



# Application of information theory in systems biology

Shinsuke Uda<sup>1</sup>

Received: 15 January 2020 / Accepted: 25 February 2020 / Published online: 6 March 2020  
© The Author(s) 2020

## Abstract

Over recent years, new light has been shed on aspects of information processing in cells. The quantification of information, as described by Shannon's information theory, is a basic and powerful tool that can be applied to various fields, such as communication, statistics, and computer science, as well as to information processing within cells. It has also been used to infer the network structure of molecular species. However, the difficulty of obtaining sufficient sample sizes and the computational burden associated with the high-dimensional data often encountered in biology can result in bottlenecks in the application of information theory to systems biology. This article provides an overview of the application of information theory to systems biology, discussing the associated bottlenecks and reviewing recent work.

**Keywords** Information processing · Systems biology · Information theory

## Introduction

Systems biology has contributed greatly to our understanding of life phenomena by helping elucidate their underlying mechanisms from a systems perspective. Various systems approaches can be applied, such as the consideration of mechanics or of biochemical reactions, but in recent years there has been increased focus on information processing in cells (Uda and Kuroda 2016) (Levchenko and Nemenman 2014) (Tkačik and Bialek 2016). That is not to say that investigating cells as information processing systems is entirely novel; in neuroscience, there have been various studies of the mechanisms of information processing in neurons and neuronal networks (Rieke et al. 1997; Timme and Lapish 2018), and the notion of information processing in cells, especially cellular signal transduction, has previously been described (Azeloglu and Iyengar 2015).

In most cells, information processing is implemented by biochemical processes; in neurons, it is implemented through electrical signaling. Collecting quantitative data is generally more difficult for biochemical processes than for the electrical signals in neurons, and this has hampered the analysis of information processing in cells. However, recent developments

in technology have made it possible to quantitatively measure various biochemical processes, and thus to investigate information processing, in a single cell (Cheong et al. 2011; Gregor et al. 2007; Keshelava et al. 2018; Ozaki et al. 2010; Selimkhanov et al. 2014).

Quantitative analysis of information processing requires a definition of the amount of information. This is provided by Shannon's information theory (Cover and Thomas 2006; Shannon 1948). In this theory, information is defined and formulated in the context of communication between a sender and receiver; the definitions are general and can be applied not only to communication, but also to fields such as statistics, machine learning, computer science, and gambling. Systems biology is no exception; the application of Shannon's information theory and its definition of information allows the quantitative analysis of mechanisms of information processing. However, this raises some concerns in terms of sample size, the dimension of the data vector, and the interpretation of the analysis results. In this article, the author describes the efficacy of applying Shannon's information theory to systems biology and discusses notable points specific to biological situations.

✉ Shinsuke Uda  
uda@bioreg.kyushu-u.ac.jp

<sup>1</sup> Division of Integrated Omics, Research Center for Transomics Medicine, Medical Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan

## Information quantification

Data have become increasingly important in this Information Age. Data sets are generally used with the expectation they

contain useful information, but how can the amount of information they contain be quantified?

Consider data sampled from a statistical population, where a random variable  $X$  is generated from a distribution and the outcome  $x$  of  $X$  is a part of the data. All information on  $X$  is contained in the distribution of  $X$  because  $x$  can be generated if the distribution is known. Intuitively, the amount of information contained might be expected to be proportional to the length of the sequence representing the data items sampled from the population. However, this is not necessarily the case. For example, take the example of a coin flip, where  $x = 1$  and  $x = 0$  represent an outcome of heads and tails, respectively. The data set of the observed outcomes for coin flips repeated  $n$  times can be written as a sequence of length  $n$ :  $\{x_i\}_{i=1}^n$ . If, however, the representations of heads and tails are changed to  $x = 11$  and  $x = 00$ , this doubles the length of the sequence. Similarly, the length of the sequence can be increased without limit through the redundant representation of the data, with no change in the amount of information contained in the data. It therefore seems to be essential that measuring the amount of information contained in a data set requires non-redundant representation of the data.

Let  $l_i$  be the description length of event  $i$ , which occurs with probability  $p_i$ . The average description length of the data has the following lower bound (Cover and Thomas 2006):

$$\sum_i p_i l_i \geq -\sum_i p_i \log p_i \equiv H(p) \tag{1}$$

This lower bound corresponds to the entropy  $H$  of distribution  $\{p_i\}$ . Inequality (1) means that the average description length for arbitrary representation of event  $i$  cannot be reduced to less than the value of entropy. In the case of a continuous random variable, the summation is replaced with the equivalent integral. Thus, entropy can be used as a measure to quantify the amount of information contained in a distribution. The most commonly used unit of information is the “bits,” which represents the length of a binary sequence, because most data can be transformed to binary representation.

Equation (1) indicates that a data set of length  $n$  with entropy  $H(p)$  has an average number of states of  $2^{nH(p)}$ , where the base of logarithms is 2 and the unit of entropy is bits. Thus, the entropy can also be interpreted as an index of uncertainty, with larger values indicating greater uncertainty. It may not be immediately intuitive that a measure of uncertainty would be an indicator of the amount of information; however, it may help to keep in mind that uncertainty is defined by the average number of states.

Consider a sender transmitting a message to a receiver. How can the amount of information reliably transmitted between the sender and receiver be quantified? Information transmission is measured according to the mutual information

to the sender’s and receiver’s data  $X$  and  $Y$ , respectively, as follows:

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) = H(Y) - H(Y|X) \\ &= \sum_{x,y} p(x,y) \log \frac{p(y|x)}{p(y)} \end{aligned} \tag{2}$$

where  $p(\cdot)$  indicates a probabilistic distribution function. The conditional distribution  $p(y|x)$  is termed a communication channel in the context of communication, and defines the distribution of  $Y$  given  $X$ ; this can be interpreted as the relationship between  $X$  as the input and  $Y$  as the output (Fig. 1). The term  $2^{I(X;Y)}$ , based on the mutual information  $I(X;Y)$ , corresponds to the average number of states of the sender’s data that the receiver can distinguish. The mutual information can be described by the difference between the entropy and the conditional entropy (Fig. 2), where the conditional entropy  $H(A|B)$  is the amount of information of  $A$  when  $B$  is known. Thus, the mutual information  $I(X; Y)$  can be interpreted as the information that is the residue of information  $X$  (or  $Y$ ) omitted the information when  $Y$  (or  $X$ ) is known. Schematically, this corresponds to the intersection of the information content of  $X$  and  $Y$ . If  $I(X; Y) = 0$ , there is no information transmission between the sender and the receiver. This is equivalent to  $X$  and  $Y$  being statistically independent.

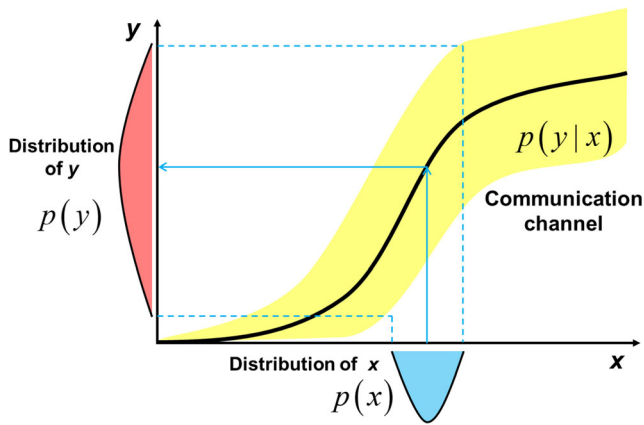
Given the distributions  $p(x)$  and  $p(y|x)$ , the mutual information is uniquely determined by Eq. (2) (Fig. 1), where  $p(x, y) = p(x)p(y|x)$  and  $p(y) = \sum_x p(y|x)p(x)$ . In the transmission process, the channel  $p(y|x)$  typically depends on the communication system used, but the input distribution  $p(x)$  can often be designed. The upper bound for the mutual information, given an input distribution  $p(x)$ , is known as the channel capacity:

$$C = \max_{p(x)} I(X; Y)$$

When applying these concepts to biological situations, the in vivo input distribution is often difficult to measure, so the channel capacity is often calculated instead of evaluating the mutual information from the input distribution.

Although, for simplicity,  $X$  and  $Y$  have so far been denoted as scalar quantities, they can more generally be represented as vectors. When the data form a time series, the information including time effects can be represented by the extension of random variables such as  $X$  and/or  $Y$  from scalar to vector. For simplicity, consider time to be discrete rather than continuous, given by  $t = \{t_i\}$ ; then, the time trajectory of  $X$  can be represented as  $\{X_t\} = (X_1, \dots, X_m)^T$ . The mutual information between time trajectory  $\{X_t\}$  and  $\{Y_t\}$  is given by:

$$I(\{X_t\}; \{Y_t\}) = \sum_{\{x_t\}, \{y_t\}} p(\{x_t\}, \{y_t\}) \log \frac{p(\{y_t\}|\{x_t\})}{p(\{y_t\})} \tag{3}$$



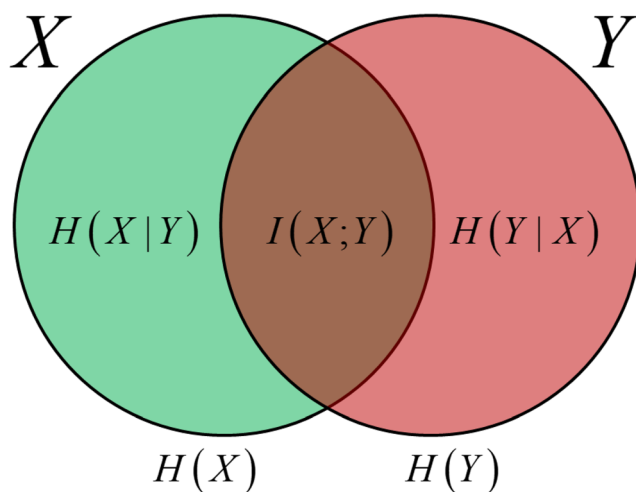
**Fig. 1** The relationship of the distributions of  $x$  as the sender and  $y$  as the receiver. The solid line is the average of  $y$  given  $x$ . The vertical height of the yellow area for a given value of  $x$  is the variability of  $y$  given  $x$

This can be considered to be the information transmission between a sender and receiver (Munakata and Kamiyabu 2006). The mutual information of the snapshot at a single time given by Eq. (2) corresponds to the lower bound of the mutual information of the time trajectory given by Eq. (3).

The mutual information  $I(X; Y)$  does not always indicate the direct statistical interaction between  $X$  and  $Y$ . For example, if  $Z$  interacts with  $X$  and  $Y$ , and  $X$  and  $Y$  are not independent, then even if  $X$  and  $Y$  do not directly interact, there is an indirect interaction between  $X$  and  $Y$  mediated through  $Z$ . The conditional mutual information:

$$I(X; Y|Z) = \sum_{x,y,z} p(x, y, z) \log \frac{p(y|x, z)}{p(y|z)}$$

allows the statistical relationship between  $X$  and  $Y$ , subtracting the effect of  $Z$  to be quantified.



**Fig. 2** Graphical representation of entropy and mutual information. The circles represent the entropies of the random variables  $X$  and  $Y$ . The area of intersection of the two circles corresponds to the mutual information between the variables. The remainder of each circle outside the intersection corresponds to each conditional entropy

Transfer entropy (Hlavackova-Schindler et al. 2007; Palus et al. 2001; Schreiber 2000):

$$\begin{aligned} TE(X \rightarrow Y) &= I(\{y_i\}_{i=t+1}^T; \{x_i\}_{i=t}^{t-\tau_x} | \{y_i\}_{i=t}^{t-\tau_y}) \\ &= \sum_{\{y_i\}_{i=t+1}^T, \{y_i\}_{i=t}^{t-\tau_y}, \{x_i\}_{i=t}^{t-\tau_x}} p(\{y_i\}_{i=t+1}^T, \{y_i\}_{i=t}^{t-\tau_y}, \{x_i\}_{i=t}^{t-\tau_x}) \\ &\quad \log \frac{p(\{y_i\}_{i=t+1}^T | \{y_i\}_{i=t}^{t-\tau_y}, \{x_i\}_{i=t}^{t-\tau_x})}{p(\{y_i\}_{i=t+1}^T | \{y_i\}_{i=t}^{t-\tau_y})} \end{aligned}$$

is a specific application of conditional mutual information to the analysis of time series, where  $T$ ,  $\tau_x$  and  $\tau_y$  indicate the number of lags. It can be interpreted as an extension of Granger causality to a non-linear relationship.

### Bottlenecks

In general, it is not easy to evaluate mutual information from data with a finite sample size. Especially in biology, sample sizes tend to be relatively small but with high dimensions. This results in bottlenecks and makes the evaluation of mutual information difficult.

Mutual information is defined by using a distribution function. Thus, the evaluation of mutual information requires the distribution function to be estimated directly or indirectly. However, estimating a distribution function from data is not easy in practice, especially in high dimensions (Hastie et al. 2009). A simple method for estimating a distribution function is by using a normalized histogram; however, the value of the resulting mutual information will vary depending on the bin size of the histogram, and selecting bin size is an unexpectedly troublesome task. The variance in the estimated distribution function increases as the bin size decreases, and its bias increases as the bin size increases. A number of methods to select bin size have been proposed (Freedman and Diaconis 1981; Scott 1979; Shimazaki and Shinomoto 2007; Sturges 1926). However, it is still not easy to control the bias–variance trade-off when selecting the bin size for practical data (Hastie et al. 2009). For high-dimensional data, the exponential increase in the number of bins makes it especially difficult to process the histogram.

An improved method for constructing histograms using B-spline functions has been proposed by Daub et al. (Daub et al. 2004). This method is more efficient than using an ordinary histogram for estimating a distribution function by the extension of bins to polynomial functions with the use of characteristics of B-spline functions, requiring the selection of the bin size and the order of the B-spline function. The kernel density estimation method (Parzen 1962) is frequently used to estimate a distribution function; this requires the selection of the bandwidth of the kernel function rather than the bin size. Various methods to select the bandwidth have been proposed

(Turlach 1993), but there remains the problem of the bias–variance trade-off for the selection of bandwidth.

Estimating a distribution function with high accuracy and precision generally requires a sample size difficult to achieve in biology experiments. The sample size to be required to estimate distribution function increases exponentially as the number of dimensions increases, so it is especially difficult to reliably estimate a distribution function in high dimensions. The computational burden for evaluating information also increases exponentially because the summations in Eqs. (1)–(3) run over all the domains of the variables. This makes the straightforward computation of information in high dimensions an intractable problem. In biology, the number of molecular species is of the order  $10^3$  to  $10^5$ , depending on the omics layer, and so the number of dimensions of biological data, especially omics data, is often large. However, reformulating the information equations in terms of the expectation of the logarithm of a distribution function or the ratio between distribution functions can be useful. When the distribution functions are known, the computation of the expectation for all domains can be approximated by the sample mean. This approximation by sample mean only needs the values of probability density at the sampling points; thus, estimating distribution function is not needed. This reduces the computational burden for the expectation to the order of the sample size, and in biology the sample size is relatively small compared with the number of dimensions. For example, in Eq. (2), the following holds:

$$E \left[ \log \frac{p(y|x)}{p(y)} \right] \approx \frac{1}{n} \sum_i \log \frac{p(y_i|x_i)}{p(y_i)}$$

The Kozachenko–Leonenko estimator (Kozachenko and Leonenko 1987; Kraskov et al. 2004) can be used to compute the quantity of information based on approximations by using the sample mean and applying the  $k$ -nearest neighbors method, even for high-dimensional data. However, the accuracy of the Kozachenko–Leonenko estimator seems to be low, especially for high dimensions, and the information such as entropy and mutual information often take negative values. The parameter  $k$  for the number of nearest neighbors needs to be selected, but as yet there is no theoretical criterion on how to do this. In addition, there is the problem of the bias–variance trade-off for the selection of parameter  $k$ .

Thus, sample sizes and computational burden could be bottlenecks to evaluate the quantity of information.

## The application of information theory to systems biology

Information theory has contributed to systems biology in two main ways: the analysis of information transmission in cells

and the inference of the network structure of molecular species. In the former, information theory has been used to quantify information transmission (previous studies are summarized in Table 1). In the latter, information theory has been used to examine the presence or absence of statistical relationships.

In recent studies of information transmission, information theory has been used to examine signal transduction, considered to be a core mechanism in cellular information processing (Tkačik et al. 2008a, b; Cheong et al. 2011; Lestas et al. 2010; Levchenko and Nemenman 2014; Selimkhanov et al. 2014; Tostevin and ten Wolde 2009; Uda et al. 2013; Waltermann and Klipp 2011; Yu et al. 2008). Tkačik et al. examined the channel capacity and mutual information between an upstream transcription factor, Bicoid, and a downstream target gene product, Hunchback, during early embryogenesis in *Drosophila* flies (Tkačik et al. 2008b). As discussed earlier, mutual information is generally calculated from the input distribution and the conditional distribution. Unlike most biological experiments, in which the measurement of the input distribution in vivo is typically difficult, it is possible to measure the in vivo distribution of Bicoid concentrations (Gregor et al. 2007). Tkačik et al. reported that the mutual information was almost 1.5 bits, close to the channel capacity of almost 1.7 bits. These findings indicated that in vivo Bicoid/Hunchback system uses a distribution of the input, Bicoid, that results in the channel capacity. This is interesting, with the close agreement in the values of mutual information and the channel capacity possibly implying a design principle in cellular information processing.

Cheong et al. examined the channel capacity between upstream tumor necrosis factor and the downstream nuclear factor(NF)  $\kappa$ B or activating transcription factor–2 (ATF-2) and found that information transmission increased through a combination of the effects of NF $\kappa$ B and ATF-2 (Cheong et al. 2011). In addition, they defined two models, the bush model and the tree model, which differ in terms of the network structure for information transmission. In the bush model, information is transmitted downstream by branched pathways directly. In the tree model, information is transmitted via the pathways, which go through a common downstream molecule. They investigated the characteristics of information transmission by comparing two models under the assumption of a Gaussian distribution.

Uda et al. examined channel capacities between growth factors and either signaling molecules or immediate early genes (Uda et al. 2013). They demonstrated that each channel capacity is almost 1 bit for each growth factor, but each growth factor uses a specific pathway to transmit information through the use of multivariate mutual information. In addition, the information transmission was generally more robust than the average signal intensity, despite pharmacological perturbations, and compensation for information transmission

**Table 1** Summary of previous studies on information transmission in biological systems (this table is modified from Uda and Kuroda (2016))

Authors	Measurement technique	Sender	Receiver	Biological System	Main result
Tkačik et al.	Snapshot	Bicoid	Hunchback	Transcription factor, gene expression	Comparing information transmission in vivo to channel capacity
Cheong et al.	Snapshot	TNF	NFκB, ATF-2	Nuclear translocation, protein phosphorylation	Information transmission by multiple molecular species
Uda et al.	Snapshot	Growth factors, ERK, CREB	ERK, CREB, c-FOS, EGR1	Protein phosphorylation, gene production	Robustness and compensation of information transmission
Selimkhanov et al.	Live imaging	EGF ATP LPS	ERK Ca <sup>2+</sup> NFκB	Protein phosphorylation, small molecule, nuclear translocation	Information transmission by temporal pattern
Keshelava et al.	Live imaging	Acetylcholine	Ca <sup>2+</sup>	G protein-coupled receptor signaling	Information transmission at a single cell level

occurred. The mechanism or mathematical conditions underlying this robustness and compensation have not been fully elucidated; further study is needed. Robust information transmission has also been identified in the spines of neurons by numerical stochastic simulation; despite the signal being noisy, the information transmission in spines with small volumes is efficient compared with that for those with large volumes, and is sensitive to the input (Fujii et al. 2017; Tottori et al. 2019a, 2019b).

Tkačik et al., Cheong et al., and Uda et al. acquired data sets for estimating the distributions in their studies by snapshots taken at a series of time points. With this method, the combined effects of multiple time points are omitted from the calculation of mutual information, which results in the information transmission being underestimated. Generally, calculating mutual information and the channel capacity between a stimulus and the associated time trajectory, which consists of multiple time points, increases the computational difficulty because of the high-dimensional summation involved. Tostevin and ten Wolde theoretically analyzed the mutual information between time trajectories generated by biochemical reactions based on a Gaussian approximation (Tostevin and ten Wolde 2009).

Selimkhanov et al. measured the continuous time course of molecular concentrations by live imaging and efficiently calculated the mutual information and channel capacity between the stimulus and the associated time trajectory of downstream molecules, including the combined effect at multiple time points, by applying the *k*-nearest neighbors method (Selimkhanov et al. 2014). The combined effect can increase the channel capacity compared with using only a snapshot at a single time point, as can combining molecular species.

A method using machine learning for evaluating mutual information has been proposed (Cepeda-Humerez et al. 2019). In this, the time trajectories and inputs, which represent the stimuli and experimental conditions, are regarded as the explanatory and response variables, respectively. After

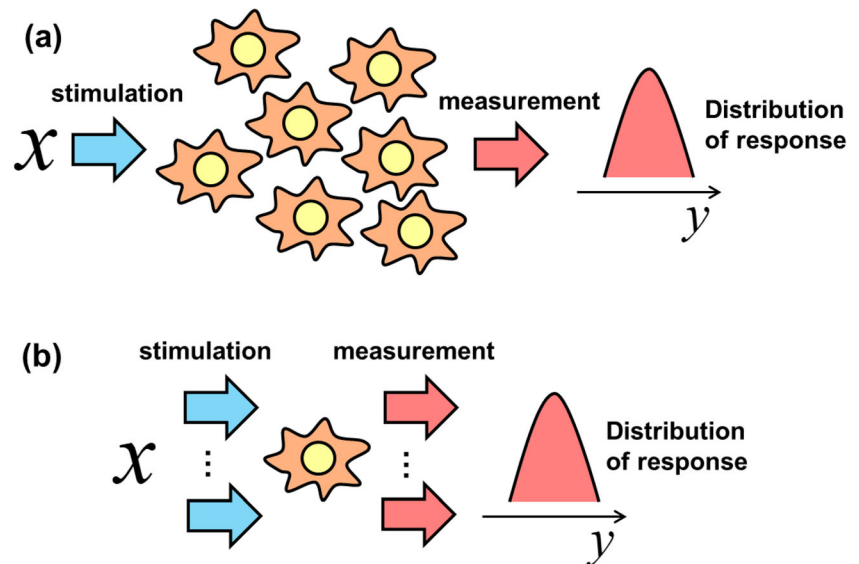
training the model for classification or regression with a training data set, mutual information was estimated from the prediction error by using a test data set. This method is based on the intuitive idea that prediction error is reduced by increasing the amount of information the time trajectories include about the input corresponding to the stimulation.

The data to estimate information transmission is usually acquired by stimulating a single cell once. Because this estimate requires a large sample size for the data set, as discussed earlier, many single cells need to be measured. When this approach is used, as in the studies described above, the author refers here to the resulting information as “information at the population level” (Fig. 3a). Conversely, a single cell can be stimulated and measured repeatedly to obtain the data. This requires the characteristics of the single cell system to remain unchanged with the repeated stimulation and measurements. The author refers to the information collected in this way as “information at the single cell level” (Fig. 3b). Information at the population level assumes that the cell systems do not differ between the cells or that the response at the population level does not vary; this can be interpreted as the receiver not discriminating between the signals in specific single cells. Conversely, information at the single cell level assumes that the cell system can vary between cells. The interpretation of the information depends on the problem settings involved in encode and decode systems.

Keshelava et al. (Keshelava et al. 2018) repeatedly stimulated and measured the responses of single cells by using a live imaging technique, and evaluated the channel capacity at the single cell level. This varied between single cells, suggesting that the system of information transmission differs between individual cells. The average channel capacity at the single cell level (almost 2 bits) was larger than the channel capacity at the population level.

Information theory has also been employed to infer the network structure of molecular species. In this case, information is used to examine statistical relationships between

**Fig. 3** Schematic interpretation of information at the population level and information at the single cell level. **a** Information at the population level is evaluated from a distribution obtained from the responses of a population of cells where each cell is stimulated once. **b** Information at the single cell level is evaluated from a distribution obtained from the responses of a single cell stimulated repeatedly



molecular species. Several methods to infer the structures of biological networks have been reported. Here, the author focuses on those based on an information theory approach. The ARACNE (Margolin et al. 2006) and CLR (Faith et al. 2007) network inference algorithms determine the presence of an edge between nodes by calculating the mutual information of the two nodes. ARACNE employs data processing inequalities to eliminate the weakest associations in every closed triplet of nodes. Intuitively, suppose a relay of information transmission, data processing inequality means that the quantity of information of end point cannot increase more than that of relay point. This procedure is exact when the network has a tree structure. In contrast, CLR compares the values of mutual information for a particular pair of nodes to the background distribution, which is empirically estimated from the two sets of values of mutual information: the set of values of mutual information between one of the pairs and all nodes, and the set of values of mutual information between the other of the pairs and all nodes. CLR is based on the assumption that the empirical distribution provides background information about the absence of edges. On the other hand, a statistical hypothesis test also provides a threshold of mutual information to determine the absence of edges. Even when  $X$  and  $Y$  are independent,  $I(X; Y) = 0$  does not always hold because of the sampling error resulting from the finite sample size. The permutation test is effective for examining the statistical significance of the null hypothesis  $I(X; Y) = 0$  and the alternative hypothesis  $I(X; Y) \neq 0$  (Daub et al. 2004).

The non-negative decomposition of multivariate mutual information has been applied to the inference of network structure from single-cell transcriptome data (Chan et al. 2017). Transfer entropy has been used to infer the connectivity of a neuronal network from time series data of neuronal activity (Vicente et al. 2011; Terada et al. 2019). Entropy has also

been used to characterize a population of differentiated cells (Grun et al. 2016).

## Summary and perspectives

Information theory is a powerful tool for quantifying information transmission in cells and inferring the network structure of molecular species and the connectivity of neuronal networks. The cost of data acquisition in biology can be high; nevertheless, the evaluation of quantity of information requires a large sample size because this is defined by distribution functions. In addition, when the data set is high dimensional, such as with omics data and time series, the computational burden increases exponentially with the increase in dimensions. When the sample size is small and the data are high dimensional, this can result in a bottleneck to applying information quantification; however, many computational methods are being developed to avoid such a bottleneck. It is important to choose suitable methods to evaluate the quantity of information on problem setting such as biological situations and experimental conditions. Bottlenecks could potentially be avoided by the future development of suitable experimental measurement techniques and computational methods, allowing information theory to be applied more widely to systems biology.

The interpretation of information transmission quantified by information quantification methods is not yet fully established, possibly because of its short research history. Information transmission refers essentially to the potential amount of information that can be transmitted. For example, if the information transmission of a pathway within a cell is 2 bits, this means that four states can be controlled by the pathway; however, the cell needs to control only two states

of survival or differentiation. The value of mutual information does not always correspond to the amount of biologically meaningful information. The elucidation of information transmission linked with a decoding system is needed to clarify the biological meaning of information.

The development of computational methods for information transmission on a time trajectory would allow the examination of how to encode information to a time trajectory. For example, epidermal growth factor stimulation has been reported to induce the transient phosphorylation of extracellular signal-regulated kinase (pERK) and cell proliferation, whereas nerve growth factor stimulation induces the sustained production of pERK and cell differentiation (Marshall 1995; Gotoh et al. 1990; Qiu and Green 1992; Traverse et al. 1992). This means that information for the distinct growth factors is encoded into the specific time trajectory of pERK, which is selectively decoded by the downstream pathways, resulting in the appropriate cell fate decisions (Sasagawa et al. 2005). This suggests that the time trajectories for the transient and sustained conditions enhance the information transmission. Extracting information about the enhancing part or pattern of such time trajectories could help elucidate the encoding mechanism underlying information transmission.

A drawback of information theory in the inference of network structures of molecular species is the difficulty of inferring the direction of edges because information quantification is symmetric for  $X$  and  $Y$ . One method for inferring the direction of edges is to use the transfer entropy. Although it remains difficult to infer the structure of a large network, such as in an omics data set, the structure of a partial network can be inferred from a time series data set of multiple molecular species.

Currently, the main contribution of studies of information transmission is limited to cellular signal transduction. However, the elucidation of the entire spectrum of life phenomena across multi-omic layers, that is, by using transomics analysis, is attracting research attention (Yugi et al. 2016; Yugi and Kuroda 2018). However, the transomics analysis of information transmission is difficult under the present circumstances because of the difficulty of data acquisition, the high-dimensional nature of the data, and the multiple time-scales involved. Nevertheless, from the long-term perspective, elucidation of how information is transmitted across multi-omic layers would be highly interesting. The author expects that the future development of measurement technology and analysis methods based on information theory could address this problem.

**Acknowledgments** The author thanks our laboratory members for helpful discussions.

**Funding information** This work was supported by the Creating information utilization platform by integrating mathematical and information sciences, and development to society, CREST (JPMJCR1912) from the Japan Science and Technology Agency (JST) and by the Japan Society

for the Promotion of Science (JSPS) (KAKENHI Grant Numbers JP18H04801, JP18H02431) and Kayamori Foundation of Informational Science Advancement.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Azeloglu EU, Iyengar R (2015) Signaling networks: information flow, computation, and decision making. *Cold Spring Harb Perspect Biol* 7(4):a005934
- Cepeda-Humerez SA, Ruess J, Tkačik G (2019) Estimating information in time-varying signals. *Plos Comput Biol* 15(9):e1007290
- Chan TE, Stumpf MPH, Babbie AC (2017) Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst* 5:251–4
- Cheong R, Rhee A, Wang CJ, Nemenman I, Levchenko A (2011) Information transduction capacity of noisy biochemical signaling networks. *Science* 334:354–358. <https://doi.org/10.1126/science.1204553>
- Cover TM, Thomas JA (2006) *Elements of information theory*, 2nd edn. Wiley, New York
- Daub CO, Steuer R, Selbig J, Kloska S (2004) Estimating mutual information using B-spline functions - an improved similarity measure for analysing gene expression data. *Bmc Bioinformatics* 5:118. <https://doi.org/10.1186/1471-2105-5-118>
- Faith JJ et al (2007) Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol* 5:54–66
- Freedman D, Diaconis P (1981) On the histogram as a density estimator - L2 theory. *Z Wahrscheinlichkeit* 57:453–476. <https://doi.org/10.1007/Bf01025868>
- Fujii M, Ohashi K, Karasawa Y, Hikichi M, Kuroda S (2017) Small-volume effect enables robust, sensitive, and efficient information transfer in the spine. *Biophys J* 112:813–826. <https://doi.org/10.1016/j.bpj.2016.12.043>
- Gotoh Y, Nishida E, Yamashita T, Hoshi M, Kawakami M, Sakai H (1990) Microtubule-associated-protein (MAP) kinase activated by nerve growth factor and epidermal growth factor in PC12 cells. Identity with the mitogen-activated MAP kinase of fibroblastic cells. *Eur J Biochem* 193:661–669
- Gregor T, Tank DW, Wieschaus EF, Bialek W (2007) Probing the limits to positional information. *Cell* 130:153–164. <https://doi.org/10.1016/j.cell.2007.05.025>
- Grun D et al (2016) De novo prediction of stem cell identity using single-cell transcriptome data. *Cell Stem Cell* 19:266–277

- Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning, 2nd edn. Springer-Verlag, New York
- Hlavackova-Schindler K, Palus M, Vejmelka M, Bhattacharya J (2007) Causality detection based on information-theoretic approaches in time series analysis. *Phys Rep* 441:1–46
- Keshelava A, Solis GP, Hersch M, Koval A, Kryuchkov M, Bergmann S, Katanaev VL (2018) High capacity in G protein-coupled receptor signaling. *Nat Commun* 9:876. <https://doi.org/10.1038/s41467-018-02868-y>
- Kozachenko LF, Leonenko NN (1987) Sample estimate of the entropy of a random vector. *Probl Inf Transm* 23:95–101
- Kraskov A, Stogbauer H, Grassberger P (2004) Estimating mutual information. *Phys Rev E* 69:066138
- Lestas I, Vinnicombe G, Paulsson J (2010) Fundamental limits on the suppression of molecular fluctuations. *Nature* 467:174–178. <https://doi.org/10.1038/nature09333>
- Levchenko A, Nemenman I (2014) Cellular noise and information transmission. *Curr Opin Biotechnol* 28:156–164. <https://doi.org/10.1016/j.copbio.2014.05.002>
- Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7:S7. <https://doi.org/10.1186/1471-2105-7-S1-S7>
- Marshall CJ (1995) Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell* 80:179–185
- Munakata T, Kamiyabu M (2006) Stochastic resonance in the FitzHugh-Nagumo model from a dynamic mutual information point of view. *Eur Phys J B* 53:239–243
- Ozaki Y, Uda S, Saito TH, Chung J, Kubota H, Kuroda S (2010) A quantitative image cytometry technique for time series or population analyses of signaling networks. *PLoS One* 5:e9955. <https://doi.org/10.1371/journal.pone.0009955>
- Palus M, Komarek V, Hmrcir Z, Sterbova K (2001) Synchronization as adjustment of information rates: detection from bivariate time series. *Phys Rev E* 63:046211
- Parzen E (1962) Estimation of a probability density-function and mode. *Ann Math Stat* 33:1065–1076
- Qiu MS, Green SH (1992) PC12 cell neuronal differentiation is associated with prolonged p21ras activity and consequent prolonged ERK activity. *Neuron* 9:705–717
- Rieke F, Warland D, Ruyter van Steveninck R, Bialek W (1997) Spikes: exploring the neural code. MIT Press, Cambridge
- Sasagawa S, Ozaki Y, Fujita K, Kuroda S (2005) Prediction and validation of the distinct dynamics of transient and sustained ERK activation. *Nat Cell Biol* 7:365–373. <https://doi.org/10.1038/ncb1233>
- Schreiber T (2000) Measuring information transfer. *Phys Rev Lett* 85:461–464
- Scott DW (1979) Optimal and data-based histograms. *Biometrika* 66:605–610. <https://doi.org/10.1093/biomet/66.3.605>
- Selimkhanov J et al (2014) Systems biology. Accurate information transmission through dynamic biochemical signaling networks. *Science* 346:1370–1373. <https://doi.org/10.1126/science.1254933>
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423
- Shimazaki H, Shinomoto S (2007) A method for selecting the bin size of a time histogram. *Neural Comput* 19:1503–1527. <https://doi.org/10.1162/neco.2007.19.6.1503>
- Sturges HA (1926) The choice of a class interval case I computations involving a single. *J Am Stat Assoc* 21:65–66
- Terada Y, Obuchi T, Isomura T, Kabashima Y (2019) Objective and efficient inference for couplings in neuronal network. *J Stat Mech Theory Exp* 12:124010
- Timme NM, Lapish C (2018) A tutorial for information theory in neuroscience. *eNeuro*:5. <https://doi.org/10.1523/ENEURO.0052-18.2018>
- Tkačik G, Bialek W (2016) Information processing in living systems. *Annu Rev Condens Ma P* 7:89–117
- Tkačik G, Callan CG Jr, Bialek W (2008a) Information capacity of genetic regulatory elements. *Phys Rev E Stat Nonlin Soft Matter Phys* 78:011910
- Tkačik G, Callan CG Jr, Bialek W (2008b) Information flow and optimization in transcriptional regulation. *Proc Natl Acad Sci U S A* 105:12265–12270. <https://doi.org/10.1073/pnas.0806077105>
- Tostevin F, ten Wolde PR (2009) Mutual information between input and output trajectories of biochemical networks. *Phys Rev Lett* 102:218101
- Tottori T, Fujii M, Kuroda S (2019a) NMDAR-mediated Ca(2+) increase shows robust information transfer in dendritic spines. *Biophys J* 116:1748–1758. <https://doi.org/10.1016/j.bpj.2019.03.030>
- Tottori T, Fujii M, Kuroda S (2019b) Robustness against additional noise in cellular information transmission. *Phys Rev E* 100:042403. <https://doi.org/10.1103/PhysRevE.100.042403>
- Traverse S, Gomez N, Paterson H, Marshall C, Cohen P (1992) Sustained activation of the mitogen-activated protein (MAP) kinase cascade may be required for differentiation of PC12 cells. Comparison of the effects of nerve growth factor and epidermal growth factor. *Biochem J* 288(Pt 2):351–355
- Turlach BA (1993) Bandwidth selection in kernel density estimation: a review. Discussion Paper 9307, Humboldt-Univ, Berlin
- Uda S, Kuroda S (2016) Analysis of cellular signal transduction from an information theoretic approach. *Semin Cell Dev Biol* 51:24–31
- Uda S et al (2013) Robustness and compensation of information transmission of signaling pathways. *Science* 341:558–561. <https://doi.org/10.1126/science.1234511>
- Vicente R, Wibral M, Lindner M, Pipa G (2011) Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *J Comput Neurosci* 30:45–67
- Waltermann C, Klipp E (2011) Information theory based approaches to cellular signaling. *Biochim Biophys Acta* 1810:924–932. <https://doi.org/10.1016/j.bbagen.2011.07.009>
- Yu RC et al (2008) Negative feedback that improves information transmission in yeast signalling. *Nature* 456:755–761. <https://doi.org/10.1038/nature07513>
- Yugi K, Kuroda S (2018) Metabolism as a signal generator across trans-omic networks at distinct time scales. *Curr Opin in Syst Biol* 8:59–66. <https://doi.org/10.1016/j.coisb.2017.12.002>
- Yugi K, Kubota H, Hatano A, Kuroda S (2016) Trans-omics: how to reconstruct biochemical networks across multiple 'omic' layers. *Trends Biotechnol* 34:276–290

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.