UCLA
UCLA Previously Published Works

Title
Quantifying information of intracellular signaling: progress with machine learning

Permalink
https://escholarship.org/uc/item/3mw2s57j

Journal
Reports on Progress in Physics, 85(8)

ISSN
0034-4885

Authors
Tang, Ying
Hoffmann, Alexander

Publication Date
2022-08-01

DOI
10.1088/1361-6633/ac7a4a

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Rep Prog Phys. Author manuscript; available in PMC 2023 July 12.

Published in final edited form as:

Rep Prog Phys. ; 85(8): . doi:10.1088/1361-6633/ac7a4a.

Quantifying information of intracellular signaling: progress with machine learning

Tang Ying[1,2,3,*], **Hoffmann Alexander**[1,2,*]

[1]Institute for Quantitative and Computational Biosciences, University of California, Los Angeles, CA 90095, USA

[2]Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, CA 90095, USA

[3]International Academic Center of Complex Systems, Beijing Normal University, Zhuhai 519087, China

Abstract

Cells convey information about their extracellular environment to their core functional machineries. Studying the capacity of intracellular signaling pathways to transmit information addresses fundamental questions about living systems. Here, we review how information-theoretic approaches have been used to quantify information transmission by signaling pathways that are functionally pleiotropic and subject to molecular stochasticity. We describe how recent advances in machine learning have been leveraged to address the challenges of complex temporal trajectory datasets and how these have contributed to our understanding of how cells employ temporal coding to appropriately adapt to environmental perturbations.

Keywords

information processing; immune responses; mutual information; machine learning

Introduction

Cells are capable of sensing changes to their external environment and adapting their functions appropriately, a process known as cellular decision-making [1,2]. This involves the transmission of information gathered by molecular sensors or receptors through biochemical signaling pathways that can be functionally pleiotropic [3,4]. Signaling transduction is also subject to stochastic noise that affects molecular activities and mediates biological information transfer [5–9]. Cells may evolve to fine-tune noise levels to maximize information transmission [10], to distinguish stimulus conditions with specificity [11,12], or to allow for a degree of indeterminacy in decision-making within a population in a physiological strategy referred as bet-hedging [13]. An accurate quantification of the information flow within living systems is critical for characterizing such cellular behaviors and how their decision-making plays a role in physiology and pathology.

*Corresponding authors: Ying Tang, Alexander Hoffmann, jamestang23@gmail.com, ahoffmann@ucla.edu.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

At a conceptual level, “information” is a quantification on the amount of uncertainty. To formally quantify information, information theory was originally developed for the digital information transmitted through noisy channels [14]. Then, studies of electrical dynamics in neurons pioneered the application of information theory to biological systems [15–19]. Quantitative investigations addressed the information transmitted by neural spike trains elicited by the stimulus [20,21] and the information encoded in temporal patterns of firing [22]. Inspired by these studies, information-theoretic approaches have further been applied and adapted to study intracellular signaling processes [23,24] of immune cells responding to noxious substances [25,26].

For intracellular signaling, identifying signaling channels requires careful biological measurements to classify the groups of signaling molecules as pathways [4]. Thus, the term “signaling channel” needs to be accompanied by particular signaling molecules or defined by having separate time scales. By viewing biochemical signaling as an information channel, the sender is usually regarded as an environmental stimulus that is perceived by a receptor or sensor molecule. What is considered as the receiver, however, varies among studies and depends on what the experimental approach actually measures. Some studies measure the activities or subcellular localization of major signaling molecules, others measure the stimulus-induced expression of genes, and others measure cellular scale responses such as growth, division, movement, or death. Recent progress in measuring gene expression and signaling activities in individual cells [27] enables a quantitative investigation of intracellular information transmission based on experimental data. Information-theoretic analysis of such data may be used to quantify the extent of stimulus discrimination by the cell [28], as a focused biological problem of this review.

Among the many information measures, mutual information (Eq. 3) is a special one with important properties. It is a measure of correlation satisfying a set of requirements in Shannon’s theory [29]. When the measured variables are nonlinearly correlated, computing mutual information is still convenient. In addition, mutual information provides a likelihood for model inference [30]. This is especially useful when writing a likelihood function is hard, as prevailing in biological systems, where the intervening steps between measured species typically do not have quantitative models. For the stimulus discrimination process, mutual information has a clear biological meaning (Box 1); specifically, it indicates the amount of stimulus that cells can effectively discriminate by the intracellular signaling. Thus, mutual information is a major quantity to be reviewed.

Information-theoretic approaches are data-driven and involve a statistical estimation of the probability and entropy of the data. However, accurate estimation is hindered by some of the following properties of data. First, signaling-response data may be high dimensional, especially when based on imaging methods. Second, regulatory networks [31] of signaling molecules are capable of complex temporal patterns due to interdigitated feedback loops. Third, the data are affected by multiple sources that cause variability: preexisting biological heterogeneity within the population of cells, stochastic molecular noise that affects the signaling process, and technical imperfections in measurements. These make the estimation of the biologically relevant effective information capacity challenging.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Recent works have shown that these challenges can be approached by machine-learning approaches, where a class of models are trained by data to recognize patterns in the data, to infer probabilities and to inform the way unseen data are treated. In recent decades, machine learning has had great success in image classification, speech recognition, and more [32–35]. Machine-learning models use labeled data, known as training data, to learn the complex distribution of data. The model can also cluster training data into different categories, classify new data that are not seen during training to the corresponding category, and assign an accurate probability to new data when given a sufficient amount of training data. Among the many applications of machine learning, two specific tasks related to intracellular information processing include pattern classification and time series analysis.

First, for pattern classification where a specific number of potential answers exist and training data have been labeled, a machine-learning model can be trained to correctly classify unseen data. An example is the MNIST dataset of handwritten digits, where the trained model, such as a deep neural network [33], performs well on the task of recognizing new digits. Second, time series analysis aims to extract meaningful statistics from the time series data, including stock prices, climate change, and speech. Machine-learning models such as recurrent neural networks [36,37], a class of neural networks where connections between nodes form a temporal sequence, are able to exhibit temporal dynamical behavior, learn the patterns in the time series data, and further predict future values, known as time series forecasting.

The machine-learning approaches for the above two tasks are applicable for evaluating the information content of single-cell signaling response data. The pattern classification may classify distinct single-cell signaling responses when cells encounter different ligands or concentrations of the same ligand [11]. The trained model is then used to classify the measured signaling responses from unknown stimulation conditions. The truth table of the classification enables us to estimate the intracellular information transmission. In addition, since the signaling processes happen in a time course, time series analysis is helpful to extract biologically meaningful statistics from the measured data and evaluate the information transmission over time.

Extending these approaches of machine learning to intracellular signaling, the past five years have seen new advances both in theory and in application. Although previous reviews have described studies of intracellular information processing [24,30,38,39] and specific applications in immuno-oncology [40] or other biological problems [25,41,42], the new advances have yet to be summarized and put into context. Here, we summarize the foundational work on information-theoretic quantities and then describe recent advances in leveraging mathematical modeling and machine-learning approaches to quantify information transmission via biochemical signaling pathways (Figure 1). A summary of data sources (Table 2) and existing numerical packages (Box 2) to estimate the information-theoretic quantities is provided to help interested readers learn more and contribute to this field.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Foundational work on information theory for intracellular signaling

We begin with an overview of the fundamental information-theoretic quantities (Table 1). First, we list the basic definitions in information theory [45], which have also been summarized in past reviews [40,51]. We also review mathematical modeling to study intracellular information processing, where the model can also be used to generate data for the data-driven approaches of quantifying intracellular information transmission. Thus, the survey of this section on the basic qualities and mathematical modeling prepares us to review the data-driven approaches in the next sections.

Basic definitions of information quantities

Shannon entropy.—Historically, four major types of entropy have been formulated, each of which provides a way of understanding the probabilistic nature of random variables. First, originating from the understanding of gas laws in the mid-1800s, Clausius introduced the concept entropy as the ratio between heat and temperature [52]. Second, based on the frequentist view of probabilities, Boltzmann formulated the entropy with maximum multiplicity of the macroscopic states [53] to obey the second law of thermodynamics at equilibrium, justifying the Maxwell–Boltzmann distribution [54]. Gibbs further developed this formulation as an ensemble of options [55]. Third, following Shannon's information theory [14], Jaynes reframed statistical thermodynamics as inferences with the least possible bias under limited data [29]. Fourth, Shore and Johnson proved the principle of entropy maximization as requirements to be satisfied by any distribution function [44]. We refer readers to [56,57] for more detailed discussions.

We start with the formulation of entropy in Shannon's theory, as it is more appropriate to provide biologically sound interpretations for the major information measure of this review, the mutual information introduced below. The Shannon entropy [14] for a discrete random variable x with possible states X and a discrete probability distribution $P(x)$ is defined as:

$$H(X) = -\sum_{x\in X} P(x)\log_2 P(x)\,. \tag{1}$$

Shannon entropy has a close connection to statistical physics [29], providing a likelihood for inference on models given data. Examples of such inferences in quantitative biology include the protein 3D structure from genomic sequences [58,59], the prevalence landscape of mutated viral sequences [60], and the diversity of the antibody sequence repertoire [61].

Differential entropy.—For a continuous probability density $p(x)$, the Shannon entropy is defined as differential entropy [45]:

$$H_{diff}(X) = -\int_{-\infty}^{+\infty} dx\; p(x)\log_2 p(x)\,. \tag{2}$$

The estimate of entropy depends on estimating the probability distribution. One typically uses the frequency from finitely measured samples as an estimate of the probability to calculate the entropy of the probability distribution. Such an estimation of probabilities leads

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

to an error in calculating the entropy, which is proportional to the number of states and scales as 1 over the sample size [30].

Given N finite data, the cumulative probability distribution in the prefactor of Eq. (2) can be approximated by its sampling frequency [62]: $H_{diff}(X) = -\sum_{j=1}^{N} \delta_j \log_2 p(x_j)$, where δ_j is the frequency of observing the j-th event. When the number of sampled data is infinitesimal compared with the number of total configurations, the sampling frequency δ_j can be chosen as uniform for each sampled event, $\delta_j = 1/N$, giving an averaged entropy (Boltzmann entropy $-\log_2 p(x_j)$ for the configuration x_j) of the finite samples. When the sampled data are sufficient to cover the frequent configurations of the full probability distribution, the averaged entropy from the samples approximates the entropy of the full distribution. This approximation was found useful to produce an accurate estimation of the mutual information of intracellular signaling [62].

Kullback–Leibler (KL)-divergence.—For two discrete probability distributions $P(x)$, $Q(x)$, the KL-divergence [43] $D_{KL}(P \parallel Q) = \sum_{x \in X} P(x) \log_2[P(x)/Q(x)]$ quantifies the dissimilarity between the two distributions. The KL divergence is also named the relative entropy. As a symmetrized KL divergence, the Jensen–Shannon divergence plays a similar role in measuring the similarity between two probability distributions. The divergence can be used to quantify the similarity between the data distribution and the distribution generated from the model in various scientific disciplines. In biology, it has been applied to quantify distributional dissimilarity, including that between genes in tumors and healthy samples [63] and that between transcriptional states of T lymphocytes [64].

Cross entropy.—The cross entropy [44] for discrete probability distributions is $CE(P \parallel Q) = D_{KL}(P \parallel Q) + H(P)$, where $H(P)$ denotes the entropy of the probability distribution P as in Eq. (1). It quantifies the information across the two probability distributions, which is extendable to the continuous case similarly to the differential entropy. Cross-entropy between distributions of data and models may serve as a loss function in machine learning. Its application in biology is similar to the KL divergence.

Mutual information.—Given another random variable y with possible states Y, the mutual information [45,65,66] between the two random variables is:

$$I(Y; X) = H(Y) - H(Y \mid X), \tag{3}$$

where the conditional entropy $H(Y \mid X) = -\sum_{x \in X, y \in Y} P(x, y) \log_2[P(x, y)/P(x)]$. Mutual information quantifies the mutual dependence between the random variables, i.e., the amount of information about one random variable through observing the other. It being zero is equivalent to the two random variables being independent.

Mutual information is symmetric and represents the correlation of two variables, which is termed “cooperativity” in physical biochemistry. In practice, one usually cares about the maximum mutual information (channel capacity), and maximization is conducted for only one variable, such that the interpretation becomes asymmetrical. The maximization is done

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

by inferring the probability distributions from limited data under certain constraints, such as the probability normalization condition. Thus, mutual information is closely related to the type of entropy formulated by Jaynes [14,29].

In addition, mutual information can also be regarded as a Kullback–Leibler (KL) divergence between the conditional distributions and the prior distribution P_X: $I(Y;X) = E_Y\left[D_{KL}\left(P_{X\mid Y} \mid P_X\right)\right]$ where E_Y is the expected value over the random variable Y. That is, mutual information is the expectation of the KL divergence of the univariate distribution P_X from the conditional distribution $P_{X\mid Y}$ given Y. The more different the two distributions are on average, the greater the information gain.

The mutual information is widely useful. It helps disentangle interactions between a system's internal variables and their coupling to changing environments [67]. It has also been extended to various contexts, for example, the renormalized mutual information for continuous variables with a deterministic dependence [68]. More importantly, it is the mutual information rather than the entropy that is more often used as a likelihood for the model inference [30,65].

Channel capacity.—Channel capacity is obtained by maximizing mutual information between the input and output distributions $P_X(x)$ and $P_Y(y)$, which measures the rate at which information is transmitted over a communication channel. The maximum mutual information is formally obtained as:

$$I_{max}(Y;X) = \max_{P_X(x)} I(Y;X), \tag{4}$$

where the maximization is with respect to the input marginal distribution $P_X(x)$.

Mutual information and channel capacity are essential to quantify the stimulus discrimination process by intracellular signaling (Box 1). Therefore, estimating the probabilities from limited measurements requires dedicated approaches, which will be elaborated in the next sections.

Pointwise information measures

We now review the pointwise information measures and measures that consider two consecutive timepoints. We denote two time series (trajectories) by $x_{1:n}, y_{1:n}$, where the subscript represents the timepoints, 1 to n. The dynamics of the time series can be incorporated by the transition probabilities, i.e., the conditional probabilities of the time series. For clarity, we consider a system with the Markov property: the conditional probability $p(x_n \mid x_{1:n-1}) = p(x_n \mid x_{n-1})$. The information measures can be extended to the case of a stationary Markov process with higher order, i.e., longer memory.

Transfer entropy.—Given the transition probabilities, the transfer entropy [46] measures the amount of directed transfer of information between two time series, which can distinguish the driving and responding elements. It is defined as follows:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$T_{X\to Y} = H(y_n \mid y_{n-1}) - H(y_n \mid y_{n-1}, x_{n-1}), \tag{7}$$

$T_{X\to Y} = H(y_n \mid y_{n-1}) - H(y_n \mid y_{n-1}, x_{n-1})$, where the conditional entropies are for the time series. Transfer entropy is a conditional mutual information $T_{X\to Y} = I(y_n; x_{n-1} \mid y_{n-1})$, which has the history of the variable y_{n-1} in the condition. Transfer entropy is a finite version of directed information [74]. Restricted directed information was used to infer the causal relation between genes from single-cell RNA sequence data [75]. Similarly, for a single time series, the excess entropy [76] measures the amount of uncertainty in the future explained by the past information.

General information metric for two timepoints.—The strength of causal influences for two timepoints i, j between two time series $x_{1:n}, y_{1:n}$ can be demonstrated by a unified framework of information measures [47]. To derive the framework, the authors approximated the joint probability distribution $p(x_{i,j}, y_{i,j})$ by another probability distribution $q(x_{i,j}, y_{i,j})$. The causal influences between two time series $ci(x_i \to y_j)$ can be quantified by minimizing the KL divergence between the two probability distributions $p(X_{1:n}, Y_{1:n}), q(X_{1:n}, Y_{1:n})$:

$$ci(x_i \to y_j) = \min_{q(x_{i,j}, y_{i,j})} D_{KL}\left[p(x_{i,j}, y_{i,j}) \parallel q(x_{i,j}, y_{i,j})\right], \tag{8}$$

under the constraint of the Markov condition: $q(x_i, y_j \mid y_i) = q(x_i \mid y_i)q(y_j \mid y_i)$ [47]. Note that here the two timepoints are denoted by the subscript, whereas it is denoted by x, y in [47]. This general information metric can be reduced to various information measures [77], including mutual information and transfer entropy. It also generates integrated information [78] that quantifies the extent of synergistic causal influences between the two series and the stochastic interaction [79] that measures the mixed strength of causal and simultaneous influences. We expect that it will have applications in understanding the full information transfer between two dynamical variables of biological systems.

Trajectory-wise information measures

The above measures do not estimate the mutual information from an entire trajectory. To this end, the trajectory-wise information measures will be covered as follows.

Trajectory entropy.—Given the trajectory's probability $p(y_{1:n})$ for the observed trajectory $y_{1:n}$, the trajectory entropy for each single trajectory is given by [48]:

$$H(y_{1:n}) = -\log_2 p(y_{1:n}). \tag{9}$$

The trajectory entropy is for a single trajectory, rather than the average on the trajectory ensemble [80]. The trajectory entropy was originally formulated for mesoscopic nonequilibrium systems [81]. Based on the trajectory entropy, a set of thermodynamical quantities can be formulated [56,82]. In addition, the principle of maximum caliber [56]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

extends the principle of entropy maximization to trajectories, and thus the maximization with respect to trajectories can be conducted in a similar procedure.

The trajectory entropy itself may be a stochastic quantity, and different experimental realizations lead to different distributions of the trajectory entropy. However, when the experimental condition is fixed and only repeated measurements are conducted, the entropy's distribution is fixed and should be fully determined by a fixed distribution of the trajectory probability. Each trajectory configuration has a probability and an entropy value. In this case, the trajectory entropy can be inferred in the same way as the entropy for static variables, and the concept of information is defined similarly.

The probability of trajectory is not well defined mathematically in continuous-time space because the trajectory configurations are infinite and the total probability volume is infinite. Thus, discrete time is required to rigorously define the probability space. In practice, one can use the frequency of the trajectory with discrete time and finite state as an estimate of the probability and employ the differential entropy Eq. (2) to approximate the averaged trajectory entropy for an ensemble of trajectories. For example, the probability can be calculated by inferring a stochastic model from the data of signaling responses and is useful to quantify the mutual information from the time series data of intracellular signaling responses [49].

Mutual information in trajectory space.—In trajectory space, mutual information can be formulated as in [49,83]. We consider the mutual information between the input set (X) and the output trajectory set $(Y_{1:n})$, where $n = 1, 2, 3, \ldots$ denotes the timepoint. Up to each timepoint n,

$$I(Y_{1:n}; X) = H(Y_{1:n}) - H(Y_{1:n} \mid X), \tag{10}$$

where $H(Y_{1:n} \mid X)$ and $H(Y_{1:n})$ are the conditional and unconditional trajectory entropy based on Eq. (9). When the trajectory probability is generated from a dynamical model, the probability depends on the dynamics. Then, the trajectory entropy and the mutual information also depend on the dynamics, such that the information embedded in dynamical patterns of trajectories can be revealed by this mutual information. The maximization for $I(Y_{1:n}; X)$ is done at each time point, which is a quantification of the maximum extent of distinguishing the stimuli cumulatively (see subsection "The stochastic model-based method").

Mutual information rate in Fourier-frequency space.—The mutual information rate at which the information between trajectories increases with time has been formulated in the Fourier-frequency space [50,84]. The authors considered two ensembles of time series at steady state with each obeying Gaussian statistics. The coupling between ensembles can be linearized. Under the assumptions, the joint probability distribution of the two series fluctuates around the steady state mean values, and $x_{1:n}, y_{1:n}$ is given by $\rho(z) = \exp\left(-v^T Z^{-1} v/2\right)/\left[(2\pi)^N |Z|^{1/2}\right]$, where the vector $z \doteq (x_{1:n}, y_{1:n})$. The covariance

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

matrix Z has the matrix blocks $C^{xx}, C^{xy}, C^{yx}, C^{yy}$, where each is defined as $C_{ij}^{xx} = \langle x_i x_j \rangle$ with $\langle\rangle$ denoting the noise average.

In the continuous-time limit at a fixed time interval, the mutual information rate between the two trajectory ensembles $I_R(x_{1:n}; y_{1:n}) \doteq \lim_{n \to \infty} I(x_{1:n}; y_{1:n})/n$ is calculated as:

$$I_R(x_{1:n}; y_{1:n}) = -\frac{1}{4\pi}\int_{-\infty}^{\infty} d\omega \ln\left[1 - \frac{|S^{xy}(\omega)|^2}{S^{xx}(\omega)S^{yy}(\omega)}\right], \tag{11}$$

where $S^{xx}(\omega)$, $S^{xy}(\omega)$, $S^{yy}(\omega)$ is the power spectrum from the Fourier transform of C^{xx}, C^{xy}, C^{yy}. The mutual information rate reveals the information transmission from the ligand concentration to the flagellar motor in the chemotaxis network of *E. coli* [50].

Information theory to intracellular signaling with mathematical modeling

With the above information quantities, information transmission through signaling networks has been characterized with the help of mathematical models. A number of mathematical models have been constructed to model signaling networks [31] and analyze the information flow in the networks [85]. More specifically, the information flow was estimated in models of gene regulation [26,86–88]. Optimal information processing strategies have been studied in different network topologies of gene regulation [89–95] using the data on noise levels of gene expression [10]. Information transmission in the MAPK/ERK pathway [96] and in the bacterial quorum sensing signaling network [97] was analyzed. The channel capacity was calculated from a discrete-time Markov model on the signaling transduction [98], and the mutual information was evaluated through chemical reaction networks [99–101].

In addition to quantifying information transmitted through one signaling molecule, the information flow through shared network components for multiple inputs and outputs was studied in interferon signaling [102] and with a Boolean network of fibroblast signal transduction [103]. The contribution of duplicated components in the signaling pathway to channel capacity was investigated [104]. Information transmission was found to be maximized by synergistic control in noisy gene regulatory networks [105]. The information transfer between dynamical system components was formulated for both continuous and discrete systems [106], as well as stochastic dynamical systems [107,108], where noise was tuned to improve information transmission [109]. Furthermore, information theory was used in deterministic dynamical systems to infer the structure of signaling networks [110]. The decoding of signaling information to determine downstream gene expression was explored [111,112].

First data-driven approaches of information theory to intracellular signaling

Henceforth, we focus on the mutual information between the extracellular stimulus conditions and the intracellular single-cell signaling responses, which provides an estimate of the amount of information about the stimulus identity and dose (Box 1). We mainly review the methods using single-cell measurements of signaling molecules by live-cell

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

imaging, as it provides real-time tracking of the signaling activities that are crucial to quantify information transmission.

In this section, we review the first set of data-driven approaches in historical developments. The prominent statistical approaches in quantifying information transmission from the live-cell imaging data are listed in Figure 2. A pioneering work employed a single-timepoint measurement [113,114] (time-point method). A second approach [62] evaluated the information encoded in the signaling time course from the multivariate measurement (vector method), including a further extension by considering dynamical features of the signaling responses [115]. Extracting information transmission from long time series of signaling responses requires alternative approaches, which will be reviewed in the next section.

The time-point method

As a pioneering work in quantifying information transmission from measured single-cell signaling activity, the authors in [113] estimated the mutual information and channel capacity at a single timepoint. At each timepoint, the data from single-cell measurement under one stimulus condition led to a distribution of signaling activity across cells, and the distributions under various stimulus conditions provided mutual information for stimulus discrimination.

The estimated mutual information is affected by noise [28,116] and the feedback of regulators [114,117–119]. The analysis has been extended to multiple signaling molecules, enabling the noise decomposition of biochemical signaling networks [120]. For measurements at multiple timepoints, the method is applicable to each timepoint separately, without taking into account the time course of signaling responses. Thus, the information transmission over a time course through the signaling molecule may be lost.

The vector method applied to measurements

Remarkable progress was made in [62] to quantify the information transmission over the time course of signaling responses. The method treated the time series data from each single cell as a multivariate vector and used the k-nearest neighbor to estimate the probability of the time series [121,122]. The performance of the k-nearest neighbor estimator depends on the metric of the distance and the value of k [123], which may need to be fine-tuned for each dataset. The error bars and bias of this estimated mutual information were evaluated [124,125], and the accuracy was improved by kernel estimation [126]. Furthermore, information was coded by a combination of time series and molecular species [127].

Although the method can evaluate the information of the time course, the current limitation on the number of cells from live-cell imaging data restricts the length of the time course for an accurate estimation. As sampling the vectorial distribution suffers from a combinatorial explosion, the estimation becomes inaccurate when the number of timepoints increases over ~10 timepoints [49]. In addition, treating the time series as vectors makes the density estimation independent of the ordering of timepoints and thus does not distinguish dynamical patterns encoded in the time series.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The vector method applied to dynamical features

The dynamical features of the temporal signaling responses transmit information [128], such as through amplitude and frequency regulation of transcription factor activity [129]. Information transmission via the dynamical features of signaling responses has been quantified [115] (Figure 3). The effect of representative features has been analyzed by adding one or a few features [130]. By adding dynamical features simultaneously, more information encoded in the trajectories was extracted [115]. This analysis uncovers the most informative features that optimize stimulus discrimination. Note that calculating each dynamical feature is subject to noise, which may alter the estimation of mutual information.

Recent approaches for estimating information transmission

To extract the information transmission from the long time course of the signaling responses [131–134], recent works have employed machine-learning methods. Below, we review three representative examples. They include the decoding-based approach [130] that uses the machine-learning classifier, the statistical learning-based method using logistic regression, and the stochastic model-based method which employs the hidden Markov model. In each case, the machine-learning methods help estimate the probability and information from the time series data. A comparison of the data-driven approaches is provided in the end.

The decoding-based approach

One approach used a machine learning decoder to calculate the mutual information [130] (Figure 4). This method first trained a classifier given the time series of signaling responses under each stimulus condition and used the classifier to separate new data of signaling responses into the group with the best match. It provided a lower bound on the mutual information, and the deviation depended on the accuracy of the classifier. When classifiers employ linear principal components, they may be inadequate for discriminating oscillatory and nonoscillatory trajectories. To overcome this issue, various machine-learning models, such as neural networks, can be used for classifiers to improve estimates [99]. In addition to the lower bound, an upper bound on the mutual information was derived [135].

The classifier can also use the most informative features to discriminate stimulus conditions, where the top-ranked features termed signaling codewords are identified by information-theoretic analysis [115]. The codewords were further used to construct a decision tree to classify the stimulus conditions binarily by specific dynamical features. In addition, the dynamical signaling patterns to realize the optimal transmission of information were obtained by optimal control theory [136].

The statistical learning-based method

As an efficient method, a statistical learning-based estimation of mutual information (SLEMI) was proposed in [125,137] (Figure 5). The method is applicable to high-dimensional time series of signaling responses, without restriction on the number of timepoints. The numerical package [125] enables a broader use to various datasets, generating mutual information, channel capacity and probabilities of correct pairwise discrimination. SLEMI used a Bayesian framework based on logistic regression to estimate

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the probabilities of stimuli given measured trajectories. As logistic regression assumes a linear fitting on the trajectories to calculate the ratio of the trajectory probabilities between stimulus conditions, it is not clear whether this approach can fully account for the complex dynamical patterns of observed signaling trajectories, such as oscillatory behavior [115,138]. Thus, the estimated mutual information could be less accurate when applied to complex trajectories, where logistic regression may be replaced by a more advanced Bayesian classifier [125].

The stochastic model-based method

Inspired by the trajectory entropy defined along a single trajectory [81], the data can be viewed from the trajectory perspective. Then, stochastic dynamical models, such as the hidden Markov model that was used for speech recognition [139], can be applied to learn and reproduce the time course of the signaling responses [49] (Figure 6). The hidden Markov model (or the time-inhomogeneous Markov model) captures the time-inhomogeneity of the trajectories and represents the trajectory ensemble with approximately 80% accuracy. The model further generates trajectory probabilities to calculate mutual information. The limited number of measured cells and timepoints in live-cell imaging may alter the accuracy of the model training and the subsequent mutual information estimation.

This framework provides an estimate of the information encoded in the signaling dynamics over time. The estimated information accumulation over time reveals the temporal ordering of the discriminating different stimuli and may decrease when the stimuli induce similar signaling responses in a certain time regime that diminish the extent of the stimulus discrimination. It also indicates the temporal phases of information transmission that can be mapped to the functionality of the regulatory circuit and the amount of information accumulation available to immune response genes [49].

A comparison of the data-driven approaches

Applying each of the approaches above to the NFκB signaling responses under 13 different immune stimulus conditions characterizes their properties (Figure 7). All the methods give the maximum mutual information of approximately 1~2 bits, smaller than $\log_2 13 \approx 3.7$ bits under perfect transmission. The loss of information may be caused by molecular noise in signaling responses. For each method, the mutual information calculated from a single timepoint [113] ignores the information from time courses. The vector method [62] is ineffective when there are more than approximately 10 timepoints because it becomes inaccurate to sample the vectorial distribution from the measured data. Both methods do not distinguish the dynamical patterns with timepoints aligned properly.

The decoding-based method may not fully count the information over long time courses, as mutual information estimates are saturated after a handful of measurements. This can be improved with the performance of the classifier [99] and by using the optimal input distribution instead of the uniform distribution [130]. The random permutation of timepoints does not significantly alter the estimation, indicating an incomplete discrimination of the dynamical patterns of signaling responses and a lack of tracking information over time.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Thus, the scope of applying the decoding-based method depends on the complexity of the time series and the quality of the accessible classifiers.

Both the SLEMI [125] and stochastic model-based method give increasing mutual information, implying distinct temporal patterns of signaling responses at all times, as consistently observed in the data [49]. After the random permutation of timepoints, the mutual information from the two methods decreases, corresponding to the information under the genuine order of timepoints where the distinct stimuli become less distinguishable. The stochastic model-based method provides continuously increasing mutual information over time, even after the random permutation of timepoints, as permuted signaling responses have persistent differences in response amplitudes [49]. However, mutual information, such as in the early time regime, may be underestimated if the searched optimal number of parameters has an underestimation or the stochastic model does not accurately learn the dynamics.

Outlook

We have reviewed studies on quantifying the information transmission of intracellular signaling with the aid of mathematical modeling and machine learning. Several outstanding questions are guiding current and future studies.

To improve the quantification of information transmission, advanced models of machine learning [32,34] may be employed to learn the time course of signaling responses with higher performance. They may also enable the extraction of diverse useful information from the time series. Specifically, the recurrent neural network [36,37,140] has achieved great success in learning the dynamics of time series. The transformer with an attention-based architecture [141] performs well in learning complex time series because it can capture long-range dependencies between input and output by designing neural networks with functional gates for memorizing and updating. The application of these models to single-cell signaling responses may have a better performance in reproducing data and predicting future responses.

In addition to information transmission by signaling molecules in single cells, the reliability of signal transduction is affected by cell populations [142,143]. Cell subpopulations can independently transmit information that gives graded responses to stimuli [144]. Fractional response analysis by using Rényi information further reveals that changes in fractions of cells under various response levels scale linearly with the log of the cytokine dose [145]. It is also attractive to quantify the information content of the signaling process in more realistic contexts [146], such as under mixed natural signals. The mutual information estimates under time-varying signals reveal the information flow when cells are subject to environmental changes [99,147]. The information flow can be optimized by controlling the environment via reinforcement learning [148], which models how agents take action in an environment to maximize a cumulative reward, such as the information gain. In addition, the positional information from the spatially distributed signaling molecules has been evaluated by mathematical modeling [149,150] and by constructing the decoder [151]. Quantifying the information transmission over large spatial and long time scales [152] awaits further developments, such as by convolution neural networks [153].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The increase of the single-cell data would continue to motivate future work of evaluating information content in a data-driven manner, and vice versa. In addition to live-cell imaging, applying machine-learning models to other single-cell data would reveal more insights into intracellular information processing. For example, the causal relation between genes has been inferred from single-cell RNA sequence data by using restricted directed information [75]. Both scRNA and smFISH (Table 2) data can measure downstream gene expression of the signaling molecule, providing a platform to test the hypothesis on conveying information of the signaling molecule to gene expression [49,112]. The autoencoder [35] may help learn meaningful representations from these multigene data. Furthermore, predictions from information-theoretic approaches [112] can be tested experimentally by optogenetic approaches [131,154,155]. Such experimental setups avoid the coactivation of other, unknown factors involved in gene expression, providing an unambiguous way to measure information transmission by the signaling molecule to downstream genes. Exploring the decoding of the signaling information to responsive gene expression for cell fate decisions documents the actual physiological role of estimated information quantities and reveals evolutionary perspectives of cellular information processing and decision-making.

While machine-learning approaches show promising applications in understanding living information processing, we would like to remind ourselves that simply applying machine learning as a tool may have limitations. As quoted from E. T. Jaynes [156]: "*New data that we insist on analyzing in terms of old ideas (that is, old models which are not questioned) cannot lead us out of the old ideas. However, many data we record and analyze, we may just keep repeating the same old errors, and missing the same crucially important things that the experiment was competent to find. That is what ignoring prior information can do to us; no amount of analyzing coin tossing data by a stochastic model could have led us to discovery of Newtonian mechanics, which alone determines those data*."[1] Therefore, machine learning and information theory should be taken as frameworks to help design the experiment, such as mutual information, which can frame an inference problem for modeling biological systems [30] and provide a new angle to understand and predict biological processes beyond existing data [157]. We anticipate that the cross-feeding between quantitative biology, information theory, and machine learning [158] will lead to significant advances in these areas.

Sources of single-cell data

We list sources of single-cell data useful for information-theoretic analysis, including mathematical model simulations, single-cell RNA (scRNA) sequences [159], single-molecule fluorescence in situ hybridization (smFISH), and live-cell imaging [160] (Table 2). In the main text, we have mainly reviewed the approaches using live-cell imaging, but other types of data may find increasingly important roles in future studies. See a complementary review on the data source for studying intracellular signaling [27].

[1]We thank an anonymous reviewer for mentioning this quote.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Data simulated from mathematical models.

To evaluate information transmission, the simulated data of signaling molecules can be generated from differential equations for modeling signaling transduction [62,109,112]. The trajectories of chemical species were also simulated from chemical reaction networks [99,100,111] by the stochastic simulation algorithm (or the Gillespie algorithm) [161]. To generate trajectories that accurately simulate the real time-course of signaling activities, the mathematical model needs to be experimentally calibrated and verified, which may require massive measurements on the modeled molecules and exploration of the model parameters [162,163].

Single-cell RNA sequencing (scRNA-seq).

The development of scRNA-seq has increased in recent years, as it generates the sequence profiles of all transcripts with their relative abundances in single cells. However, scRNA-seq data from methodologies such as droplet sequencing are subject to nonnegligible noise, and accurately measured genes are sparse. Thus, the data typically need dimension reduction to generate useful statistics and do not meet the high resolution required for the information-theoretic approaches of quantifying intracellular information transmission. The specifically designed measurement, e.g., targeted scRNA-seq, may be more suitable, with a tradeoff between the number of measured genes and the control on the noise level.

In addition, scRNA-seq technologies initially measure gene expression at individual timepoints and do not track the transcriptome over time. To overcome this limitation, the pseudotime can be inferred to map out the trajectories (e.g., developmental trajectories of gene expression) for single cells [164,165], with multiple-timepoint measurements [166,167]. However, the accuracy of trajectory inferences depends on the dynamics of gene expression [168] and needs to be verified, such as by real-time tracking. The underlying dynamical equations governing the cell state transition can be inferred [169], which may provide high-resolution augmentation of the noisy distribution of signaling molecules over time to estimate intracellular information transmission.

Single molecule fluorescence in situ *hybridization (smFISH)*.

As an imaging-based technique, smFISH enables the measurement of the expression of endogenous genes from ~10000 cells. A recent technique (MERFISH) can simultaneously image 100 to 1000 RNA species in single cells [170]. Nevertheless, smFISH needs to fix the sample and only measure it at a single timepoint, which prohibits its usage in quantifying information transmission over time.

Live-cell imaging.

Live-cell imaging is a direct method to measure the signaling activity of living cells in real time [62,62,115,130,133,138,171,172]. The time resolution reaches the time scale of minutes, which can continue for days. Approximately one thousand cells were measured in each experiment. This technique has a limitation on the number of signaling molecules measured simultaneously, which typically allows one or two molecules to be probed to date. Live-cell imaging and smFISH are complementary based on their pros and cons.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Software packages

To calculate the information-theoretic quantities, a number of software packages are available (Box 2). Some of these are also listed in [40].

Acknowledgments

We thank Roy Wollman, Eric J. Deeds, Adewunmi Adelaja, Katherine Sheu, and Haripriya Vaidehi Narayanan for discussions. We acknowledge Michał Komorowski and Peter S. Swain for their careful reading and helpful comments. The work was funded by NIH Grant R01AI127864 (to A.H.). Y.T. is supported by the Collaboratory Fellowship at UCLA and the National Natural Science Foundation of China (12105014).

References

[1]. Balázsi G, van Oudenaarden A and Collins JJ 2011 Cellular Decision Making and Biological Noise: From Microbes to Mammals Cell 144 910–25 [PubMed: 21414483]

[2]. Bowsher CG and Swain PS 2014 Environmental sensing, information transfer, and cellular decision-making Current Opinion in Biotechnology 28 149–55 [PubMed: 24846821]

[3]. Behar M, Barken D, Werner SL and Hoffmann A 2013 The Dynamics of Signaling as a Pharmacological Target Cell 155 448–61 [PubMed: 24120141]

[4]. Natarajan M, Lin K-M, Hsueh RC, Sternweis PC and Ranganathan R 2006 A global analysis of cross-talk in a mammalian cellular signalling network Nat Cell Biol 8 571–80 [PubMed: 16699502]

[5]. Elowitz MB, Levine AJ, Siggia ED and Swain PS 2002 Stochastic Gene Expression in a Single Cell Science 297 1183–6 [PubMed: 12183631]

[6]. Raj A and van Oudenaarden A 2008 Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences Cell 135 216–26 [PubMed: 18957198]

[7]. Raser JM and O'Shea EK 2004 Control of Stochasticity in Eukaryotic Gene Expression Science 304 1811–4 [PubMed: 15166317]

[8]. Swain PS, Elowitz MB and Siggia ED 2002 Intrinsic and extrinsic contributions to stochasticity in gene expression PNAS 99 12795–800 [PubMed: 12237400]

[9]. Bressloff PC 2014 Stochastic Processes in Cell Biology (Springer)

[10]. Tkačik G, Callan CG and Bialek W 2008 Information flow and optimization in transcriptional regulation PNAS 105 12265–70 [PubMed: 18719112]

[11]. Sheu KM, Luecke S and Hoffmann A 2019 Stimulus-specificity in the responses of immune sentinel cells Current Opinion in Systems Biology 18 53–61 [PubMed: 32864512]

[12]. François P, Voisinne G, Siggia ED, Altan-Bonnet G and Vergassola M 2013 Phenotypic model for early T-cell activation displaying sensitivity, specificity, and antagonism Proceedings of the National Academy of Sciences 110 E888–97

[13]. Veening J-W, Smits WK and Kuipers OP 2008 Bistability, Epigenetics, and Bet-Hedging in Bacteria Annual Review of Microbiology 62 193–210

[14]. Shannon CE 1948 A mathematical theory of communication The Bell System Technical Journal 27 379–423

[15]. Rieke F, Warland D, de Ruyter van Steveninck R and Bialek W 1997 Spikes: Exploring the Neural Code vol 7 (MIT Press, Cambridge)

[16]. Borst A and Theunissen FE 1999 Information theory and neural coding Nature Neuroscience 2 947–57 [PubMed: 10526332]

[17]. Paninski L 2003 Estimation of Entropy and Mutual Information Neural Computation 15 1191–253

[18]. Quian Quiroga R and Panzeri S 2009 Extracting information from neuronal populations: information theory and decoding approaches Nature Reviews Neuroscience 10 173–85 [PubMed: 19229240]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[19]. Nemenman I, Shafee F and Bialek W 2001 Entropy and inference, revisited Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic NIPS'01 (Cambridge, MA, USA: MIT Press) pp 471–8

[20]. Strong SP, Koberle R, de Ruyter van Steveninck RR and Bialek W 1998 Entropy and Information in Neural Spike Trains Phys. Rev. Lett 80 197–200

[21]. Victor JD 2002 Binless strategies for estimation of information from neural data Phys. Rev. E 66 051903

[22]. Reinagel P and Reid RC 2000 Temporal Coding of Visual Information in the Thalamus J. Neurosci 20 5392–400 [PubMed: 10884324]

[23]. Lan G and Tu Y 2016 Information processing in bacteria: memory, computation, and statistical physics: a key issues review Rep. Prog. Phys 79 052601

[24]. Levchenko A and Nemenman I 2014 Cellular noise and information transmission Current Opinion in Biotechnology 28 156–64 [PubMed: 24922112]

[25]. Waltermann C and Klipp E 2011 Information theory based approaches to cellular signaling Biochimica et Biophysica Acta (BBA) - General Subjects 1810 924–32 [PubMed: 21798319]

[26]. Tkačik G and Walczak AM 2011 Information transmission in genetic regulatory networks: a review J. Phys.: Condens. Matter 23 153102

[27]. Patange S, Girvan M and Larson DR 2018 Single-cell systems biology: Probing the basic unit of information flow Current Opinion in Systems Biology 8 7–15 [PubMed: 29552672]

[28]. Rhee A, Cheong R and Levchenko A 2012 The application of information theory to biochemical signaling systems Phys. Biol 9 045011

[29]. Jaynes ET 1957 Information Theory and Statistical Mechanics Phys. Rev 106 620–30

[30]. Tkačik G and Bialek W 2016 Information Processing in Living Systems Annual Review of Condensed Matter Physics 7 89–117

[31]. Alon U 2019 An Introduction to Systems Biology: Design Principles of Biological Circuits (CRC Press)

[32]. Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L and Zdeborová L 2019 Machine learning and the physical sciences Rev. Mod. Phys 91 045002

[33]. LeCun Y, Bengio Y and Hinton G 2015 Deep learning Nature 521 436–44 [PubMed: 26017442]

[34]. Mehta P, Bukov M, Wang C-H, Day AGR, Richardson C, Fisher CK and Schwab DJ 2019 A high-bias, low-variance introduction to Machine Learning for physicists Phys. Rep 810 1–124 [PubMed: 31404441]

[35]. Goodfellow I, Bengio Y and Courville A 2016 Deep Learning (MIT Press)

[36]. Hochreiter S and Schmidhuber J 1997 Long Short-Term Memory Neural Computation 9 1735–80 [PubMed: 9377276]

[37]. Vlachas PR, Pathak J, Hunt BR, Sapsis TP, Girvan M, Ott E and Koumoutsakos P 2020 Backpropagation algorithms and Reservoir Computing in Recurrent Neural Networks for the forecasting of complex spatiotemporal dynamics Neural Networks 126 191–217 [PubMed: 32248008]

[38]. Zielińska KA and Katanaev VL 2019 Information Theory: New Look at Oncogenic Signaling Pathways Trends in Cell Biology 29 862–75 [PubMed: 31630880]

[39]. Topolewski P and Komorowski M 2021 Information-theoretic analyses of cellular strategies for achieving high signaling capacity—dynamics, cross-wiring, and heterogeneity of cellular states Current Opinion in Systems Biology 27 100352

[40]. Karolak A, Branciamore S, McCune JS, Lee PP, Rodin AS and Rockne RC 2021 Concepts and Applications of Information Theory to Immuno-Oncology Trends in Cancer 7 335–46 [PubMed: 33618998]

[41]. Mian IS and Rose C 2011 Communication theory and multicellular biology Integrative Biology 3 350–67 [PubMed: 21424025]

[42]. Uda S 2020 Application of information theory in systems biology Biophys Rev 12 377–84 [PubMed: 32144740]

[43]. Kullback S and Leibler RA 1951 On Information and Sufficiency The Annals of Mathematical Statistics 22 79–86

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[44]. Shore J and Johnson R 1980 Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy IEEE Transactions on Information Theory 26 26–37

[45]. Cover TM and Thomas JA 2012 Elements of Information Theory (John Wiley & Sons)

[46]. Schreiber T 2000 Measuring Information Transfer Phys. Rev. Lett 85 461–4 [PubMed: 10991308]

[47]. Oizumi M, Tsuchiya N and Amari S 2016 Unified framework for information integration based on information geometry PNAS 113 14817–22

[48]. Seifert U 2005 Entropy Production along a Stochastic Trajectory and an Integral Fluctuation Theorem Phys. Rev. Lett 95 040602

[49]. Tang Y, Adelaja A, Ye FX-F, Deeds E, Wollman R and Hoffmann A 2021 Quantifying information accumulation encoded in the dynamics of biochemical signaling Nat. Commun 12 1272 [PubMed: 33627672]

[50]. Tostevin F and ten Wolde P R 2009 Mutual Information between Input and Output Trajectories of Biochemical Networks Phys. Rev. Lett 102 218101

[51]. Xu S, Böttcher L and Chou T 2020 Diversity in biology: definitions, quantification and models Phys. Biol 17 031001

[52]. Clausius R 1850 Ueber die bewegende Kraft der Wärme und die Gesetze, welche sich daraus für die Wärmelehre selbst ableiten lassen Annalen der Physik 155 368–97

[53]. Boltzmann L 1896 Vorlesungen über Gastheorie: Th. Theorie des Gase mit einatomigen Molekülen, deren dimensionen gegen die Mittlere weglänge verschwinden (J. A. Barth)

[54]. Brush SG 1975 The Kind of Motion We Call Heat: A History of the Kinetic Theory of Gases in the 19th Century (North-Holland Publishing Company)

[55]. Gibbs JW 1902 Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundations of Thermodynamics (C. Scribner's sons)

[56]. Pressé S, Ghosh K, Lee J and Dill KA 2013 Principles of maximum entropy and maximum caliber in statistical physics Rev. Mod. Phys 85 1115–41

[57]. Qian H 2021 Thermodynamic Behavior of Statistical Event Counting in Time: Independent and Correlated Measurements arXiv:2109.12806 [cond-mat]

[58]. Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R and Sander C 2011 Protein 3D Structure Computed from Evolutionary Sequence Variation PLOS ONE 6 e28766

[59]. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, Zecchina R, Onuchic JN, Hwa T and Weigt M 2011 Direct-coupling analysis of residue coevolution captures native contacts across many protein families Proceedings of the National Academy of Sciences 108 E1293–301

[60]. Chakraborty AK and Barton JP 2017 Rational design of vaccine targets and strategies for HIV: a crossroad of statistical physics, biology, and medicine Rep. Prog. Phys 80 032601

[61]. Mora T, Walczak AM, Bialek W and Callan CG 2010 Maximum entropy models for antibody diversity Proceedings of the National Academy of Sciences 107 5405–10

[62]. Selimkhanov J, Taylor B, Yao J, Pilko A, Albeck J, Hoffmann A, Tsimring L and Wollman R 2014 Accurate information transmission through dynamic biochemical signaling networks Science 346 1370–3 [PubMed: 25504722]

[63]. Ramakrishnan N and Bose R 2017 Analysis of healthy and tumour DNA methylation distributions in kidney-renal-clear-cell-carcinoma using Kullback–Leibler and Jensen–Shannon distance measures IET Systems Biology 11 99–104 [PubMed: 28518060]

[64]. Arsenio J, Kakaradov B, Metz PJ, Kim SH, Yeo GW and Chang JT 2014 Early specification of CD8+ T lymphocyte fates during adaptive immunity revealed by single-cell gene-expression analyses Nat Immunol 15 365–72 [PubMed: 24584088]

[65]. Kinney JB and Atwal GS 2014 Equitability, mutual information, and the maximal information coefficient PNAS 111 3354–9 [PubMed: 24550517]

[66]. Kraskov A, Stögbauer H and Grassberger P 2004 Estimating mutual information Phys. Rev. E 69 066138

[67]. Nicoletti G and Busiello DM 2021 Mutual Information Disentangles Interactions from Changing Environments Phys. Rev. Lett 127 228301

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[68]. Sarra L, Aiello A and Marquardt F 2021 Renormalized Mutual Information for Artificial Scientific Discovery Phys. Rev. Lett 126 200601

[69]. Mora T and Bialek W 2011 Are Biological Systems Poised at Criticality? J Stat Phys 144 268–302

[70]. Slonim N, Atwal GS, Tkačik G and Bialek W 2005 Information-based clustering PNAS 102 18297–302 [PubMed: 16352721]

[71]. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Favera RD and Califano A 2006 ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context BMC Bioinformatics 7 S7

[72]. Krishnaswamy S, Spitzer MH, Mingueneau M, Bendall SC, Litvin O, Stone E, Pe'er D and Nolan GP 2014 Conditional density-based analysis of T cell signaling in single-cell data Science 346

[73]. Archer E, Park IM and Pillow J 2012 Bayesian estimation of discrete entropy with mixtures of stick-breaking priors Advances in Neural Information Processing Systems vol 25 (Curran Associates, Inc.)

[74]. Massey JL 1990 Causality, Feedback And Directed Information (Proc. IEEE International Symposium on Information Theory and Its Applications)

[75]. Qiu X, Rahimzamani A, Wang L, Ren B, Mao Q, Durham T, McFaline-Figueroa JL, Saunders L, Trapnell C and Kannan S 2020 Inferring Causal Gene Regulatory Networks from Coupled Single-Cell Expression Dynamics Using Scribe Cell Systems 10 265–274.e11 [PubMed: 32135093]

[76]. Crutchfield JP and Feldman DP 2003 Regularities unseen, randomness observed: Levels of entropy convergence Chaos 13 25–54 [PubMed: 12675408]

[77]. Barrett AB and Seth AK 2011 Practical Measures of Integrated Information for Time-Series Data PLOS Computational Biology 7 e1001052

[78]. Tononi G, Boly M, Massimini M and Koch C 2016 Integrated information theory: from consciousness to its physical substrate Nature Reviews Neuroscience 17 450–61 [PubMed: 27225071]

[79]. Ay N 2015 Information Geometry on Complexity and Stochastic Interaction Entropy 17 2432–58

[80]. Tang Y, Yuan R and Ao P 2014 Summing over trajectories of stochastic dynamics with multiplicative noise J. Chem. Phys 141 044125

[81]. Seifert U 2012 Stochastic thermodynamics, fluctuation theorems and molecular machines Rep. Prog. Phys 75 126001

[82]. Parrondo JMR, Horowitz JM and Sagawa T 2015 Thermodynamics of information Nature Physics 11 131–9

[83]. Hasegawa Y 2018 Multidimensional biochemical information processing of dynamical patterns Phys. Rev. E 97 022401

[84]. Munakata T and Kamiyabu M 2006 Stochastic resonance in the FitzHugh-Nagumo model from a dynamic mutual information point of view Eur. Phys. J. B 53 239–43

[85]. Kholodenko B, Yaffe MB and Kolch W 2012 Computational Approaches for Analyzing Information Flow in Biological Networks Sci. Signal 5 re1–re1 [PubMed: 22510471]

[86]. Mugler A, Walczak AM and Wiggins CH 2010 Information-Optimal Transcriptional Response to Oscillatory Driving Phys. Rev. Lett 105 058101

[87]. de Ronde W H, Tostevin F and ten Wolde P R 2010 Effect of feedback on the fidelity of information transmission of time-varying signals Phys. Rev. E 82 031914

[88]. Tkačik G, Callan CG and Bialek W 2008 Information capacity of genetic regulatory elements Phys. Rev. E 78 011910

[89]. Crisanti A, De Martino A and Fiorentino J 2018 Statistics of optimal information flow in ensembles of regulatory motifs Phys. Rev. E 97 022407

[90]. Micali G and Endres RG 2019 Maximal information transmission is compatible with ultrasensitive biological pathways Scientific Reports 9 16898 [PubMed: 31729454]

[91]. Tkačik G, Walczak AM and Bialek W 2009 Optimizing information flow in small genetic networks Phys. Rev. E 80 031920

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[92]. Tkačik G, Walczak AM and Bialek W 2012 Optimizing information flow in small genetic networks. III. A self-interacting gene Phys. Rev. E 85 041903

[93]. Walczak AM, Tkačik G and Bialek W 2010 Optimizing information flow in small genetic networks. II. Feed-forward interactions Phys. Rev. E 81 041905

[94]. Rieckh G and Tkačik G 2014 Noise and Information Transmission in Promoters with Multiple Internal States Biophysical Journal 106 1194–204 [PubMed: 24606943]

[95]. Tabbaa OP and Jayaprakash C 2014 Mutual information and the fidelity of response of gene regulatory models Phys. Biol 11 046004

[96]. Grabowski F, Czyż P, Kochańczyk M and Lipniacki T 2019 Limits to the rate of information transmission through the MAPK pathway Journal of The Royal Society Interface 16 20180792

[97]. Mehta P, Goyal S, Long T, Bassler BL and Wingreen NS 2009 Information processing and signal integration in bacterial quorum sensing Molecular Systems Biology 5 325 [PubMed: 19920810]

[98]. Thomas PJ and Eckford AW 2016 Capacity of a Simple Intercellular Signal Transduction Channel IEEE Transactions on Information Theory 62 7358–82

[99]. Cepeda-Humerez SA, Ruess J and Tkačik G 2019 Estimating information in time-varying signals PLOS Computational Biology 15 e1007290

[100]. Duso L and Zechner C 2019 Path mutual information for a class of biochemical reaction networks 2019 IEEE 58th Conference on Decision and Control (CDC) 2019 IEEE 58th Conference on Decision and Control (CDC) pp 6610–5

[101]. Sarkar S, Tack D and Ross D 2020 Sparse estimation of mutual information landscapes quantifies information transmission through cellular biochemical reaction networks Commun Biol 3 1–8 [PubMed: 31925316]

[102]. Jetka T, Nienałtowski K, Filippi S, Stumpf MPH and Komorowski M 2018 An information-theoretic framework for deciphering pleiotropic and noisy biochemical signaling Nat. Commun 9 4591 [PubMed: 30389942]

[103]. Domedel-Puig N, Rué P, Pons AJ and García-Ojalvo J 2011 Information Routing Driven by Background Chatter in a Signaling Network PLOS Computational Biology 7 e1002297

[104]. Komorowski M and Tawfik DS 2019 The Limited Information Capacity of Cross-Reactive Sensors Drives the Evolutionary Expansion of Signaling Cell Systems 8 76–85.e6 [PubMed: 30660612]

[105]. Hormoz S 2013 Cross Talk and Interference Enhance Information Capacity of a Signaling Pathway Biophysical Journal 104 1170–80 [PubMed: 23473500]

[106]. Liang XS and Kleeman R 2005 Information Transfer between Dynamical System Components Phys. Rev. Lett 95 244101

[107]. Liang XS 2008 Information flow within stochastic dynamical systems Phys. Rev. E 78 031113

[108]. Rodrigo G and Poyatos JF 2016 Genetic Redundancies Enhance Information Transfer in Noisy Regulatory Circuits PLOS Computational Biology 12 e1005156

[109]. Vazquez-Jimenez A and Rodriguez-Gonzalez J 2019 On Information Extraction and Decoding Mechanisms Improved by Noisy Amplification in Signaling Pathways Scientific Reports 9 14365 [PubMed: 31591406]

[110]. Mc Mahon S S, Sim A, Filippi S, Johnson R, Liepe J, Smith D and Stumpf MPH 2014 Information theory and signal transduction systems: From molecular information processing to network inference Seminars in Cell & Developmental Biology 35 98–108 [PubMed: 24953199]

[111]. Liu P, Wang H, Huang L and Zhou T 2017 The dynamic mechanism of noisy signal decoding in gene regulation Scientific Reports 7 42128 [PubMed: 28176840]

[112]. Maity A and Wollman R 2020 Information transmission from NFkB signaling dynamics to gene expression PLOS Computational Biology 16 e1008011

[113]. Cheong R, Rhee A, Wang CJ, Nemenman I and Levchenko A 2011 Information Transduction Capacity of Noisy Biochemical Signaling Networks Science 334 354–8 [PubMed: 21921160]

[114]. Tudelska K, Markiewicz J, Kochańczyk M, Czerkies M, Prus W, Korwek Z, Abdi A, Błoński S, Kaźmierczak B and Lipniacki T 2017 Information processing in the NF-κB pathway Scientific Reports 7 15926 [PubMed: 29162874]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[115]. Adelaja A, Taylor B, Sheu KM, Liu Y, Luecke S and Hoffmann A 2021 Six distinct NFκB signaling codons convey discrete information to distinguish stimuli and enable appropriate macrophage responses Immunity 54 916–930.e7 [PubMed: 33979588]

[116]. Uda S, Saito TH, Kudo T, Kokaji T, Tsuchiya T, Kubota H, Komori Y, Ozaki Y and Kuroda S 2013 Robustness and Compensation of Information Transmission of Signaling Pathways Science 341 558–61 [PubMed: 23908238]

[117]. Ruiz R, de la Cruz Fand Fernandez-Lopez R 2018 Negative feedback increases information transmission, enabling bacteria to discriminate sublethal antibiotic concentrations Science Advances 4 eaat5771

[118]. Voliotis M, Perrett RM, McWilliams C, McArdle CA and Bowsher CG 2014 Information transfer by leaky, heterogeneous, protein kinase signaling systems PNAS 111 E326–33 [PubMed: 24395805]

[119]. Yu RC, Pesce CG, Colman-Lerner A, Lok L, Pincus D, Serra E, Holl M, Benjamin K, Gordon A and Brent R 2008 Negative feedback that improves information transmission in yeast signalling Nature 456 755–61 [PubMed: 19079053]

[120]. Rhee A, Cheong R and Levchenko A 2014 Noise decomposition of intracellular biochemical signaling networks using nonequivalent reporters PNAS 111 17330–5 [PubMed: 25404303]

[121]. Potter GD, Byrd TA, Mugler A and Sun B 2017 Dynamic Sampling and Information Encoding in Biochemical Networks Biophysical Journal 112 795–804 [PubMed: 28256238]

[122]. Papana A and Kugiumtzis D 2009 Evaluation of mutual information estimators for time series Int. J. Bifurcation Chaos 19 4197–215

[123]. Khan S, Bandyopadhyay S, Ganguly AR, Saigal S, Erickson DJ, Protopopescu V and Ostrouchov G 2007 Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data Phys. Rev. E 76 026209

[124]. Holmes CM and Nemenman I 2019 Estimation of mutual information for real-valued data with error bars and controlled bias Phys. Rev. E 100 022404

[125]. Jetka T, Nienałtowski K, Winarski T, Błoński S and Komorowski M 2019 Information-theoretic analysis of multivariate single-cell signaling responses PLOS Computational Biology 15 e1007132

[126]. Zeng X, Xia Y and Tong H 2018 Jackknife approach to the estimation of mutual information Proceedings of the National Academy of Sciences 115 9956–61

[127]. Uda S and Kuroda S 2016 Analysis of cellular signal transduction from an information theoretic approach Seminars in Cell & Developmental Biology 51 24–31 [PubMed: 26710655]

[128]. Makadia HK, Schwaber JS and Vadigepalli R 2015 Intracellular Information Processing through Encoding and Decoding of Dynamic Signaling Features PLOS Computational Biology 11 e1004563

[129]. Hansen AS and O'Shea EK 2015 Limits on information transduction through amplitude and frequency regulation of transcription factor activity ed N Barkai eLife 4 e06559

[130]. Granados AA, Pietsch JMJ, Cepeda-Humerez SA, Farquhar IL, Tkačik G and Swain PS 2018 Distributed and dynamic intracellular organization of extracellular information PNAS 115 6088–93 [PubMed: 29784812]

[131]. Chen SY, Osimiri LC, Chevalier M, Bugaj LJ, Nguyen TH, Greenstein RA, Ng AH, Stewart-Ornstein J, Neves LT and El-Samad H 2020 Optogenetic Control Reveals Differential Promoter Interpretation of Transcription Factor Nuclear Translocation Dynamics Cell Systems 11 336–353.e24 [PubMed: 32898473]

[132]. Hao N and O'Shea EK 2012 Signal-dependent dynamics of transcription factor translocation controls gene expression Nature Structural & Molecular Biology 19 31–9

[133]. Purvis JE and Lahav G 2013 Encoding and Decoding Cellular Information through Signaling Dynamics Cell 152 945–56 [PubMed: 23452846]

[134]. Sen S, Cheng Z, Sheu KM, Chen YH and Hoffmann A 2020 Gene Regulatory Strategies that Decode the Duration of NFκB Dynamics Contribute to LPS- versus TNF-Specific Gene Expression Cell Systems 10 169–182.e5 [PubMed: 31972132]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[135]. Hledík M, Sokolowski TR and Tkačik G 2019 A Tight Upper Bound on Mutual Information 2019 IEEE Information Theory Workshop (ITW) 2019 IEEE Information Theory Workshop (ITW) pp 1–5

[136]. Hasegawa Y 2016 Optimal temporal patterns for dynamical cellular signaling New J. Phys 18 113031

[137]. Billing U, Jetka T, Nortmann L, Wundrack N, Komorowski M, Waldherr S, Schaper F and Dittrich A 2019 Robustness and Information Transfer within IL-6-induced JAK/STAT Signalling Communications Biology 2 1–14 [PubMed: 30740537]

[138]. Stewart-Ornstein J, Cheng HW (Jacky) and Lahav G 2017 Conservation and Divergence of p53 Oscillation Dynamics across Species Cell Systems 5 410–417.e4 [PubMed: 29055670]

[139]. Bahl L, Brown P, de Souza P and Mercer R 1986 Maximum mutual information estimation of hidden Markov model parameters for speech recognition ICASSP ’86. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP ’86. IEEE International Conference on Acoustics, Speech, and Signal Processing vol 11 pp 49–52

[140]. Chen P, Liu R, Aihara K and Chen L 2020 Autoreservoir computing for multistep ahead prediction based on the spatiotemporal information transformation Nat. Commun 11 4568 [PubMed: 32917894]

[141]. Lim B, Arik SO, Loeff N and Pfister T 2020 Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting arXiv:1912.09363

[142]. Suderman R, Bachman JA, Smith A, Sorger PK and Deeds EJ 2017 Fundamental trade-offs between information flow in single cells and cellular populations PNAS 114 5755–60 [PubMed: 28500273]

[143]. Chevalier M, Venturelli O and El-Samad H 2015 The Impact of Different Sources of Fluctuations on Mutual Information in Biochemical Networks PLOS Computational Biology 11 e1004462

[144]. Zhang Q, Gupta S, Schipper DL, Kowalczyk GJ, Mancini AE, Faeder JR and Lee REC 2017 NF-κB Dynamics Discriminate between TNF Doses in Single Cells Cell Systems 5 638–645.e5 [PubMed: 29128333]

[145]. Nienałtowski K, Rigby RE, Walczak J, Zakrzewska KE, Głów E, Rehwinkel J and Komorowski M 2021 Fractional response analysis reveals logarithmic cytokine responses in cellular populations Nat Commun 12 4175 [PubMed: 34234126]

[146]. Achar SR, Bourassa FXP, Rademaker TJ, Lee A, Kondo T, Salazar-Cavazos E, Davies JS, Taylor N, François P and Altan-Bonnet G 2022 Universal antigen encoding of T cell activation from high-dimensional cytokine dynamics Science 376 880–4 [PubMed: 35587980]

[147]. Rivoire O and Leibler S 2011 The Value of Information for Populations in Varying Environments J Stat Phys 142 1124–66

[148]. Reddy G, Wong-Ng J, Celani A, Sejnowski TJ and Vergassola M 2018 Glider soaring via reinforcement learning in the field Nature 562 236–9 [PubMed: 30232456]

[149]. Gregor T, Tank DW, Wieschaus EF and Bialek W 2007 Probing the Limits to Positional Information Cell 130 153–64 [PubMed: 17632062]

[150]. Mugler A, Tostevin Fand ten Wolde PR 2013 Spatial partitioning improves the reliability of biochemical signaling PNAS 110 5927–32 [PubMed: 23530194]

[151]. Petkova MD, Tkačik G, Bialek W, Wieschaus EF and Gregor T 2019 Optimal Decoding of Cellular Identities in a Genetic Network Cell 176 844–855.e15 [PubMed: 30712870]

[152]. Linding R and Klipp E 2021 Shapes of Cell Signaling Current Opinion in Systems Biology 100354

[153]. Dong C, Loy CC, He K and Tang X 2016 Image Super-Resolution Using Deep Convolutional Networks IEEE Transactions on Pattern Analysis and Machine Intelligence 38 295–307 [PubMed: 26761735]

[154]. Toettcher JE, Weiner OD and Lim WA 2013 Using Optogenetics to Interrogate the Dynamic Control of Signal Transmission by the Ras/Erk Module Cell 155 1422–34 [PubMed: 24315106]

[155]. Wilson MZ, Ravindran PT, Lim WA and Toettcher JE 2017 Tracing Information Flow from Erk to Target Gene Induction Reveals Mechanisms of Dynamic and Combinatorial Control Molecular Cell 67 757–769.e5 [PubMed: 28826673]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[156]. Jaynes ET 2003 Probability Theory: The Logic of Science (Cambridge University Press)

[157]. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P and Hassabis D 2021 Highly accurate protein structure prediction with AlphaFold Nature 596 583–9 [PubMed: 34265844]

[158]. Rademaker TJ, Bengio E and François P 2019 Attack and Defense in Cellular Decision-Making: Lessons from Machine Learning Phys. Rev. X 9 031012

[159]. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC and Teichmann SA 2015 The Technology and Biology of Single-Cell RNA Sequencing Molecular Cell 58 610–20 [PubMed: 26000846]

[160]. Vera M, Biswas J, Senecal A, Singer RH and Park HY 2016 Single-Cell and Single-Molecule Analysis of Gene Expression Regulation Annu. Rev. Genet 50 267–91 [PubMed: 27893965]

[161]. Gillespie DT 1977 Exact stochastic simulation of coupled chemical reactions J. Phys. Chem 81 2340–61

[162]. Dixit PD, Lyashenko E, Niepel M and Vitkup D 2020 Maximum Entropy Framework for Predictive Inference of Cell Population Heterogeneity and Responses in Signaling Networks Cell Systems 10 204–212.e8 [PubMed: 31864963]

[163]. Loos C and Hasenauer J 2019 Mathematical modeling of variability in intracellular signaling Current Opinion in Systems Biology 16 17–24

[164]. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA and Trapnell C 2017 Reversed graph embedding resolves complex single-cell trajectories Nature Methods 14 979–82 [PubMed: 28825705]

[165]. Saelens W, Cannoodt R, Todorov H and Saeys Y 2019 A comparison of single-cell trajectory inference methods Nature Biotechnology 37 547–54

[166]. Fischer DS, Fiedler AK, Kernfeld EM, Genga RMJ, Bastidas-Ponce A, Bakhti M, Lickert H, Hasenauer J, Maehr R and Theis FJ 2019 Inferring population dynamics from single-cell RNA-sequencing time series data Nature Biotechnology 37 461–8

[167]. Schiebinger G, Shu J, Tabaka M, Cleary B, Subramanian V, Solomon A, Gould J, Liu S, Lin S, Berube P, Lee L, Chen J, Brumbaugh J, Rigollet P, Hochedlinger K, Jaenisch R, Regev A and Lander ES 2019 Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming Cell 176 928–943.e22

[168]. Weinreb C, Wolock S, Tusi BK, Socolovsky M and Klein AM 2018 Fundamental limits on dynamic inference from single-cell snapshots PNAS 115 E2467–76 [PubMed: 29463712]

[169]. Qiu X, Zhang Y, Martin-Rufino JD, Weng C, Hosseinzadeh S, Yang D, Pogson AN, Hein MY, Hoi (Joseph) Min K, Wang L, Grody EI, Shurtleff MJ, Yuan R, Xu S, Ma Y, Replogle JM, Lander ES, Darmanis S, Bahar I, Sankaran VG, Xing J and Weissman JS 2022 Mapping transcriptomic vector fields of single cells Cell 185 690–711.e45 [PubMed: 35108499]

[170]. Chen KH, Boettiger AN, Moffitt JR, Wang S and Zhuang X 2015 Spatially resolved, highly multiplexed RNA profiling in single cells Science 348

[171]. Lahav G, Rosenfeld N, Sigal A, Geva-Zatorsky N, Levine AJ, Elowitz MB and Alon U 2004 Dynamics of the p53-Mdm2 feedback loop in individual cells Nat. Genet 36 147–50 [PubMed: 14730303]

[172]. Regot S, Hughey JJ, Bajar BT, Carrasco S and Covert MW 2014 High-Sensitivity Measurements of Multiple Kinase Activities in Live Single Cells Cell 157 1724–34 [PubMed: 24949979]

[173]. Lindner M, Vicente R, Priesemann V and Wibral M 2011 TRENTOOL: A Matlab open source toolbox to analyse information flow in time series data with transfer entropy BMC Neuroscience 12 119 [PubMed: 22098775]

[174]. Mayner WGP, Marshall W, Albantakis L, Findlay G, Marchman R and Tononi G 2018 PyPhi: A toolbox for integrated information theory PLOS Computational Biology 14 e1006343

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Box 1.

The formulation of mutual information applied to stimulus discrimination via intracellular signaling.

Mutual information for stimulus discrimination.

Despite the interest in noise [5–8], mutual information serves as a fundamental quantity to understand information processing in biology [69] and has been widely used in biological systems [24,30,38]. For example, mutual information leads to a method to cluster genes [70], reconstruct the network of gene interactions [71], and quantify strengths of the influence between proteins [72].

For the focus of this review, intracellular signaling process, mutual information can be employed to measure the stimulus discrimination by signaling responses, where one random variable is the categorical stimulus set and the other is a set of the signaling responses under each stimulus. Specifically, the mutual information between M conditions chosen in an experiment as the stimuli set (S) and the signaling responses set (R) is:

$$I(R;S) = H(R) - H(R \mid S), \tag{5}$$

where $H(R \mid S)$ and $H(R)$ are the conditional and unconditional entropy from the definition in Eq.(3). The mutual information between the extracellular stimulus conditions and the intracellular signaling responses quantifies the amount of information about the stimulus conditions (identity and dose).

Channel capacity of intracellular signaling channels.

With the probability distribution of the M stimulus conditions $\boldsymbol{q} = \{q_1, q_2, \ldots, q_M\}$, the maximum mutual information is obtained by maximization with respect to this probability distribution:

$$I_{max}(R;S) = \max_{\boldsymbol{q}} I(R;S), \tag{6}$$

under the constraint of $q_1 + q_2 + \cdots + q_M = 1$ and $q_i \geq 0$. The maximization is useful especially when the stimulus distribution is empirically unknown. This is a Bayesian approach [30], where the prior for the probability distribution is typically assumed uniform, as uniform priors seem especially effective [19,73].

The maximum mutual information depends on the stimulus conditions under consideration: if M distinct conditions were considered, perfectly transmitted information leads to $\log_2 M$ bits, corresponding to the prior of a uniform distribution. A smaller value implies that the cells cannot fully discriminate the stimuli via the signaling response. With increased stimulus conditions in an experiment, the maximum mutual information approximates to the channel capacity through the signaling molecules. In addition, mutual information in trajectory space $I(R_{1:n}; S)$ can be formulated similarly, where signaling responses are time series data. The maximization in $I_{max}(R_{1:n}; S)$ can be

conducted by using data up to the timepoint n, which quantifies the maximum extent of information transmission cumulatively.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Box 2.

Software packages of calculating information-theoretic quantities for intracellular signaling.

The estimation of entropy, Kullback–Leibler divergence, mutual Information and channel capacity can be found in the R package (http://strimmerlab.org/software/entropy) and the MATLAB package (https://github.com/maximumGain/information-theory-tool). Another Python and MATLAB packages for calculating entropy and mutual information can be found (https://github.com/robince). The MATLAB toolbox to evaluate the transfer entropy are provided by (https://figshare.com/articles/code/MuTE_toolbox_to_evaluate_Multivariate_Transfer_Entropy/1005245) and (https://github.com/trentool/TRENTOOL3) [173]. A Python package (https://github.com/wmayner/pyphi) computes the integrated information [174].

For evaluating the mutual information from the time series data, the k-nearest-neighbor approach [62,66] was in a Python package (https://github.com/pawel-czyz/channel-capacityestimator). The decoding-based method [130] was implemented by a MATLAB package and a R package (https://github.com/swainlab/mi-by-decoding). The Statistical Learning-based Estimation of Mutual Information (SLEMI) [125,137] has a R package in CRAN (https://github.com/sysbiosig/SLEMI). The approach by using stochastic dynamical models [49] has a MATLAB package (https://github.com/signalingsystemslab/dMI).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

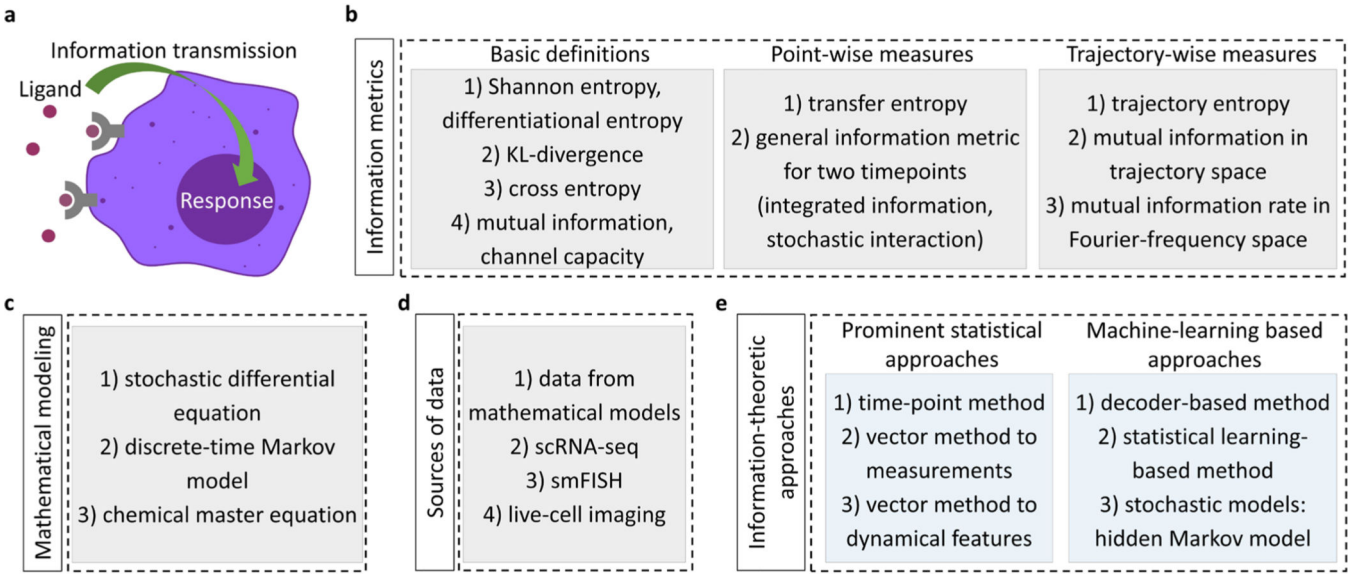

Figure 1. A summary on the major ingredients of information-theoretic approaches to be reviewed.

(a) A schematic figure of the biological question of transmitting environmental information through intracellular signaling. (b-e) We summarize the ingredients of quantitatively studying information transmission: (b) the information metrics, (c) the mathematical models, (d) the data sources, and (e) the data-driven approaches, which are further categorized as traditional statistical approaches and more recent machine-learning based methods. Each of the topics will be covered in the following sections.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

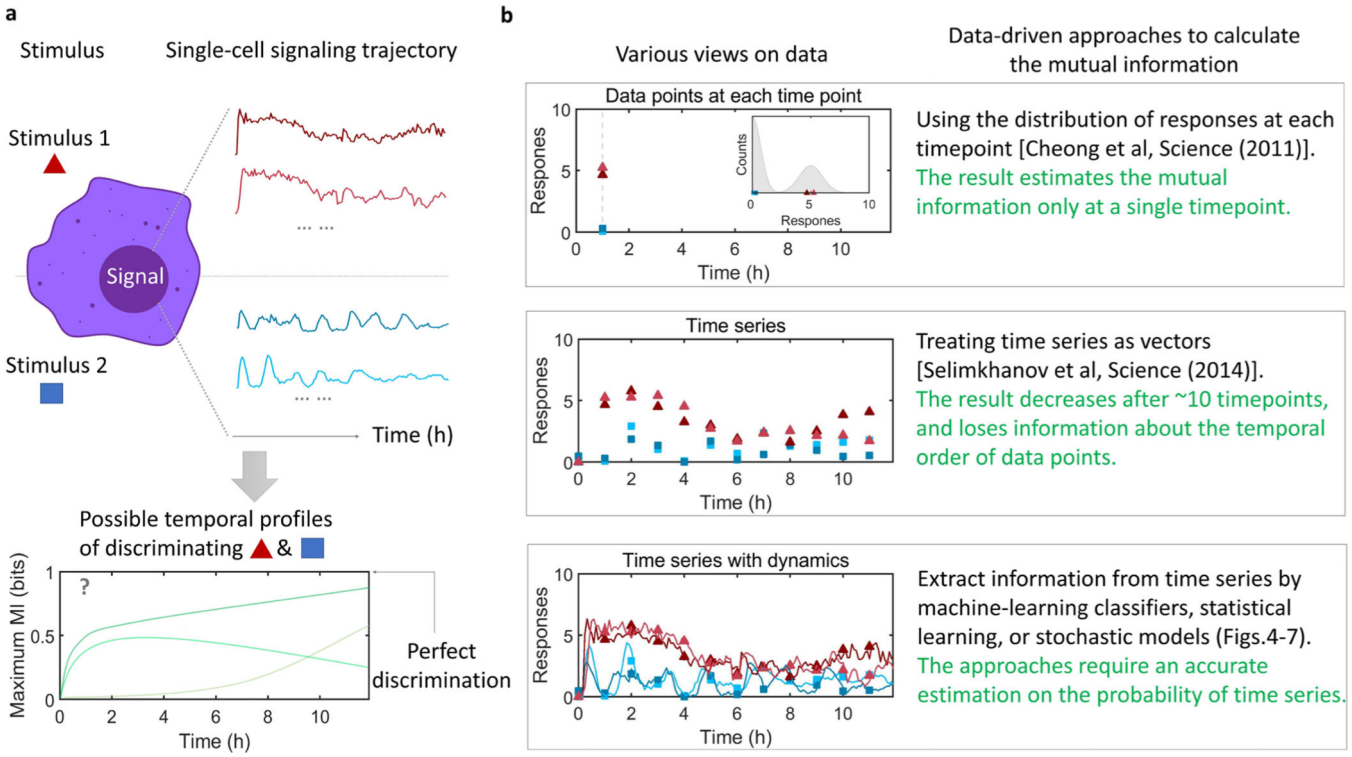

Figure 2. The approaches for estimating information transmission by using single-cell signaling responses.

(a) A schematic figure on using the single-cell live imaging measurement on signaling responses to calculate mutual information, which quantifies the stimulus discrimination. (b) A schematic on the methods of using (upper) a single-timepoint data, (middle) a few timepoints, and (below) long time series. Reproduced from [49]. Copyright (2021), from Springer Nature.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

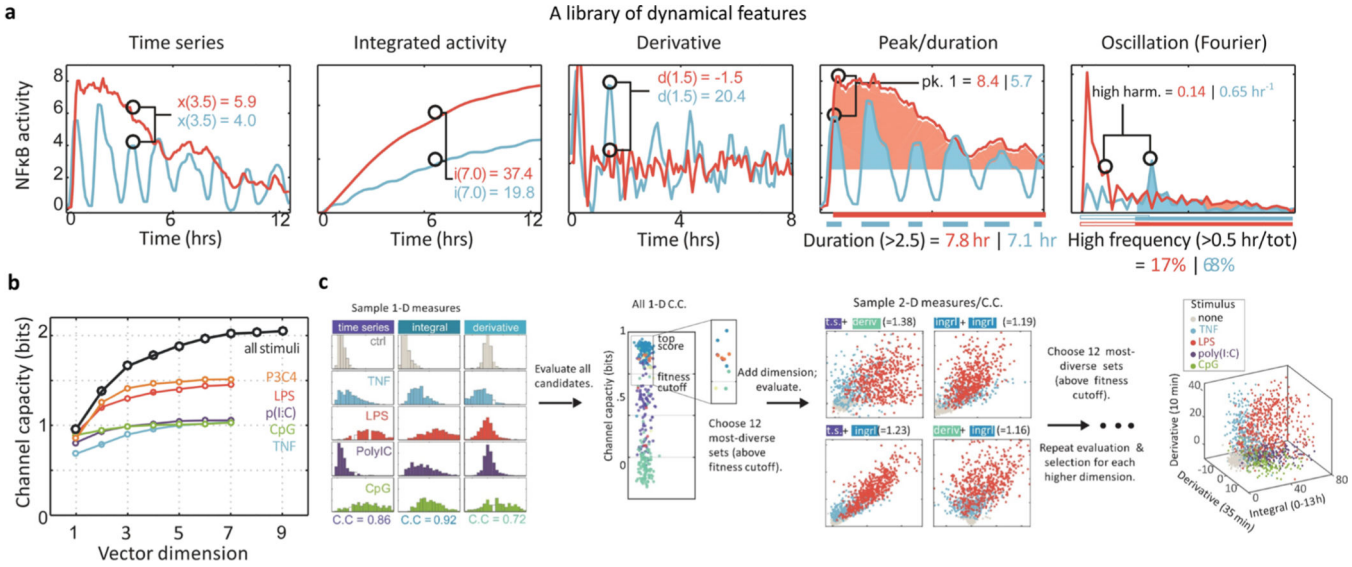

Figure 3. The vector method with dynamical features on estimating information transmission.

(a) A library of dynamical features was calculated for the long time series data of NFκB signaling responses. (b) The channel capacity was evaluated by the k-nearest neighbor estimation on the most informative dynamical features, for all stimuli and for different doses of one stimulus, as indicated. (c) The protocol of searching for the most informative combination of features. Reproduced from [115]. Copyright (2021), from Elsevier.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

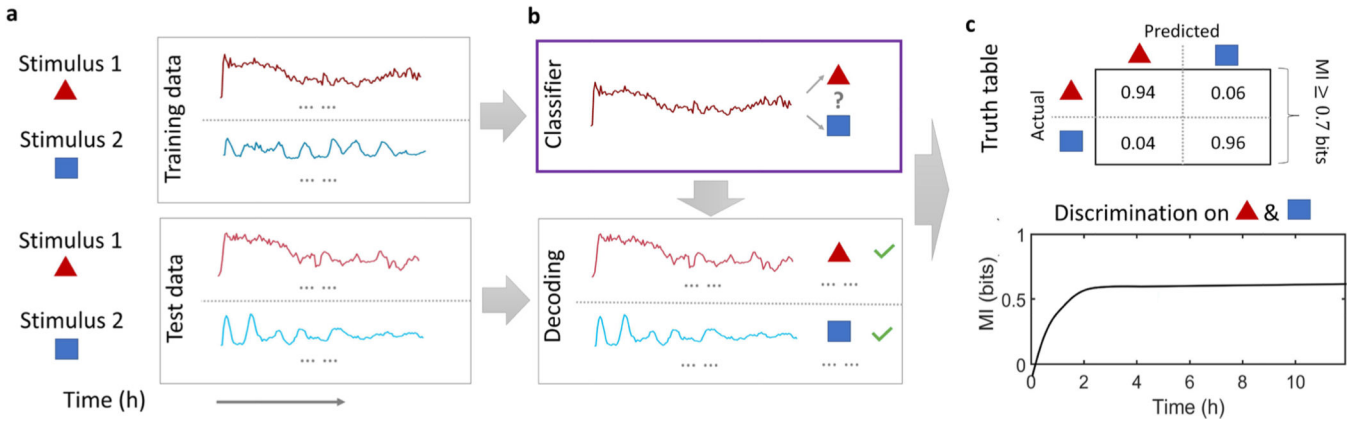

Figure 4. The decoding-based approaches to calculate the mutual information from time series of signaling responses.

(a) The measured signaling responses under various stimulus conditions are used as the training data and the test data. (b) A classifier is trained by the training data and used to recognize the stimulus condition for the test data. (c) The truth table by the classifier gives an estimated lower bound on the mutual information for the stimulus discrimination. Reproduced with permission from [130]. Copyright (2018), from PNAS.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

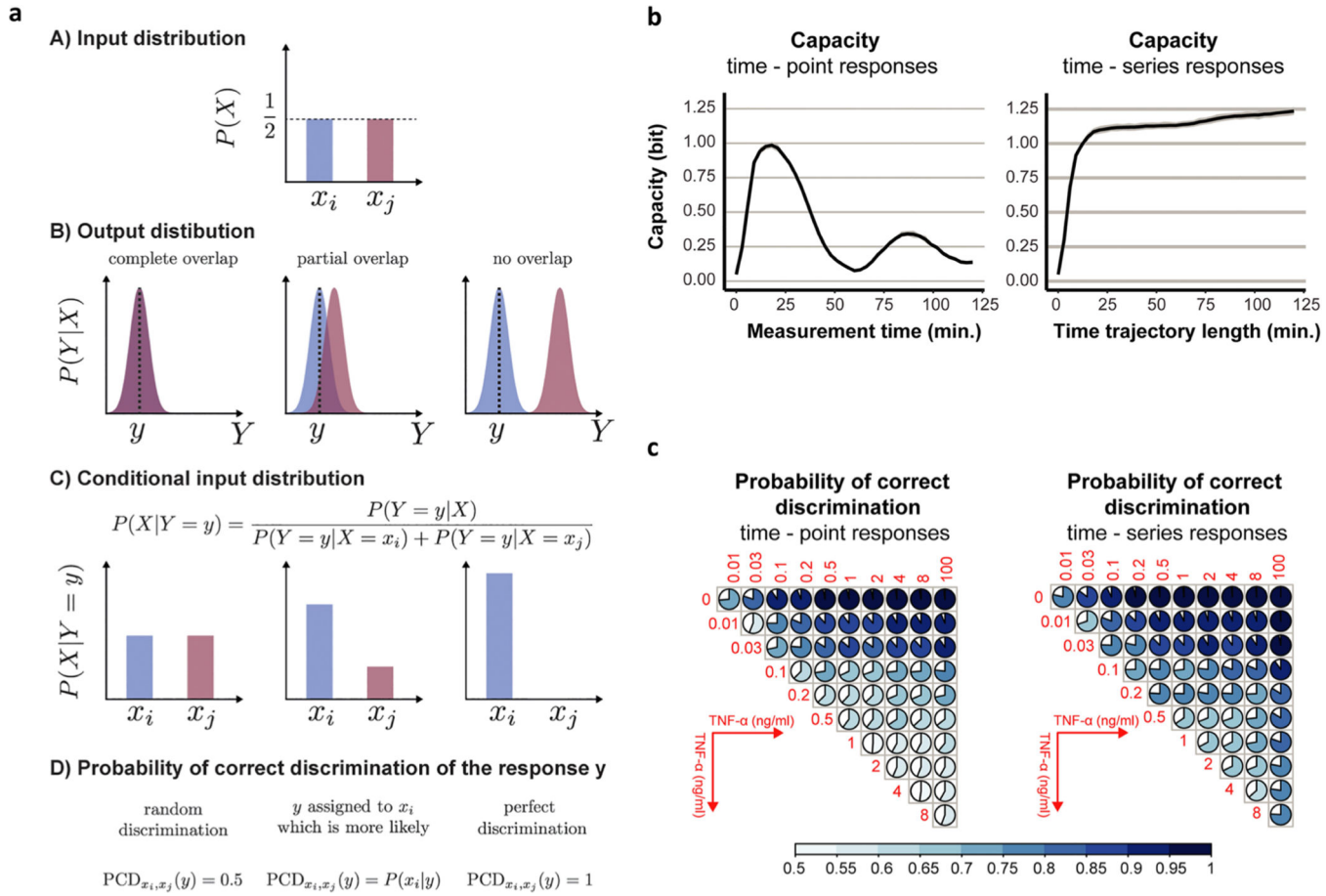

Figure 5. The statistical learning-based estimation of mutual information.

(a) A schematic figure for the probabilities of discriminating two inputs. The input distribution $P(X)$ and the conditional output probabilities $P(Y \mid X)$ lead to the conditional input distributions $P(X \mid Y)$ by Bayes formula. (b, c) Information-theoretic analysis of NFκB signaling responses to the TNFα stimulus. (b) The channel capacity as a function of time by using a single-timepoint data individually and time series. (c) The probabilities (color filled fraction of the circle marks) of correct pairwise discrimination between TNFα concentrations for the 21-minute responses and time series. See a full description on the figure and symbols in the original paper. Adapted from [125]. Copyright (2021), from Elsevier.

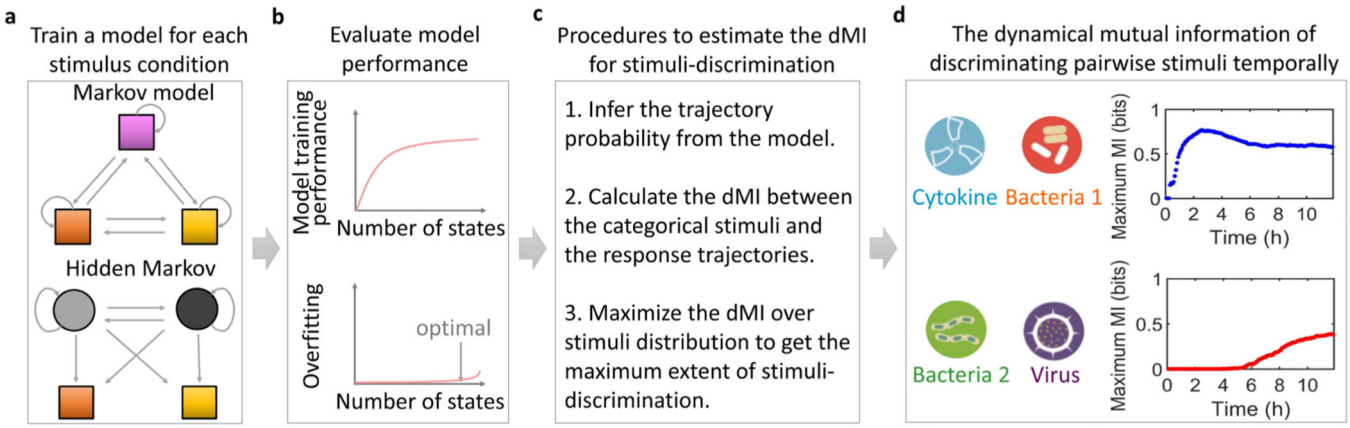

Figure 6. Quantifying the dynamical mutual information by using the stochastic models.

Stochastic models such as the hidden Markov model can be used to learn the signaling dynamics, reproduce data, infer the trajectory probabilities, and evaluate the mutual information. (a) The stochastic models to learn the data. (b) The evaluation on the model performance when identifying the proper number of parameters. (c) The procedures on calculating the mutual information from the trained stochastic models. (d) The estimated mutual information encoded in dynamics reveal the temporal ordering of discriminating certain stimuli pairs. Reproduced from [49]. Copyright (2021), from Springer Nature.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

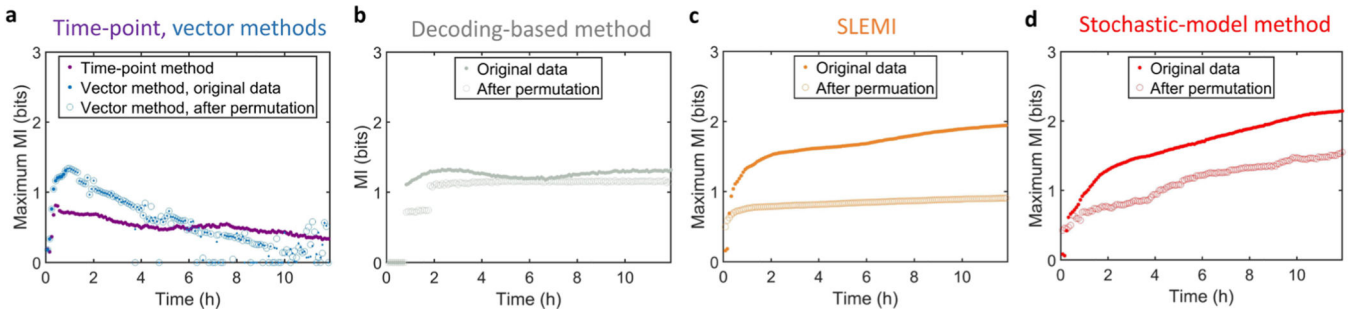

Figure 7. A comparison on the data-driven approaches for time series data.

The data is the NFkB signaling responses under 13 different stimulus conditions (17 conditions in total with replicates) [49]. The y-axis is labeled as “Maximum MI”, except for the decoding-based method providing a lower bound (y-axis is “MI” without “Maximum”) (a) The time-point method [113] and the vector method [62]. (b) The decoding-based method [130] by using the first 10 principle components and default parameters. (c) The statistical learning-based estimation of mutual information (SLEMI) [125], with parameters “boot_num”=10, “boot_prob”=0.8, ”testing_cores”=4 in the numerical package. (d) The stochastic model-based method [49] with 64 hidden states and 32 emission states for the hidden Markov model. The computational time of one bootstrap for the five methods is ~10 minutes, ~1 hour, ~10 minutes, ~10 minutes, ~10 hours on personal desktop with intel(R) core(tm) i7–8700 CPU@3.7 GHz. Reproduced from [49]. Copyright (2021), from Springer Nature.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

The fundamental information quantities useful for intracellular information processing. The three categories of information metrics: the basic definitions in information theory, the pointwise information measures, and the trajectory-wise measures.

Information quantities	Mathematical definition	References
Shannon entropy, differentiational entropy	$H(X) = -\sum_{x\in X} P(x)\log_2 P(x)$, $H_{diff}(X) = -\int_{-\infty}^{+\infty} dx\ p(x)\log_2 p(x)$	[14]
KL-divergence	$D_{KL}(P \parallel Q) = \sum_{x\in X} P(x)\log_2[P(x)/Q(x)]$	[43]
Cross entropy	$CE(P \parallel Q) = D_{KL}(P \parallel Q) + H(P)$	[44]
Mutual information, channel capacity	$I(Y;X) = H(Y) - H(Y \mid X)$, $I_{max}(Y;X) = \max_{P_X(x)} I(Y;X)$	[45]
Transfer entropy	$T_{X\to Y} = H(y_n \mid y_{n-1}) - H(y_n \mid y_{n-1}, x_{n-1})$	[46]
General information metric for two timepoints (integrated information, stochastic interaction)	$ci(x_i \to y_j) = \min_{q(x_{i,j}, y_{i,j})} D_{KL}[p(x_{i,j}, y_{i,j}) \parallel q(x_{i,j}, y_{i,j})]$	[47]
Trajectory entropy	$H(y_{1:n}) = -\log_2 p(y_{1:n})$	[48]
Mutual information in trajectory space	$I(Y_{1:n}; X) = H(Y_{1:n}) - H(Y_{1:n} \mid X)$	[49]
Mutual information rate in Fourier-frequency space	$I_R(x_{1:n}; y_{1:n}) = -\frac{1}{4\pi}\int_{-\infty}^{\infty} d\omega \ln\left[1 - \frac{\lvert S^{xy}(\omega)\rvert^2}{S^{xx}(\omega)S^{yy}(\omega)}\right]$	[50]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Sources of single-cell data. Reproduced from [27].

Typical measurements	Data from mathematical models	scRNA-seq	smFISH	Live-cell imaging
# of cells	Model-specific	~100,000	~10,000	~1,000
# of molecules	Model-specific	~10,000	~1,000	~1 or2
Timepoints vs. time series	Time series	Timepoints	Timepoints	Time series