with probability } 1-\rho \end{cases}$$

That is, under the hypothesis H_i, the transmitted signal is equal to $m_{1,i}$ with probability ρ_i and it is equal to $m_{2,i}$ with probability $1-\rho_i$. Similarly, under the hypothesis H_i, the noise is z_{1i} with probability ρ_i and it is z_{2i} with probability $1-\rho_i$, where $z_{1,i}$ is a zero-mean Gaussian random variable with variance $\sigma_{1,i}^2$ and z_{2i} is a zero-mean Gaussian random variable with variance $\sigma_{2,i}^2$.

Then, for a Gaussian mixture with two components, the SNR can be expressed as:

$$SNR = \frac{\sum_{i=0}^{M-1}(\rho_i m_{1,i}^2 + (1-\rho_i)m_{2,i}^2)}{\sum_{i=0}^{M-1}(\rho_i \sigma_{1,i}^2 + (1-\rho_i)\sigma_{2,i}^2)}.$$

Finally, we generalize the above expression for the multivariate signal case, where under the hypothesis H_i, the transmitted signal is equal to $\boldsymbol{m}_{1,i}$ with probability ρ_i and it is equal to $\boldsymbol{m}_{2,i}$ with probability $1-\rho_i$. The noise corrupting the signal $\boldsymbol{m}_{1,i}$ is a zero-mean multivariate Gaussian noise with covariance matrix $\boldsymbol{\Sigma}_{1,i}$, and the noise corrupting the signal $\boldsymbol{m}_{2,i}$ is a zero-mean multivariate Gaussian noise with covariance matrix $\boldsymbol{\Sigma}_{2,i}$. Under these assumptions, the SNR becomes:

$$SNR = \frac{\sum_{i=0}^{M-1}(\rho_i \boldsymbol{m}_{1,i}\boldsymbol{m}_{1,i}^T + (1-\rho_i)\boldsymbol{m}_{2,i}\boldsymbol{m}_{2,i}^T)}{\sum_{i=0}^{M-1}(\rho_i Trace(\boldsymbol{\Sigma}_{1,i}) + (1-\rho_i)Trace(\boldsymbol{\Sigma}_{2,i}))}.$$

All communication model simulations and analyses were performed in MATLAB. Correlation and statistical analyses were performed in Prism Graphpad 10.4.0. Note that it is customary to express SNR in decibels (dB), where the relation between linear SNR and SNR in dB is given by $SNR_{dB} = 10\log_{10} SNR$.

References

1. L. Wang, D. L. Simons, X. Lu, T. Y. Tu, C. Avalos, A. Y. Chang, F. M. Dirbas, J. H. Yim, J. Waisman, P. P. Lee, Breast cancer induces systemic immune changes on cytokine signaling in peripheral blood monocytes and lymphocytes. *EBioMedicine* **52**, 102631 (2020).

2. A. L. Wang, A. K. Miyahira, D. L. Simons, X. Lu, Y. Andrew, C. Wang, M. A. Suni, V. C. Maino, F. M. Dirbas, J. Yim, J. Waisman, P. P. Lee, IL-6 Signaling in Peripheral Blood T Cells Predicts Clinical Outcome in Breast Cancer. *Cancer Res.*, 1119–1127 (2017).

3. R. J. Critchley-Thorne, D. L. Simons, N. Yan, A. K. Miyahira, F. M. Dirbas, D. L. Johnson, S. M. Swetter, R. W. Carlson, G. A. Fisher, A. Koong, S. Holmes, P. P. Lee, Impaired interferon signaling is a common immune defect in human cancer. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9010–9015 (2009).

4. D. A. Harrison, The Jak/STAT pathway. *Cold Spring Harb. Perspect. Biol.* **4**, a011205–a011205 (2012).

5. G. R. Stark, J. E. Darnell, The JAK-STAT Pathway at Twenty. *Immunity* **36**, 503–514 (2012).

6. W. X. Li, Canonical and non-canonical JAK-STAT signaling. *Trends Cell Biol.* **18**, 545–551 (2008).

7. D. Moodley, H. Yoshida, S. Mostafavi, N. Asinovski, A. Ortiz-Lopez, P. Symanowicz, J.-B. Telliez, M. Hegen, J. D. Clark, D. Mathis, C. Benoist, Network pharmacology of JAK inhibitors. *Proceedings of the National Academy of Sciences* **113**, 201610253 (2016).

8. N. Nandagopal, L. A. Santat, L. LeBon, D. Sprinzak, M. E. Bronner, M. B. Elowitz, Dynamic ligand discrimination in the Notch signaling pathway. *Cell* **172**, 869-880.e19 (2018).

9. Y. Tang, A. Adelaja, F. X.-F. Ye, E. Deeds, R. Wollman, A. Hoffmann, Quantifying information accumulation encoded in the dynamics of biochemical signaling. *Nat. Commun.* **12**, 1272 (2021).

10. A. Emadi, T. Lipniacki, A. Levchenko, A. Abdi, "A Decision Making Model Where the Cell Exhibits Maximum Detection Probability: Statistical Signal Detection Theory and Molecular Experimental Data" in *2023 57th Annual Conference on Information Sciences and Systems (CISS)* (2023), pp. 1–4.

11. C. Waltermann, E. Klipp, Information theory based approaches to cellular signaling. *Biochim. Biophys. Acta* **1810**, 924–932 (2011).

12. A. Rhee, R. Cheong, A. Levchenko, The application of information theory to biochemical signaling systems. *Phys. Biol.* **9**, 045011 (2012).

13. I. Habibi, R. Cheong, T. Lipniacki, A. Levchenko, E. S. Emamian, A. Abdi, Computation and measurement of cell decision making errors using single cell data. *PLoS Comput. Biol.* **13**, 1–17 (2017).

14. M. T. Laub, R. Sender, S. Fuchs, R. Milo, W. A. Lim, B. Mayer, T. Pawson, E. J. Capra, M. T. Laub, M. T. Laub, M. Goulian, J. M. Skerker, M. S. Prasol, B. S. Perchuk, E. G. Biondi, M. T. Laub, J. M. Skerker, B. S. Perchuk, A. Siryaporn, E. A. Lubin, O. Ashenberg, M. Goulian, Y. Gotoh, Y. Eguchi, T. Watanabe, S. Okamoto, A. Doi, R. Utsumi, S. K. Checa, M. D. Zurbriggen, F. C. Soncini, P. Creixell, A. Palmeri, C. J. Miller, H. J. Lou, C. C. Santini, M. Nielsen, P. Creixell, E. M. Schoof, C. D. Simpson, J. Longden, C. J. Miller, H. J. Lou, Keeping

Signals Straight: How Cells Process Information and Make Decisions. *PLoS Biol.* **14**, e1002519 (2016).

15. A. Cui, T. Huang, S. Li, A. Ma, J. L. Pérez, C. Sander, D. B. Keskin, C. J. Wu, E. Fraenkel, N. Hacohen, Dictionary of immune responses to cytokines at single-cell resolution. *Nature*, 1–8 (2023).

16. C. E. Shannon, A Mathematical Theory of Communication. *Bell System Technical Journal*, doi: 10.1002/j.1538-7305.1948.tb01338.x (1948).

17. M. Salehi, J. Proakis, *Digital Communications* (McGraw-Hill Education, 2007).

18. A. Karolak, S. Branciamore, J. S. McCune, P. P. Lee, A. S. Rodin, R. C. Rockne, Concepts and Applications of Information Theory to Immuno-Oncology. *Trends Cancer Res.* **7**, 335–346 (2021).

19. A. Adelaja, B. Taylor, K. M. Sheu, Y. Liu, S. Luecke, A. Hoffmann, Six distinct NFκB signaling codons convey discrete information to distinguish stimuli and enable appropriate macrophage responses. *Immunity* **54**, 916-930.e7 (2021).

20. A. Gambin, A. Charzyńska, A. Ellert-Miklaszewska, M. Rybiński, Computational models of the JAK1/2-STAT1 signaling. *Jak-Stat* **2**, e24672 (2013).

21. M. Ozen, T. Lipniacki, A. Levchenko, E. S. Emamian, A. Abdi, Modeling and measurement of signaling outcomes affecting decision making in noisy intracellular networks using machine learning methods. *Integr. Biol. (Camb.)* **12**, 122–138 (2020).

22. F. Schaper, T. Jetka, A. Dittrich, Decoding cellular communication: An information theoretic perspective on cytokine and endocrine signaling. *Curr. Opin. Endocr. Metab. Res.* **24**, 100351 (2022).

23. M. Son, A. G. Wang, H.-L. Tu, M. O. Metzig, P. Patel, K. Husain, J. Lin, A. Murugan, A. Hoffmann, S. Tay, NF-κB responds to absolute differences in cytokine concentrations. *Sci. Signal.* **14** (2021).

24. S. H. Ross, C. Rollings, K. E. Anderson, P. T. Hawkins, L. R. Stephens, D. A. Cantrell, Phosphoproteomic Analyses of Interleukin 2 Signaling Reveal Integrated JAK Kinase-Dependent and -Independent Networks in CD8+ T Cells. *Immunity* **45**, 685–700 (2016).

25. S. Yamada, S. Shiono, A. Joo, A. Yoshimura, Control mechanism of JAK/STAT signal transduction pathway. *FEBS Lett.* **534**, 190–196 (2003).

26. T. Quaiser, A. Dittrich, F. Schaper, M. Mönnigmann, A simple work flow for biologically inspired model reduction--application to early JAK-STAT signaling. *BMC Syst. Biol.* **5**, 30 (2011).

27. M. Zhu, S. Pleasic-Williams, T. H. Lin, D. A. Wunderlich, J. B. Cheng, J. L. Masferrer, pSTAT3: a target biomarker to study the pharmacology of the anti-IL-21R antibody ATR-107 in human whole blood. *J. Transl. Med.* **11**, 65 (2013).

28. N. Cheemalavagu, K. E. Shoger, Y. M. Cao, B. A. Michalides, S. A. Botta, J. R. Faeder, R. A. Gottschalk, Predicting gene-level sensitivity to JAK-STAT signaling perturbation using a mechanistic-to-machine learning framework. *cels* **0** (2024).

29. S. Fathi, C. R. Nayak, J. J. Feld, A. G. Zilman, Absolute ligand discrimination by dimeric signaling receptors. *Biophys. J.* **111**, 917–920 (2016).

30. W.-L. Lo, A. Weiss, Adapting T cell receptor ligand discrimination capability via LAT. *Front. Immunol.* **12**, 673196 (2021).

31. P. Binder, N. D. Schnellbächer, T. Höfer, N. B. Becker, U. S. Schwarz, Optimal ligand discrimination by asymmetric dimerization and turnover of interferon receptors. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2103939118 (2021).

32. O. Feinerman, R. N. Germain, G. Altan-Bonnet, Quantitative challenges in understanding ligand discrimination by alphabeta T cells. *Mol. Immunol.* **45**, 619–631 (2008).

33. C. Sweeney, K. L. Carraway 3rd, Ligand discrimination by ErbB receptors: differential signaling through differential phosphorylation site usage. *Oncogene* **19**, 5568–5573 (2000).

34. G. Gaud, S. Achar, F. X. P. Bourassa, J. Davies, T. Hatzihristidis, S. Choi, T. Kondo, S. Gossa, J. Lee, P. Juneau, N. Taylor, C. S. Hinrichs, D. B. McGavern, P. François, G. Altan-Bonnet, P. E. Love, CD3ζ ITAMs enable ligand discrimination and antagonism by inhibiting TCR signaling in response to low-affinity peptides. *Nat. Immunol.* **24**, 2121–2134 (2023).

Acknowledgments: The authors acknowledge the participation of the blood donors and breast cancer patients that made this work possible and thank them for their contributions to the study.

Funding: Research reported in this publication included work performed in the Analytical Cytometry and Biostatistics and Mathematical Oncology Shared Resources supported by the National Cancer Institute of the National Institutes of Health under grant numbers P30CA033572, U01CA232216. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions:

Conception and design: A. Matache, R.C. Rockne, P.P. Lee, S. Branciamore, A. Rodin

Development of methodology: A. Matache, R.C. Rockne

Acquisition of data: J. Lima Jr, P.P. Lee

Analysis and interpretation of data: A. Matache, R.C. Rockne, J. Lima Jr., M. Kuznetsov, K. Urbaniak, S. Branciamore, A. Rodin, P.P. Lee

Writing, review, and/or revision of the manuscript: A. Matache, R.C. Rockne, J. Lima Jr., M. Kuznetsov, K. Urbaniak, S. Branciamore, A. Rodin, P.P. Lee

Administrative, technical, or material support: R.C. Rockne, A. Rodin, P.P. Lee

Study supervision: R.C. Rockne, A. Rodin, S. Branciamore, P.P. Lee

Competing interests: None

Data and materials availability: Aggregate data and computational codes to reproduce all results are provided in supplementary material.

Figures

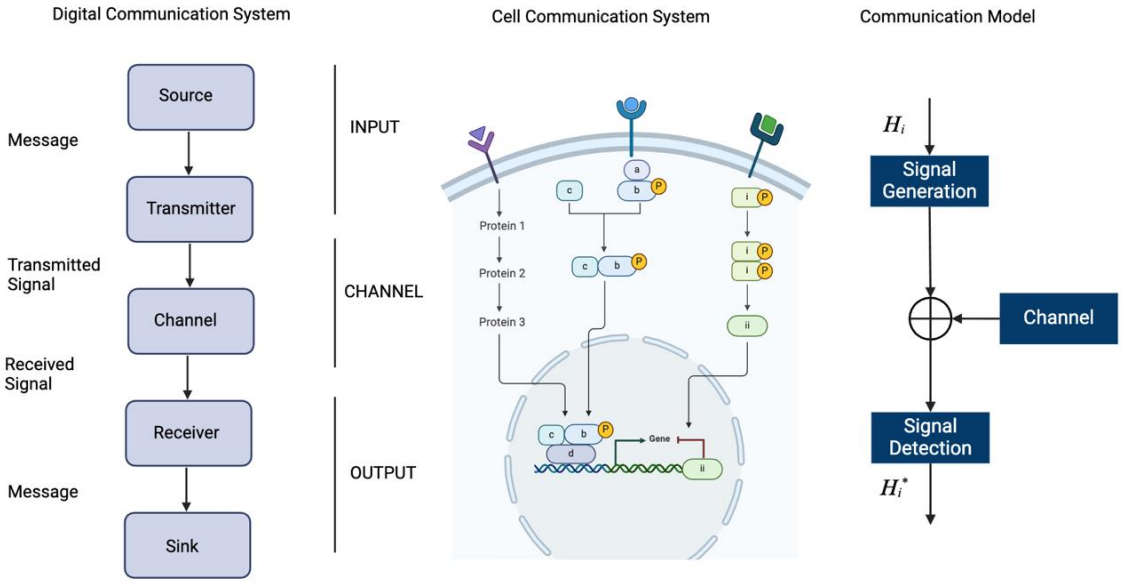


Fig. 1. Elements of a communication system and signal detection model. A digital communication system consisting of source, transmitter, channel, receiver, and sink is used to study a cellular communication system consisting of surface receptors, signaling pathways, transcription factors, and target gene expression. The communication system can be modeled as a signal detection problem consisting of signal generation and transmission through a channel before detection. Figure created with Biorender.

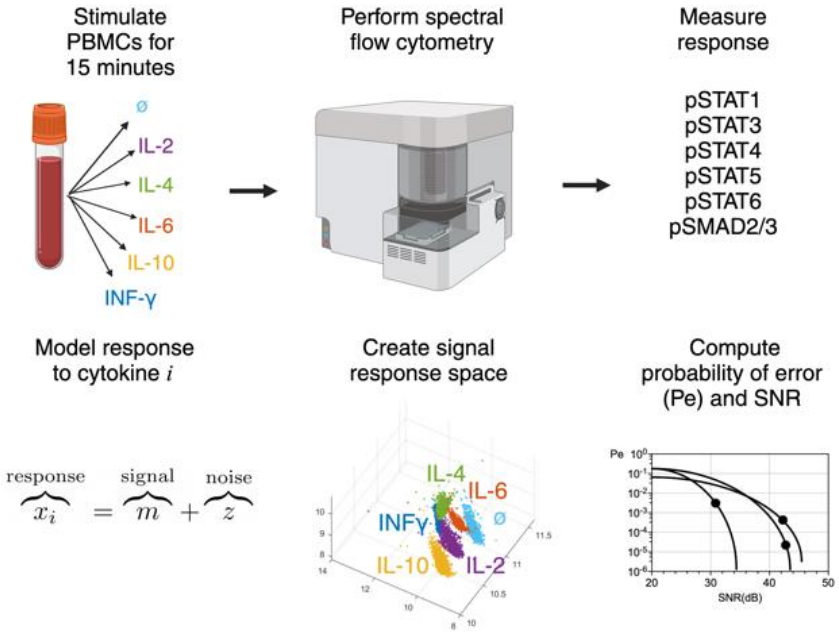


Fig. 2. Illustration of experiment, measurement, modeling, and analysis process. Peripheral blood mononuclear cells from healthy donors and ER+ breast cancer patients were subjected to stimulation with one of 5 cytokines for 15 minutes. Cell surface markers and intracellular proteins were analyzed with spectral flow cytometry to establish cell identity and to measure response to cytokine stimulation. The response to any one of the cytokines is modeled as a sum of signal and noise. All stimulations and responses are combined to create a 6-dimensional signal response space, illustrated here in 3 dimensions. The probability of error, or signal misidentification (P_e), and signal-to-noise ratio (SNR) are computed from the signal response space for each sample and compared across cell types, healthy donors, and breast cancer patient samples. Figure created with Biorender.

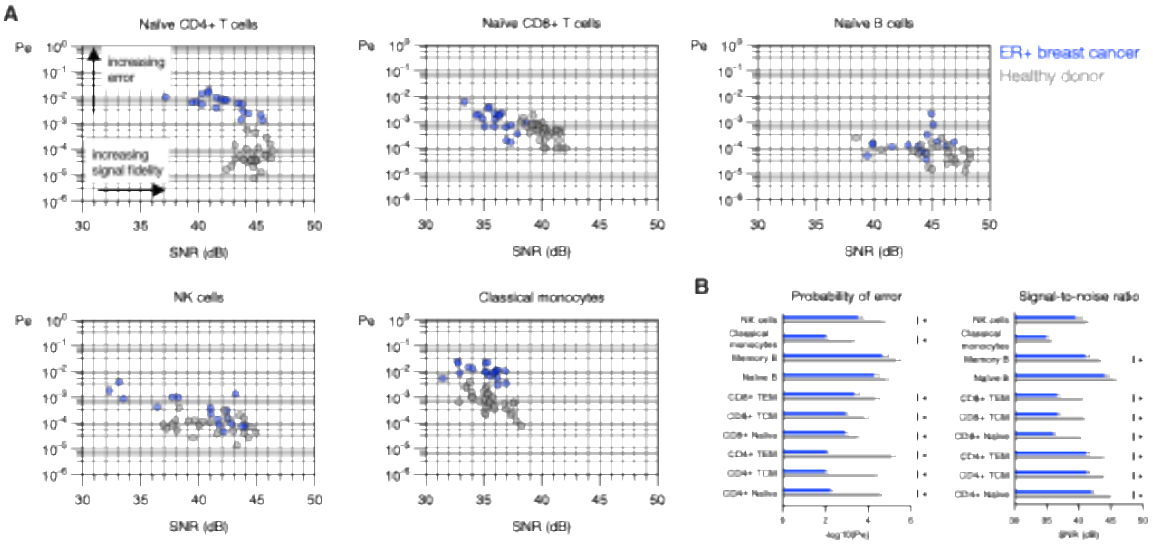


Fig. 3. Probability of error and signal-to-noise ratio are altered in peripheral blood immune cells in ER+ breast cancer patients as compared to healthy donors. A. Signal detection is characterized by P_e and SNR for naïve CD4+, CD8+, T cells, naïve B cells, NK cells, and classical monocytes in peripheral blood samples from 19 ER+ breast cancer patients and 32 healthy donors. Each datapoint corresponds to one healthy donor (grey) or ER+ breast cancer (blue) sample, integrating 6 different phosphorylation events from each of the 5 cytokine stimulations and baseline (no stimulation). **B.** Pairwise comparisons of the negative log transformed probability of error ($-\log_{10}(P_e)$) and SNR for HD and BC for each cell subtype (unpaired t-tests adjusted for multiple comparisons, $* p < 0.05$). For illustration in bar graphs, samples with $P_e = 0$ have $-\log_{10}(P_e) := 6$.

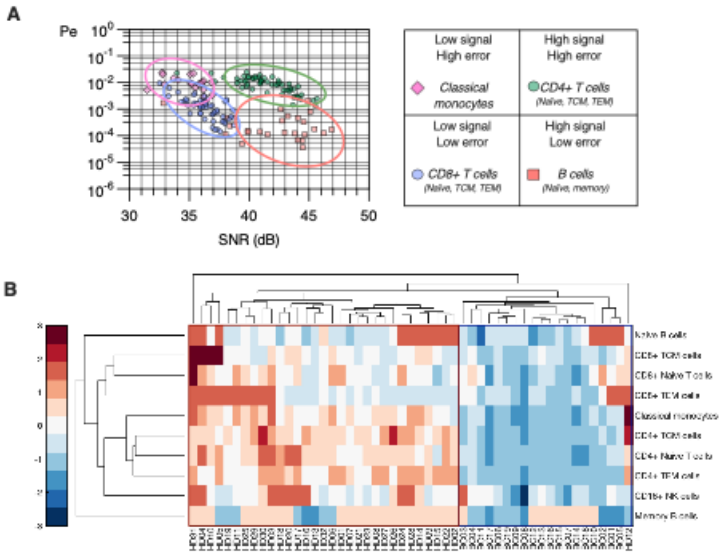


Fig. 4. Patterns of signaling error rates and SNR across immune cells in ER+ breast cancer.

A. Immune cells cluster in P_e and SNR by type and can be grouped into low and high signal and error combinations. **B.** Hierarchical clustering of signaling error across immune cell subtypes reveals patterns of error rates (columns) and increased error rates across immune cell subtypes (rows) that distinguish breast cancer samples from healthy donors. Colorbar indicates row-wise z-score.

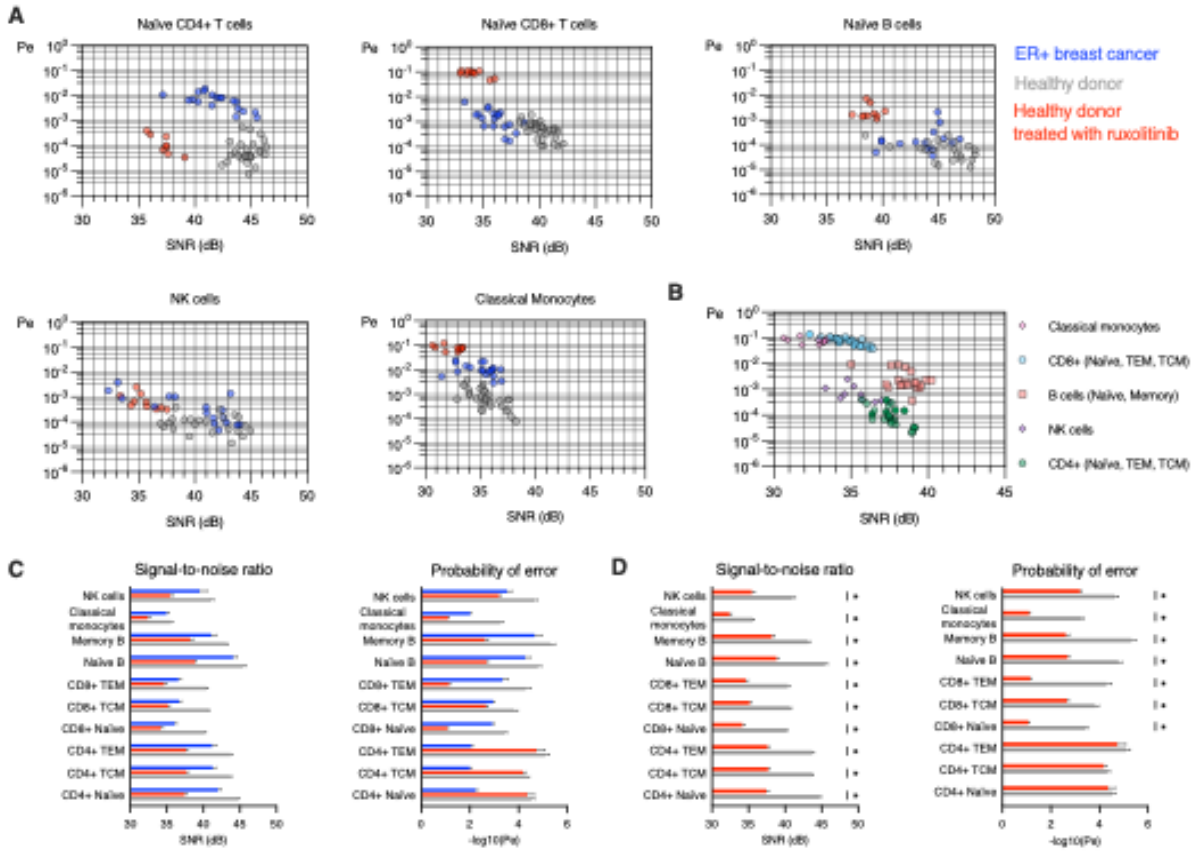
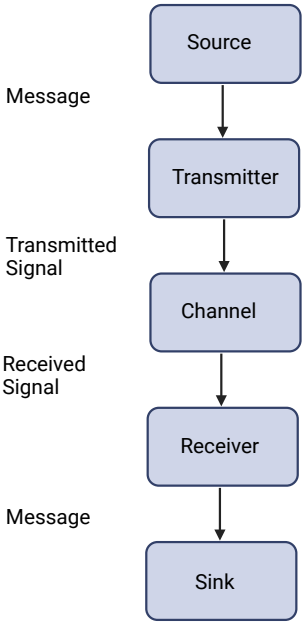
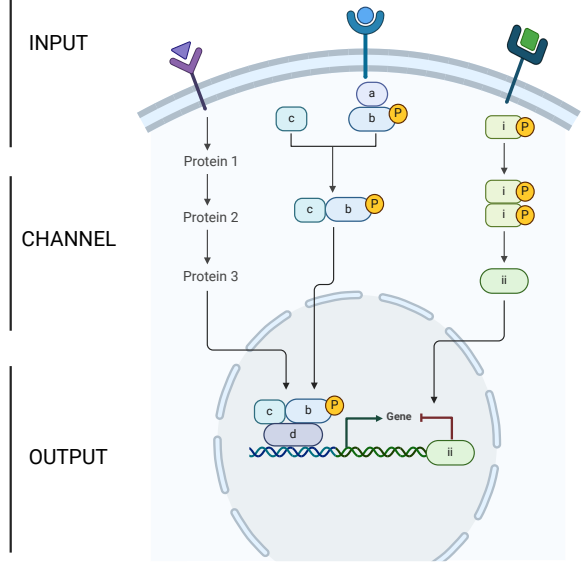


Fig. 5. JAK1/2 inhibition induces error rates and reduces SNR in healthy donors to levels comparable to breast cancer. A. Immune cell subtypes for ER+ breast cancer samples (blue), healthy donors (grey) and a subset of 10 healthy donor samples also treated with ruxolitinib (red). Signal detection error rates are increased, and signal fidelity is decreased in all cell subtypes except for the error rate in naïve CD4+ T cells. **B.** Plotting the 10 samples treated with ruxolitinib on the same SNR-P_e axes reveal relative error and signal fidelity characteristics. **C.** Comparisons of SNR and error rates for all cell subtypes in BC, HD, and HD+rux, (left) and comparison of HD and HD+rux (right) (unpaired t-tests adjusted for multiple comparisons, $* p < 0.05$). For illustration in the bar graphs, samples with $P_e = 0$ have $-\log_{10}(P_e) := 6$.

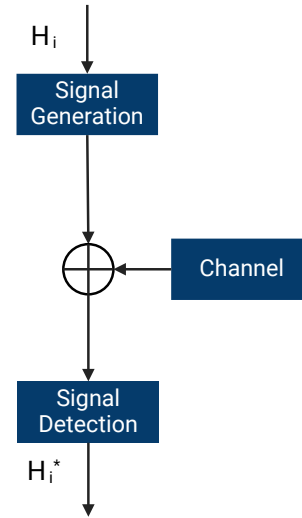
Digital Communication System



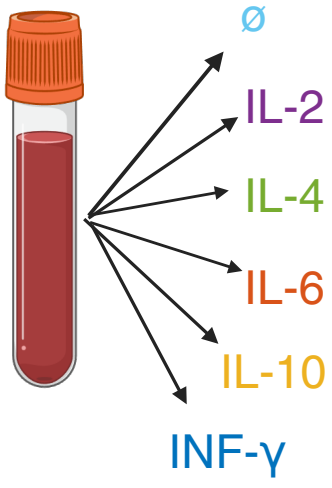
Cell Communication System



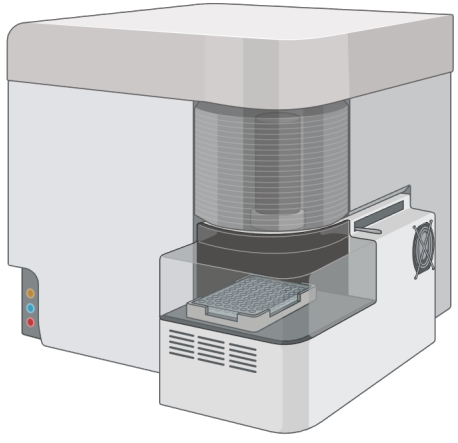
Communication Model



Stimulate PBMCs for 15 minutes



Perform spectral flow cytometry



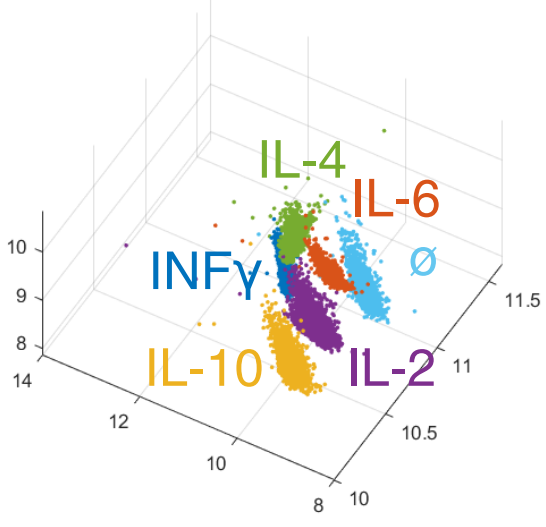
Measure response

pSTAT1
pSTAT3
pSTAT4
pSTAT5
pSTAT6
pSMAD2/3

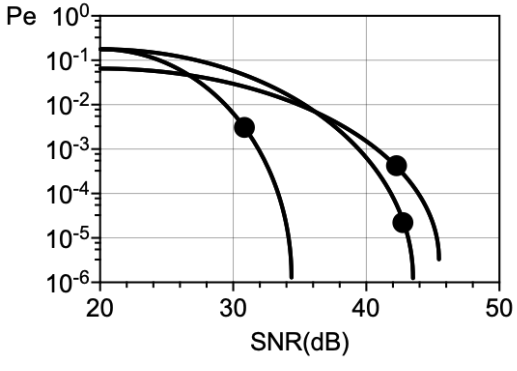
Model response to cytokine i

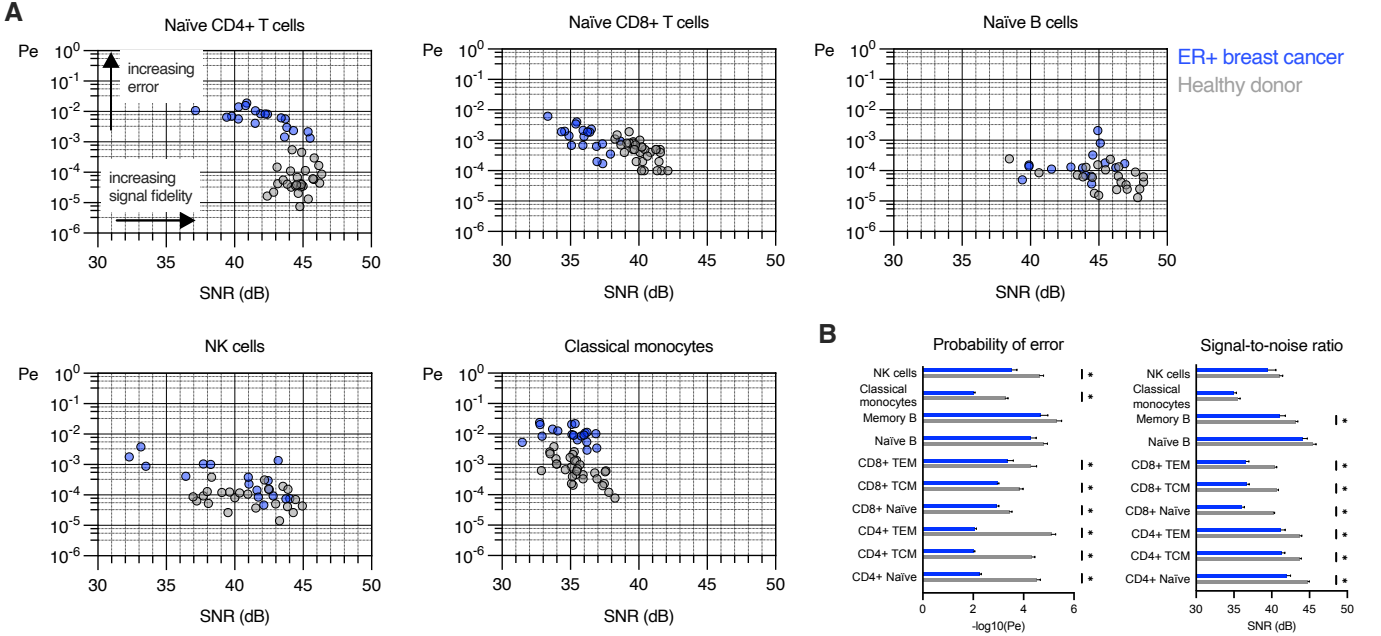
$$\overbrace{x_i}^{\text{response}} = \overbrace{m}^{\text{signal}} + \overbrace{z}^{\text{noise}}$$

Create signal response space

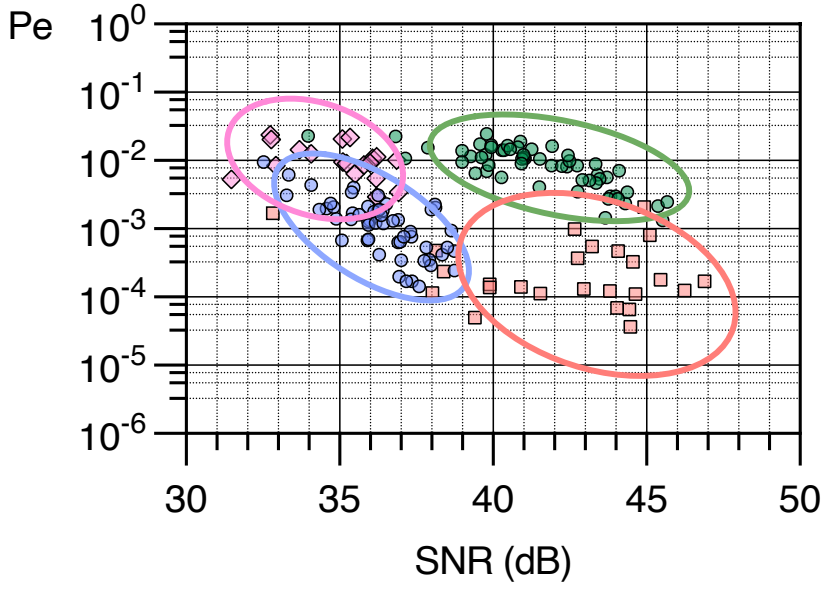


Compute probability of error (Pe) and SNR



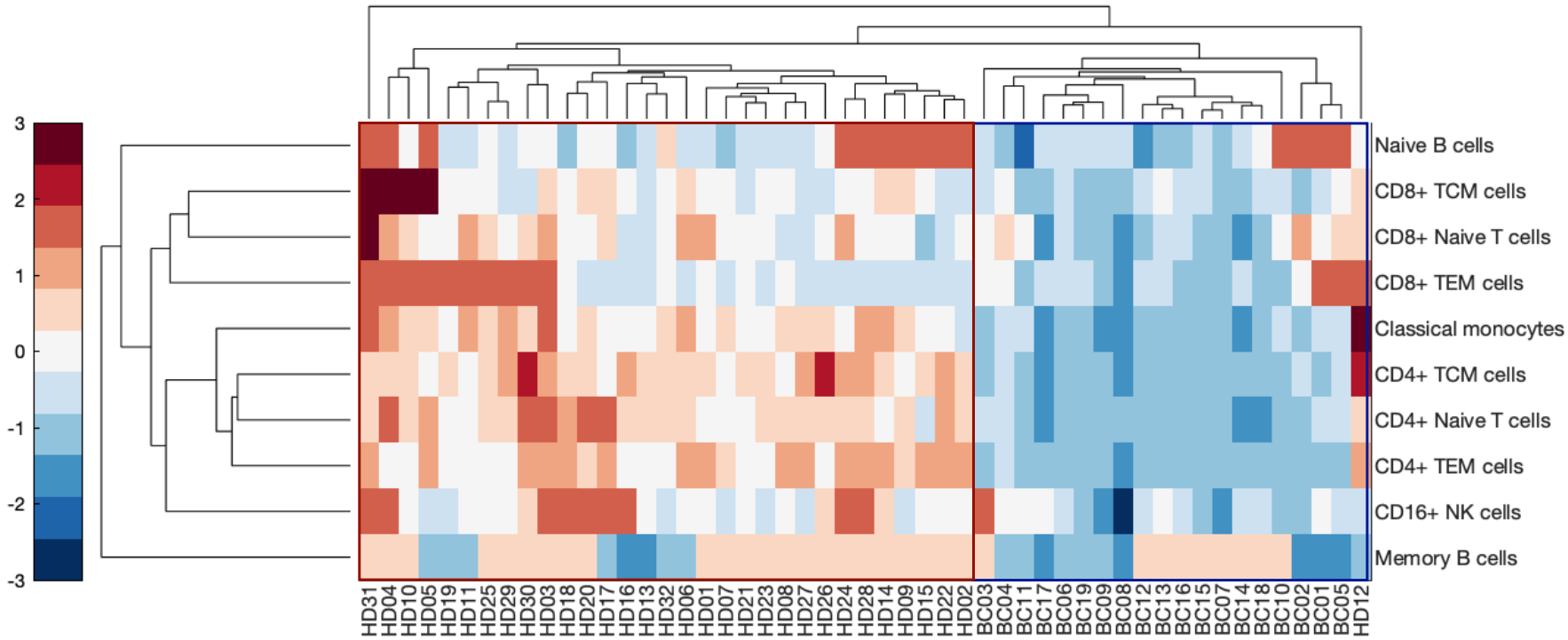

A
Naïve CD4+ T cells
Naïve CD8+ T cells
Naïve B cells
NK cells
Classical monocytes
Pe
SNR (dB)
increasing error
increasing signal fidelity
ER+ breast cancer
Healthy donor
B
Probability of error
Signal-to-noise ratio
NK cells
Classical monocytes
Memory B
Naïve B
CD8+ TEM
CD8+ TCM
CD8+ Naïve
CD4+ TEM
CD4+ TCM
CD4+ Naïve
-log10(Pe)
SNR (dB)

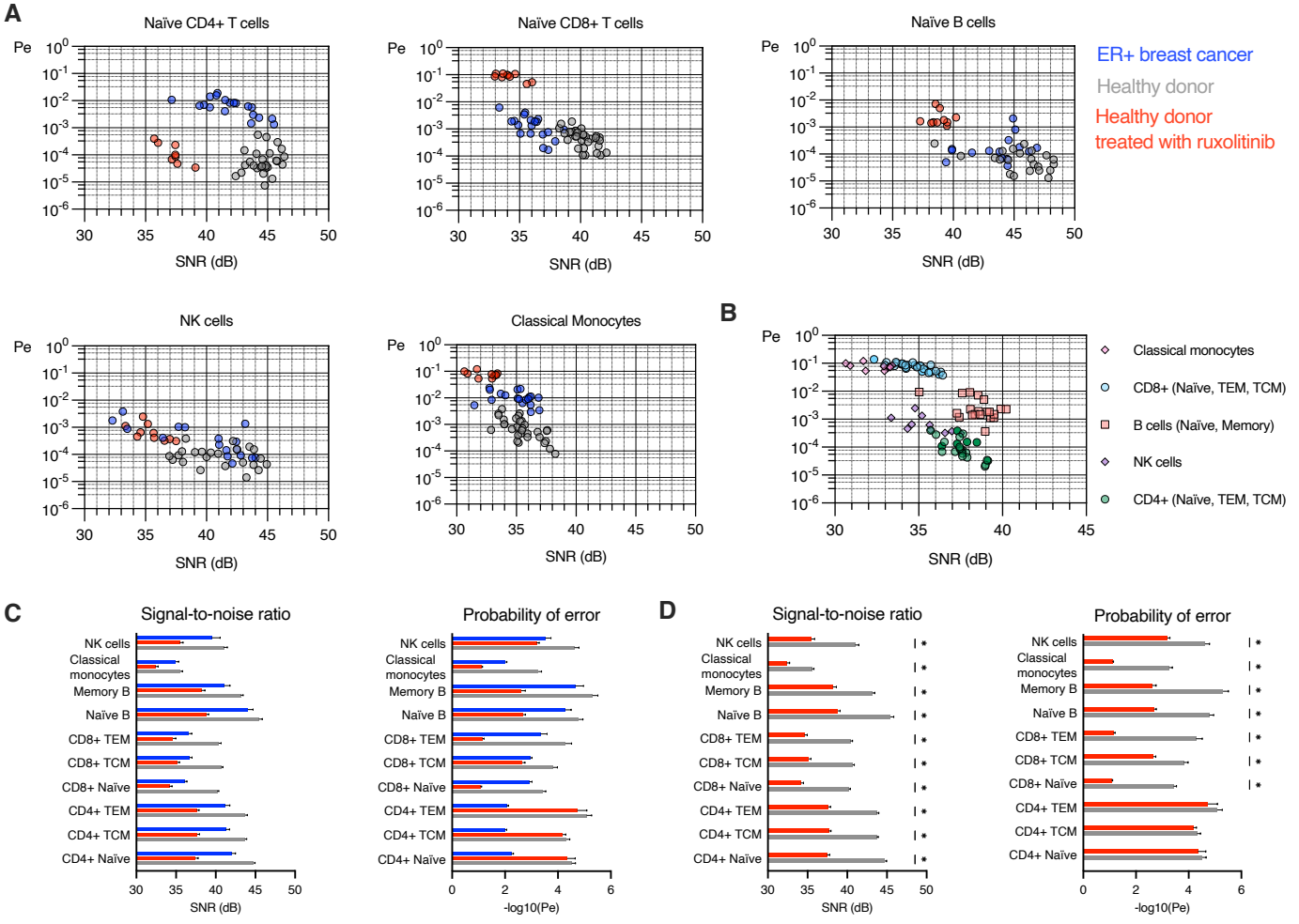
A



Low signal High error ◆ *Classical monocytes*	High signal High error ● *CD4+ T cells* *(Naïve, TCM, TEM)*
Low signal Low error ● *CD8+ T cells* *(Naïve, TCM, TEM)*	High signal Low error ■ *B cells* *(Naïve, memory)*

B




A
Naïve CD4+ T cells
Naïve CD8+ T cells
Naïve B cells
NK cells
Classical Monocytes
Pe
SNR (dB)
ER+ breast cancer
Healthy donor
Healthy donor
treated with ruxolitinib
B
Classical monocytes
CD8+ (Naïve, TEM, TCM)
B cells (Naïve, Memory)
NK cells
CD4+ (Naïve, TEM, TCM)
C
Signal-to-noise ratio
Probability of error
NK cells
Classical monocytes
Memory B
Naïve B
CD8+ TEM
CD8+ TCM
CD8+ Naïve
CD4+ TEM
CD4+ TCM
CD4+ Naïve
SNR (dB)
-log10(Pe)
D
Signal-to-noise ratio
Probability of error
*
SNR (dB)
-log10(Pe)