Title: Analysis of cellular signal transduction from an information theoretic approach

Authors: Shinsuke Uda^{a*} and Shinya Kuroda^{b,c,d,*}

Affiliations

^a Division of Integrated Omics, Research Center for Transomics Medicine, Medical Institute of Bioregulation, Kyushu University, Higashi-ku, Fukuoka, 812-8582, Japan.

^b Department of Biological Sciences, Graduate School of Science, University of Tokyo,

^c Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, University of Tokyo, ^d CREST, Japan Science and Technology Corporation, Bunkyo-ku, Tokyo 113-0033, Japan.

*Correspondence to: uda@bioreg.kyushu-u.ac.jp or skuroda@bs.s.u-tokyo.ac.jp

Abstract

Signal transduction processes the information of various cellular functions, including cell proliferation, differentiation, and death. The information for controlling cell fate is transmitted by concentrations of cellular signaling molecules. However, how much information is transmitted in signaling pathways has thus far not been investigated. Shannon's information theory paves the way to quantitatively analyze information transmission in signaling pathways. The theory has recently been applied to signal transduction, and mutual information of signal transduction has been determined to be a measure of information transmission. We review this work and provide an overview of how signal transduction transmits informational input and exerts biological output.

Keywords

Signal transduction; Information theory; Mutual information; Cellular information transmission

1. Introduction

Information processing is necessary for survival of living systems as well as the energy consumption, and it plays an important role in their adaptation to environmental conditions. Generally, information processing can be seen at multiple levels in living systems, from the single-cell level to multicellular interactions, including brain activity. In this review, we focus on the information processing of signal transduction at the single-cell level that controls cell fates such as proliferation, differentiation, and death.

Information processing at a single-cell level remains poorly understood. However, signal transduction is considered to be a core mechanism of cellular information processing, that is thought to garner information from concentrations of molecules. For example, epidermal growth factor (EGF) stimulation has been reported to induce transient phosphorylation of extracellular signal-regulated kinase (pERK) and cell proliferation, while nerve growth factor (NGF) stimulation induces sustained pERK and cell differentiation [1-4]. This means that information for distinct growth factors is encoded into distinct patterns of pERK concentration that are selectively decoded by downstream molecules and lead to proper cell fate decisions [5] (Fig. 1). For example, ERK activation can transmit information about various types of cell fates. Thus, signaling pathway can be regarded as a communication system that involves encoding, communicating, and decoding variable information.

Information transmission of a cellular system is mediated by concentrations of

molecules serving as signals, with chemical interactions between upstream molecules (senders) and downstream molecules (receivers) underlying the transmission. Thus, the concentrations of upstream molecules needs to be distinguished from concentrations of downstream molecules. More precisely, the difference in concentrations of upstream molecules rather than their absolute quantity needs to be known for information transmission to occur. In other words, the amount of information transmitted between upstream and downstream molecules is measured by how much upstream information is known downstream based on interactions between chemicals. Mutual information—that is, information from senders that receivers can interpret—is defined based on this idea, and is widely used in various fields such as communication theory and information science [6-8]. Mutual information has recently been began to be measured in signal transduction (Table 1).

The first example of mutual information being applied in a cellular system is the study of information processing in morphogenesis in *Drosophila*, in which mutual information between concentrations of upstream transcription factors and downstream gene expression based on snapshot population data at separate time points has been measured [9] [10]. Similarly, mutual information of tumor necrosis factor (TNF) dependent signal transduction [11] and ERK-dependent signal transduction [12] has been determined by snapshot population data in mammalian cells. Recently, mutual information of ERK, Calcium (Ca^{2+}), and nuclear factor kappa-light-chain-enhancer of activated B cells (NFkB) using time series data in a live cell–imaging technique has also been measured [13] (see section 5).

In this review, we assume wide readership from biologists to information scientists. The rest of this paper is organized as follows. In section 2 and 3, we explain the concept

of mutual information for signal transduction and how to calculate mutual information respectively, for biologists who are not familiar with information theory. Thus, others can skip these section. In section 4, we describe data acquisition. This section may be informative for information scientists rather than biologists. In section 5, we introduce previous works. In section 6, we describe summary and future perspective.

2. Mutual Information for Signal Transduction

Mutual information is used as a general measure of information between random variables in an array of fields, including communication theory, statistical science, and machine learning [6-8, 14]. Thus, the interpretation of mutual information depends on the problem setting in each field. For example, mutual information in statistical science is interpreted as a measure of independency between random variables—in other words, a general correlation including nonlinear relationships. In this section, we introduce mutual information as a measure of the amount of information transmission, which requires only the statistical property of signals.

Let x and y be concentrations of an upstream molecule (a sender) and a downstream molecule (a receiver) in a signal transduction pathway, respectively. If we assume that the concentrations of molecules x and y behave deterministically, the concentrations of molecule y is uniquely determined by $f(x)$, that is $y = f(x)$, whose function is given by the chemical reactions between the upstream and downstream molecules. However, in practice, the concentrations of x and y behave stochastically due to factors such as reaction noise and variable expression levels of molecules in individual cells, and they are regarded as random variables X and Y , respectively. Thus, the

relationship between the each concentration of molecules x and y is denoted by the conditional distribution $p(y|x)$, which is the distribution of molecule concentration y given the concentrations of molecule x (Fig. 2). This means that, the concentration of downstream molecule y is distributed depending on the concentration of downstream molecule x (Fig. 2, gray area). Therefore, the distribution of downstream molecule concentration $p(y)$ is determined by the distribution of upstream molecule concentration $p(x)$ and the conditional distribution $p(y|x)$. This is denoted by marginalization $p(y) = \int dx p(y|x) p(x)$. For example, in the ERK-dependent signal transduction pathway (Fig. 1), the concentrations of molecules x and y can be regarded as NGF and pERK concentrations, respectively. Then, $p(x)$ is the distribution of NGF concentrations, and $p(y)$ is the distribution of pERK concentrations at the cell population level. $p(y|x)$ is the conditional distribution of pERK concentration given the specific NGF concentration at the cell population level. Thus, the mutual information is defined by

$$I(X;Y) = \int dx p(x) \int dy p(y|x) \log \frac{p(y|x)}{p(y)} \quad (1)$$

Given that x and y are NGF and pERK concentrations, respectively, the mutual information is the amount of information reliably transmitted between NGF and pERK.

If the distribution $p(x)$ and the conditional distribution $p(y|x)$ are given, equation (1) can be solved because the distribution $p(y)$ is uniquely determined from them.

As already described, mutual information is a measure of the amount of information transmission. However, what is the amount of information at all? We intuitively think that the more words that a book contains, the more information it has; the description length corresponds to the amount of information. All information can be described by a binary number as well because computers process all information, including texts, pictures, and sounds, by binary sequences. In signal transduction, the amount of information may correspond to controllable numbers of cell fates or biological functions. The description length of a binary sequence is a logarithm of base 2, and the unit is ‘bits’. For simplicity, we assume that random variables are discrete. The requisite description length is $\log_2 1/p(x)$ to describe the information of event x from distribution $p(x)$, so that the average of description length is the minimum [8]. The more frequently an event happens, the shorter the description length of the event, making the average description length a minimum. In other words, frequent events contain less information than rare events. The average description length of an event is termed Shannon’s entropy or information.

$$H(X) = -\sum_i p(x_i) \log p(x_i) \quad (2)$$

Shannon’s entropy is also interpreted as the amount of uncertainty attached to the random variable. In the case of a continuous variable, the summation is replaced with the integral in equation (2). Since Shannon’s entropy is the average length of a description of events, it can be regarded as the order for the average number of states necessary to describe a random variable. For example, a random variable, which is 1 bit according to Shannon’s entropy, contains the information corresponding to $2^1 = 2$ states.

Conditional entropy is defined by

$$H(Y|X) = -\sum_{i,j} p(x_i) p(y_j|x_i) \log p(y_j|x_i) \quad (3).$$

The mutual information of equation (1) is denoted by subtracting the conditional entropy from Shannon's entropy

$$I(X;Y) = H(Y) - H(Y|X) \quad (4).$$

Equation (4) can be represented graphically with a Venn diagram (Fig. 3). The mutual information is non-negative and symmetrical. Thus, the mutual information takes the upper bound $\min(H(X), H(Y))$.

The mutual information is determined by the distribution $p(x)$ of a sender and the conditional distribution $p(y|x)$ that is the communication channel. With a short time scale, the communication channel, which is given by the signaling pathway between the upstream and downstream molecules, can be considered to be almost fixed. On the other hand, the distribution $p(x)$ of a sender can be optimized to maximize mutual information under the fixed communication channel. The maximum of mutual information optimized against the distribution $p(x)$ is termed 'channel capacity'.

From the point of view of the design principle of a cell, the comparison of channel capacity with mutual information using the distribution $p(x)$ of molecule concentrations *in vivo* has attracted researchers' interest. Tkačik *et al.* has reported that the mutual information calculated by *in vivo* distribution of upstream molecule concentration, Bicoid, is almost equivalent to the channel capacity [9] (see section 5).

Two to the power of mutual information is the average number of states of the sender, which the receiver can distinguish. We think a simple case transmitting 2 signal from

upstream molecule to downstream molecule for the purpose of illustration. For example, each signal implies survive or differentiation. In this case, we assume that two different distributions of upstream molecule concentration, if they are sufficiently away not from each other and does not overlap, can be regarded as two distinct signals (Fig. 4 A and B, yellow or red area along x axis). Thus, the entropy of upstream signal is approximately 1 bit, and the upper bound of mutual information between upstream molecule and downstream molecule is 1bit. If each distribution of downstream molecule concentration corresponds to each distribution of upstream molecule concentration, and the distributions of downstream molecule concentration do not overlap each other, the mutual information between upstream molecule and downstream molecule is 1bit (Fig. 4A). This means that one can distinguish 2 signals by observing concentration of downstream molecule, and the information of signal of upstream molecule is perfectly transmitted between upstream molecule and downstream molecule in the sense of information theory. By contrast, if the distribution of downstream molecule concentration overlaps each other, the mutual information between upstream molecule and downstream molecule becomes lower than 1bit, and decreases by increasing overlapping area of the distributions of downstream molecule concentration, because one cannot always distinguish each signals by observing concentration of downstream molecule (Fig. 4B). The how the distributions of downstream molecule concentration overlap each other depends on both $p(x)$ and $p(y|x)$. For accurate information transmission, the smaller width and farther distance of distribution $p(x)$ is favorable, and the smaller width of $p(y|x)$ against the range of $p(x)$ is favorable so that the each distribution of downstream molecule concentration becomes farther and

distinguishable. In practice, the number of signals becomes more than 2, the conditions for accurate information transmission becomes more complicated.

Roughly speaking, 2 to the power of the mutual information is the average number of states of the receiver that the sender can control in the context of signal transduction (Fig. 4). However, the sender does not always fully use all controllable states for information transmission of biological meaning. For example, the information of cell differentiation consists of 2 states; that is, whether a cell differentiates or not. This information can be transmitted by the pathway, which has mutual information of more than 1 bit. If a cell uses the pathway of 2 bits of mutual information, the cell redundantly assigns the 2 states of differentiation to 4 states. Thus, the large mutual information does not always translate to a large transmission of information in a biological context. Conversely, if a number of biological meanings clearly exist, the mutual information of pathways related to them needs to be sufficiently large to transmit such information. For example, if a cell tries to control 4 distinguishable cell fates, such as survival, proliferation, differentiation, and death, the mutual information of the pathways related to the cell fates needs to be more than 2 bits to control the cell fates with high average accuracy. Given that the cell fates are controlled with high accuracy on average despite the mutual information being lower than 2 bits, the information for cell fate may be transmitted not only by a one-to-one relationship of concentrations of molecules but also by a network of relationships between concentrations of molecules, including molecular species or multiple time points. In addition, unknown pathways may exist. Thus, a sufficiently large amount of mutual information is a necessary condition for transmitting information in a biological context.

The use of a network of relationships of concentrations of molecules (i.e., the use of a

combination of molecular species) may be an efficient means for information transmission of signal transduction. In such a case, although mutual information can be applied to measure information transmission by redefining mutual information in equation (1) for many variables, two problems usually arise as the number of variables increases. First, the computational cost to calculate mutual information increases exponentially. Second, the sample size to calculate mutual information with high accuracy becomes larger. Thus, calculation of mutual information for a huge number of variables is difficult, but some approximation techniques may be effective. In an analysis of mutual information for many variables, elucidating how to code the information in signal transduction seems to be important.

We assumed that the concentration of a molecule is a random variable and separated the source of stochasticity into reaction noise (intrinsic noise) and variable concentrations of molecules in individual cells (extrinsic noise). Generally, estimating the specific effects of intrinsic and extrinsic noise for information transmission is not easy. Using numerical simulation in information transmission, Selimkhanov *et al.* reported that extrinsic noise has a larger effect than intrinsic noise [13]. In addition, the effect of extrinsic noise has been reported to be larger than that of intrinsic noise as the source of variability [15], which helps to understand the nature of cellular information transmission.

3. How to Calculate Mutual Information

Mutual information is calculated from the distribution $p(x)$ of a sender and the conditional distribution $p(y|x)$ of a communication channel. One can use the

empirical distribution of an upstream molecule concentration or prior knowledge as the distribution $p(x)$. For example, Tkačik *et al.* measured *in vivo* Bicoid concentrations at a cell population level and obtained the empirical distribution $p(\text{Bicoid})$ as $p(x)$ from the data set of the cell population (Table 1, see section 5). Given that the distribution $p(x)$ is unknown, but the conditional distribution $p(y|x)$ is known, the channel capacity could be calculated by maximizing the mutual information against the distribution $p(x)$. In such a case, the mutual information is usually maximized by numerical computation [16, 17]. In the field of signal transduction, the distribution $p(x)$ is sometimes subjected to conditions such as smoothness of distribution due to aspect of controllability of concentration x [9, 12]. For example, in a previous study [12], because the NGF concentrations *in vivo* were unknown, we determined the distribution of molecule concentration $p(\text{NGF})$ as $p(x)$, so that the mutual information was numerically maximized.

Obtaining the empirical conditional distribution $p(y|x)$ depends on the possibility of experimentally adjusting the concentrations of molecule x . If the concentrations can be directly altered, in the first step, cells are stimulated by specific concentrations of molecule $x = x_i$, and the concentrations of molecule y in each cell are then measured. In this step, one can obtain the empirical conditional distribution given $x = x_i$; that is, $p(y|x = x_i)$. In the next step, the first step is iterated with various concentrations of molecule $x = x_i$. In the final step, one can obtain the empirical conditional distribution $p(y|x)$ by collecting $p(y|x = x_i)$. If concentrations of molecule x cannot be

directly manipulated, but can be changed by a stimulation s , in the first step, cells are stimulated by specific concentrations of molecule $s = s_i$, and the concentrations of molecules x and y in each cell are then measured. In this step, one can obtain the empirical conditional distribution given $s = s_i$; that is, $p(x, y | s = s_i)$. In the next step, the first step is iterated with various concentrations of molecule $s = s_i$. In the final step, one can obtain the empirical conditional distribution $p(x, y | s)$ by collecting $p(x, y | s = s_i)$, and the empirical joint distribution by the marginalization

$$p(x, y) = \sum_s p(x, y | s) p(s) \approx \frac{1}{N} \sum_i^N p(x, y | s = s_i) \quad (5)$$

where s_i is sampled from the distribution $p(s)$. The definitive method for determining the distribution $p(s)$ is not known; thus, the distribution $p(s)$ is conventionally set to a uniform distribution. b-iv) The empirical conditional distribution of communication channel is calculated by

$$p(y | x) = \frac{p(x, y)}{p(x)} = \frac{p(x, y)}{\sum_y p(x, y)}. \quad (6)$$

For example, Uda *et al.* [12] simultaneously measured pERK and c-FOS in each cell with various NGF concentrations and experimentally obtained the empirical joint distribution $p(x, y)$.

A conventional and simple method for obtaining an empirical distribution from a data set is to use a normalizing histogram. However, this method usually has drawbacks. The value of the mutual information often sensitively depends on the bin width, especially when the sample size is small. In that case, the empirical distribution is apt to contain

many false modes like spikes, and it becomes overly complicated and far from the true distribution. Thus, the mutual information calculated from the empirical distribution with a small bin width is apt to be overestimated. Conversely, if the bin width is large, the empirical distribution tends to be over smoothed and over simplified. Thus, the mutual information calculated from the empirical distribution with a large bin width may be underestimated. As the sample size becomes smaller, this tendency becomes more prominent. In addition, as the dimension of the sample space increases, the number of bins exponentially increases. This creates difficult problems involving both computational costs and memory space. Estimating a distribution from a finite size data set is generally not easy. Various methods to avoid such problems have been developed by researchers in information science and include adaptive partitioning, kernel density estimation, the k-nearest neighbor method, and assuming parametric distribution [7, 18-20]. Because each method has advantages and disadvantages, a user chooses an appropriate method based on the situation, including the sample size, the dimension of the sample space, computational costs, and so forth. In addition, the computational procedure of mutual information is closely related to the estimation method of the empirical distribution from the data set.

4. Data acquisition

An adequate sample size to calculate mutual information is generally large depending on the number of variables, and the cost of measurement in biological experiments becomes considerably high.

Ozaki *et al.* introduced quantitative image cytometry (QIC), an automated quantitative immunostaining method that combines a liquid-handling robot and image analysis, and

acquired a large data set [21]. In this case, the concentrations of molecules x and y in a single cell were estimated from information such as the summation of pixels of a cell area, which was acquired from the snapshot image. For example, Uda *et al.* [12] measured pERK and c-FOS concentrations simultaneously. The mutual information was determined by the distributions $p(x)$ and $p(y|x)$; thus, the arbitrary unit of concentrations of molecules was sufficient. The advantage of QIC is that it does not require introducing any probe into cells, and the intracellular conditions remain unperturbed. However, a disadvantage of QIC is its inability to continuously measure signals at multiple time points in the same cell because cells must be fixed at a single time point. By contrast, live cell imaging permits continuous measurement of signals at multiple time points in the same cell. The concentrations of molecules x and y are measured by the intensity of a luminescent or fluorescent probe in live cell imaging, which is estimated from the pixel images of each cell. Thus, the live cell imaging can provide the data for calculating mutual information, including the effect of the combination of multiple time points. However, the live cell imaging probe signal is *not* a direct signal of the target molecules, and uncertainty remains on what the probe signal reflects. In addition, the probe may affect cell activity and intracellular conditions. Thus, QIC and live cell imaging are complementary methods.

Two types of distributions, snapshot data at cell population level or continuous time series data in a single cell, can be used to calculate mutual information. The mutual information obtained from snapshot data at a cell population level and continuous time series data in a single cell corresponds to the amount of information transmission within a cell population and a single cell, respectively.

5. Previous work on information transmission in signal transduction

Shannon's information theory has recently been applied to analyze signal transduction [9-13, 22-26]. The problem settings of information transmission are diverse, even in the same signal transduction field, and depend on biological processes, measurement methods, and the upstream and downstream molecules. This is summarized in Table 1.

Tkačik *et al.* has calculated the channel capacity between an upstream transcription factor, Bicoid, and a downstream target gene product, Hunchback, during early embryogenesis in *Drosophila* [9]. Although the mutual information is generally calculated from the distribution $p(x)$ and the conditional distribution $p(y|x)$ of concentrations of molecules x and y , the measurement of the distribution $p(x)$ is difficult *in vivo*. However, this case was exceptional because the *in vivo* distribution $p(x)$ of Bicoid concentrations can be measured [27]. The researchers evaluated the *in vivo* mutual information and the channel capacity between Bicoid and Hunchback and compared the *in vivo* mutual information to the channel capacity. The *in vivo* mutual information was almost 1.5 bits and was close to the channel capacity, which was almost 1.7 bits. These findings indicated that *in vivo* Bicoid/Hunchback systems use an optimal distribution of input, Bicoid, that gives a channel capacity. This study is visionary and one of the pioneering works on information transmission in the field of signal transduction.

Cheong *et al.* have evaluated channel capacity between upstream TNF and downstream NFκB or activating transcription factor-2 (ATF-2) [11]. They found that

the combination of NF κ B and ATF-2 enables an increase in information transmission. In addition, under the assumption of Gaussian distribution for the molecular concentrations, the network structure related to information transmission was investigated by comparing two models, the bush model and the tree model. The bush model transmits information by branched pathways directly downstream. The tree model transmits information by the pathways, which go through a common downstream molecule.

Uda *et al.* evaluated each channel capacity between growth factors and signaling molecules and immediate early genes [12]. By use of multivariate mutual information, they demonstrated that, whereas each channel capacity is almost 1 bit against each growth factor, each growth factor uses each specific pathway to transmit information. In addition, the information transmission was generally more robust than the average signal intensity despite pharmacological perturbations, and compensation for information transmission occurred. However, the mechanism of robustness and compensation has not been sufficiently elucidated yet, although work is underway.

Tkačik *et al.* [9], Cheong *et al.* [11] and Uda *et al.* [12] measured concentrations of molecules by snapshot at each time point (Table 1). In contrast, Selimkhanov *et al.* measured the continuous time course of molecule concentrations by live imaging [13]. Thus, they evaluated the channel capacity between stimulation and the time trajectory of the concentration of downstream molecules, including the effect at multiple time points. Combining multiple time points can increase the channel capacity compared to using only snapshot data at a single time point, as can combining molecular species. Generally, calculating the channel capacity between stimulation and time trajectory increases the computational difficulty because of the high dimensional integral. Selimkhanov *et al.* calculated the mutual information efficiently by the application of k-nearest neighbor

method. In addition, Tostevin *et al.* analyzed the mutual information between time trajectories based on Gaussian approximation [23].

6. Summary and future perspective

Shannon's information theory is applicable to the quantitative analysis of information transmission in signal transduction. The mutual information is the amount of information that is reliably transmitted over a communication channel, and it can be calculated from the distribution $p(x)$ of upstream molecule concentration and the conditional distribution $p(y|x)$ as the communication channel. The amount of information may correspond to the number of controllable states of cell fates or biological functions. In addition, the amount of information corresponds to description length, and the unit is often bits. In other words, the mutual information corresponds to the binary logarithm of the average number of (controllable) states.

Cells can transmit information by coding it through a combination of multiple time points and molecular species. This concept is similar to population coding in brain science [28-30], in which an object is recognized by a combination of activated neurons. Each neuron can be activated not only by looking at a specific object but also by looking at other objects. This means that a neuron can be used for coding of multiple objects. Similarly, a combination of multiple time points or molecular species has been reported to be used for controlling multiple biological functions in living systems [5, 31-36]. Thus, population coding of multiple time points and multiple molecular species in signal transduction is conceptually similar to the population coding in brain science, which may serve as a useful reference in elucidating signal transduction.

In the previously described studies [9, 11-13], channel capacity is almost 1 to 2 bits. Given the complexity of living systems, the measured channel capacity was very low, which suggests that an unknown mechanism or pathways of information transmission may exist. However, information transmission, including the combination of both molecular species and multiple time points, has to our knowledge not been investigated yet. The simultaneous measurement of several combinations of molecular species with continuous time is difficult for current technology. In addition, the computational cost for exploring the optimal combination of molecular species and multiple time points of large channel capacity is high. More advanced technology is required.

In the field of signal transduction, one of the important properties of information transmission defined by mutual information is that information is not necessarily encoded in absolute concentration of molecule. This property would be a clue to elucidate the design principle of living system; how body shape of multicellular organisms are controlled regardless the body size. For example, a range of body size is very wide even within the same species. If the information of body shape is encoded in absolute concentration of molecule, the higher absolute concentration should be needed as the body size increases. However, such instance has not been observed yet. For example, the biggest dog, Great Dane, is five times taller in height and 40 time bigger in weight than Chihuahua, the smallest dog; however, their body shapes look similar each other regardless of their body sizes. The absolute concentrations of molecule of morphogens between Great Dane and Chihuahua has not thus far been reported to be different yet to our knowledge, and are unlikely to be different. This suggests that the information of body shape is unlikely to be encoded in absolute concentration of molecule, and that information entropy rather than absolute concentration is more likely

to be used for coding body shape information. Such information coding may be also used in general biological processes, and such a “system-size free” information coding may be a universal mechanism in living systems.

Mutual information may be useful to find a drug target of diseases. For example, in cancer cells, the pathways involved in proliferation have been reported to be dis-regulated [37]. Mutual information between upstream and downstream molecule may be more effective rather than absolute concentration of downstream molecule for evaluation of dysregulation of cell proliferation pathway in cancers. This means that, even if the absolute concentration of molecule decreased by an inhibitor, while the mutual information between upstream and downstream molecule does not decrease, the dysregulated pathway is not inhibited even with a decrease of absolute concentration of molecule [12]. Teschendorff *et al* proposed to evaluate drug sensitivity by approach of entropic analysis [38]. They used entropy as a criterion of uncertainty and robustness of network, thus, their approach is some different from information transmission defined by mutual information.

Channel capacity describes one of the major quantities of information transmission when Shannon’s information theory is applied to signal transduction. It is set as a maximization problem for mutual information underlying the conditional distribution $p(y|x)$ against the distribution $p(x)$ of a sender. The reason why channel capacity was estimated is that it gives the maximal effectiveness of information transmission and reveals the design principles of signal transductions.

Acknowledgments

We thank our laboratory members for critical reading of the manuscript. This work was

supported by the Creation of Fundamental Technologies for Understanding and Control of Biosystem Dynamics, CREST, from the Japan Science and Technology (JST); by a KAKENHI Young Scientists (B) (#25830142) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT), and by a Human Frontier Science Project (HFSP) grant (RGP0061/2011).

Figure legends

Table 1. A summary of previous works. Abbreviations: TNF, tumor necrosis factor; NF κ B, nuclear factor kappa-light-chain-enhancer of activated B cells; ATF-2, activating transcription factor-2; NGF, nerve growth factor; PACAP, pituitary adenylate cyclase-activating peptide; PMA, phorbol 12-myristate 13-acetate; ERK, extracellular signal-regulated kinase; CREB, 3'-5'-cyclic adenosine monophosphate response element-binding protein; ERGR1, early growth response protein 1; EGF, epidermal growth factor; ATP, adenosine triphosphate; LPS, lipopolysaccharide; Ca²⁺, calcium. The data set [12] is available at <http://kurodalab.bi.s.u-tokyo.ac.jp/info/>.

Fig. 1. The information transmission in signal transduction. Given that x represents pERK concentrations, pERK can be regarded as a sender for downstream gene expression, and as a receiver for upstream growth factors. Whether pERK is a sender or a receiver depends on the problem setting for analysis.

Fig. 2. The relationship of distributions of upstream molecule concentration x as a sender and downstream molecule concentration y as a receiver. The solid line is the average of y given x and corresponds to the average reaction response curve. The vertical height of gray area given x is the variability of y given x . Given that, let us assume an ERK signaling network in which x and y are NGF and pERK concentrations, respectively. $p(x)$ and $p(y)$ are the distributions of NGF and pERK concentrations, respectively. $p(y|x)$ is the conditional distribution of pERK

concentrations given a NGF concentration.

Fig. 3. The graphical representation of Shannon's entropy and mutual information. Each area corresponds to Shannon's entropy for each random variable. The area of intersection of the two circles corresponds to mutual information between random variable X and Y . The rest of each circle outside the intersection corresponds to each conditional entropy.

Fig. 4. The distinguishability of the upstream molecule concentration x from the downstream molecule concentration y . (A) In case of a smaller variation of molecule concentration y , the difference in the molecule concentration x is distinguished from the molecule concentration y . (B) In the case of a larger variation of molecule concentration y , the difference in the molecule concentration x is not always distinguished from molecule concentration y due to an overlap of each conditional distribution of molecule concentration y .

Reference list

- [1] Marshall CJ. Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell* 1995;80:179-85.
- [2] Gotoh Y, Nishida E, Yamashita T, Hoshi M, Kawakami M, Sakai H. 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* 1990;193:661-9.
- [3] Qiu MS, Green SH. PC12 cell neuronal differentiation is associated with prolonged p21ras activity and consequent prolonged ERK activity. *Neuron* 1992;9:705-17.
- [4] Traverse S, Gomez N, Paterson H, Marshall C, Cohen P. 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* 1992;288 (Pt 2):351-5.
- [5] Sasagawa S, Ozaki Y, Fujita K, Kuroda S. Prediction and validation of the distinct dynamics of transient and sustained ERK activation. *Nat Cell Biol* 2005;7:365-73.
- [6] Yeung RW. Information theory and network coding. New York: Springer; 2008.
- [7] Bishop CM. Pattern Recognition and Machine Learning: Springer; 2006.
- [8] Cover TM, Thomas JA. 2nd ed. New York: Wiley; 2006.
- [9] Tkacik G, Callan CG, Jr., Bialek W. Information flow and optimization in transcriptional regulation. *Proc Natl Acad Sci U S A* 2008;105:12265-70.
- [10] Tkačik G, Callan CG, Jr., Bialek W. Information capacity of genetic regulatory elements. *Phys Rev E Stat Nonlin Soft Matter Phys* 2008;78:011910.
- [11] Cheong R, Rhee A, Wang CJ, Nemenman I, Levchenko A. Information transduction capacity of noisy biochemical signaling networks. *Science* 2011;334:354-8.
- [12] Uda S, Saito TH, Kudo T, Kokaji T, Tsuchiya T, Kubota H, et al. Robustness and compensation of information transmission of signaling pathways. *Science* 2013;341:558-61.
- [13] Selimkhanov J, Taylor B, Yao J, Pilko A, Albeck J, Hoffmann A, et al. Systems biology. Accurate information transmission through dynamic biochemical signaling networks. *Science* 2014;346:1370-3.
- [14] Murphy KP. Machine Learning: a probabilistic perspective. Cambridge: The MIT Press; 2012.
- [15] Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, Endy D, et al. Regulated cell-to-cell variation in a cell-fate decision system. *Nature* 2005;437:699-706.
- [16] Arimoto S. An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Trans Information Theory* 1972;18:14-20.
- [17] Blahut RE. Computation of channel capacity and rate-distortion functions. *IEEE Trans Information Theory* 1972;18:460-73.
- [18] Fraser AM, Swinney HL. Independent Coordinates for Strange Attractors from Mutual Information. *Physical Review A* 1986;33:1134-40.
- [19] Darbellay GA, Vajda I. Estimation of the information by an adaptive partitioning of the observation space. *IEEE Trans Information Theory* 1999;45:1315-21.
- [20] Walters-Williams J, Li Y. Estimation of Mutual Information: A Survey. *Rough Sets and Knowledge Technology, Proceedings* 2009;5589:389-96.
- [21] Ozaki Y, Uda S, Saito TH, Chung J, Kubota H, Kuroda S. A quantitative image cytometry technique for time series or population analyses of signaling networks. *PLoS One* 2010;5:e9955.
- [22] Levchenko A, Nemenman I. Cellular noise and information transmission. *Curr Opin Biotechnol* 2014;28:156-64.
- [23] Tostevin F, ten Wolde PR. Mutual information between input and output trajectories of biochemical networks. *Phys Rev Lett* 2009;102:218101.
- [24] Lestas I, Vinnicombe G, Paulsson J. Fundamental limits on the suppression of molecular fluctuations. *Nature* 2010;467:174-8.
- [25] Yu RC, Pesce CG, Colman-Lerner A, Lok L, Pincus D, Serra E, et al. Negative feedback

- that improves information transmission in yeast signalling. *Nature* 2008;456:755-61.
- [26] Waltermann C, Klipp E. Information theory based approaches to cellular signaling. *Biochimica et Biophysica Acta (BBA) - General Subjects* 2011;1810:924-32.
- [27] Gregor T, Tank DW, Wieschaus EF, Bialek W. Probing the limits to positional information. *Cell* 2007;130:153-64.
- [28] Wu S, Nakahara H, Amari S. Population coding with correlation and an unfaithful model (vol 13, pg 775, 2001). *Neural Computation* 2001;13:1921-.
- [29] Kang K, Sompolinsky H. Mutual information of population codes and distance measures in probability space. *Physical Review Letters* 2001;86:4958-61.
- [30] Averbeck BB, Latham PE, Pouget A. Neural correlations, population coding and computation. *Nat Rev Neurosci* 2006;7:358-66.
- [31] Behar M, Hoffmann A. Understanding the temporal codes of intra-cellular signals. *Curr Opin Genet Dev* 2010;20:684-93.
- [32] Nakakuki T, Birtwistle MR, Saeki Y, Yumoto N, Ide K, Nagashima T, et al. Ligand-specific c-Fos expression emerges from the spatiotemporal control of ErbB network dynamics. *Cell* 2010;141:884-96.
- [33] Janes KA, Albeck JG, Gaudet S, Sorger PK, Lauffenburger DA, Yaffe MB. A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. *Science* 2005;310:1646-53.
- [34] Chen JY, Lin JR, Cimprich KA, Meyer T. A two-dimensional ERK-AKT signaling code for an NGF-triggered cell-fate decision. *Mol Cell* 2012;45:196-209.
- [35] Akimoto Y, Yugi K, Uda S, Kudo T, Komori Y, Kubota H, et al. The extraction of simple relationships in growth factor-specific multiple-input and multiple-output systems in cell-fate decisions by backward elimination PLS regression. *PLoS One* 2013;8:e72780.
- [36] Saito TH, Uda S, Tsuchiya T, Ozaki Y, Kuroda S. Temporal decoding of MAP kinase and CREB phosphorylation by selective immediate early gene expression. *PLoS One* 2013;8:e57037.
- [37] Blume-Jensen P, Hunter T. Oncogenic kinase signalling. *Nature* 2001;411:355-65.
- [38] Teschendorff AE, Sollich P, Kuehn R. Signalling entropy: A novel network-theoretical framework for systems analysis and interpretation of functional omic data. *Methods* 2014;67:282-93.

Table 1

Authors [ref. No]	Measure ment techniqu e	Sender	Receiver	Biological System	Main result
Tkačik <i>et al.</i> [9]	snapshot	Bicoide	Hunchba ck	Transcription factor, gene expression	Comparing information transmission in vivo to channel capacity
Cheong <i>et al.</i> [11]	snapshot	TNF	NFκB, ATF-2	Nuclear translocation, Protein phosphorylati on	Information transmission by multiple molecular species
Uda <i>et al.</i> [12]	snapshot	NGF, PACAP, PMA, ERK, CREB	ERK, CREB,c- FOS, EGR1	Protein phosphorylati on, gene production	Robustness and compensation of information transmission
Selimkhanov <i>et al.</i> [13]	Live imaging	EGF	ERK	Protein phosphorylati on, small molecule, nuclear translocation	Information transmission by temporal pattern
		ATP	Ca ²⁺		
		LPS	NFκB		

Fig.1

Fig.2

Fig.3

Fig.4

A

B