scientific reports

OPEN

Epidermal growth factor receptor cascade prioritizes the maximization of signal transduction

Kaori Kiso-Farnè[1] & Tatsuaki Tsuruyama[1,2,3,4,5]✉

Many studies have been performed to quantify cell signaling. Cell signaling molecules are phosphorylated in response to extracellular stimuli, with the phosphorylation sequence forming a signal cascade. The information gain during a signal event is given by the logarithm of the phosphorylation molecule ratio. The average information gain can be regarded as the signal transduction quantity (ST), which is identical to the Kullback–Leibler divergence (KLD), a relative entropy. We previously reported that if the total ST value in a given signal cascade is maximized, the ST rate (STR) of each signaling molecule per signal duration (min) approaches a constant value. To experimentally verify this theoretical conclusion, we measured the STR of the epidermal growth factor (EGF)-related cascade in A431 skin cancer cells following stimulation with EGF using antibody microarrays against phosphorylated signal molecules. The results were consistent with those from the theoretical analysis. Thus, signaling transduction systems may adopt a strategy that prioritizes the maximization of ST. Furthermore, signal molecules with similar STRs may form a signal cascade. In conclusion, ST and STR are promising properties for quantitative analysis of signal transduction.

Signal transduction systems are unique chain reactions involving signaling molecules in biological systems. A well-known example is the epidermal growth factor (EGF)-driven signal cascade in cancer cells[1]. Many pioneering studies recently reported that information science theory is a powerful framework for quantitatively understanding signal transduction[2]. Using this framework allows researchers to identify the information transmission strategies used by cells[3–5].

We have recently been developing a method for quantifying signal transduction involving the Kullback–Leibler divergence (KLD), a relative entropy that expresses the average difference in information entropy before and after a signal event. The difference is called information gain[6–8]. The KLD has also been used in other fields, such as clinical trial research[9], Bayesian model diagnostics[10], and bioequivalence evaluations[11].

A431 skin cancer cells can be used to evaluate EGF signaling. Stimulating the EGF receptor (EGFR) of the cultured cells causes sequential phosphorylation of Raf1, mitogen-activated protein kinase (MAPK)-extracellular signal-regulated kinase 1 (MEK1), and kinase-extracellular signal-regulated kinase 1 (ERK1), which allows for the transmission of information along with other transcription factors into the nucleus, followed by cell proliferation[12–19]. This sequential phosphorylation of signaling molecules is considered to form the Raf1-MEK1-ERK1 signaling cascade[1]. In addition, the MAPK cascade is essential for cancer cell proliferation[14–17]. In the absence of EGFR stimulation, signaling molecules are constitutively phosphorylated through other types of stimulation. Therefore, the phosphorylation level at the time of signal transduction by EGF receptor stimulation should be evaluated as the difference between the post- and pre-stimulation values. The KLD is an information gain, representing the average value of the difference in information entropy before and after signal transduction; therefore, the KLD is an appropriate approach for quantitatively evaluating differences before and after receptor

[1]Center for anatomical, pathological, and forensic medical researches, Graduate School of Medicine, Kyoto University, Kyoto 606-8501, Japan. [2]Drug and Discovery Medicine, Graduate School of Medicine, Medical Innovation Center, Kyoto University, Kyoto 606-8507, Japan. [3]Department of Physics, Graduate School of Science, Tohoku University, Aramaki, Aoba-ku 6-3, Sendai 980-8578, Japan. [4]Department of Molecular Biosciences, Radiation Effects Research Foundation, Minami-ku, Hiroshima 732-0815, Japan. [5]Department of Tumor Research, Kitano Hospital, The Tazuke Kofukai Medical Research Institute, Kita-ku, Osaka 530-8480, Japan. ✉email: tsuruyam@kuhp.kyoto-u.ac.jp; tatsuakitsuruyama@gmail.com
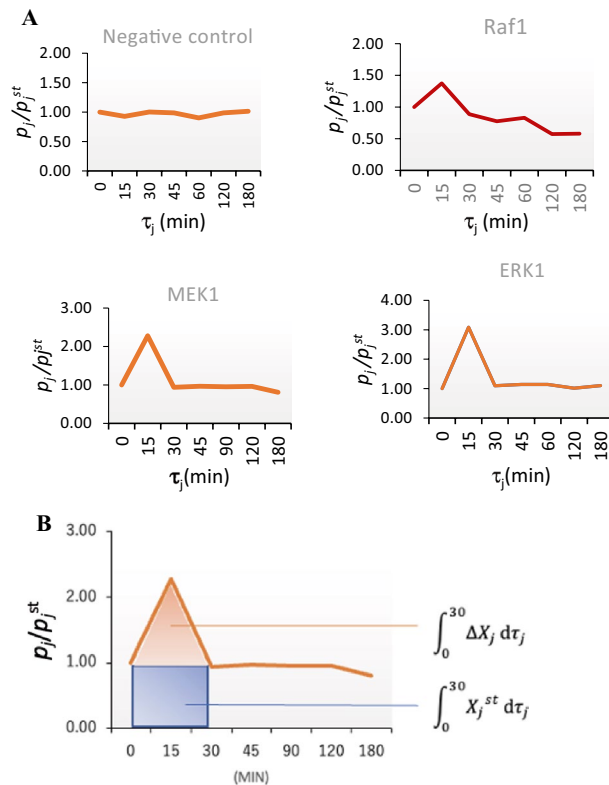
Figure 1. (**A**) Time course of the phosphorylation of signaling molecules in starved A431 cells. The ratio $p_j/p_j^{st} = X_j/X_j^{st}$ is shown on the y-axis, and the time after epidermal growth factor (EGF) stimulation (min) is shown on the x-axis. Plots of actual experimental time-course data are provided (S1 Data and S2 Data). The negative control, Raf1, MEK1, and ERK1 plots represent the ratio intensity. The typical time course pattern indicates that the phosphorylation of one signal molecule increases from $t = 5$ to 15 min after stimulation and declines from $t = 15$ to 45 min. (**B**) Time course of phosphorylation of the signal molecule X_j. Information entropy gain ΔI_j ($= \log p_j/p_j^{st}$) was calculated using the ratio $\int_0^{\tau_j} (\Delta X_j + X_j^{st}) \, dt / \int_0^{\tau_j} X_j^{st} \, dt$. In this case, $\tau_j = 30$ min.

stimulation. This study utilized the KLD to quantify signal transduction in the Raf1-MEK1-ERK1 and MAPK signaling cascades in A431 cells.

Results

Signal transduction model. First, we consider the signal cascade consisting of n steps in which the signal molecules are phosphorylated. The j-th molecule X_j denotes the signal molecule that is phosphorylated at the j-th step ($1 \le j \le n$). For example, in Raf1-MEK1-ERK1, the 1st, 2nd, and 3rd molecules are Raf1, MEK1, and ERK1, respectively. The initial proportion of the j-th signal molecules participating in signal transduction before EGF stimulation is expressed as follows:

$$p_j^{st} = X_j^{st}/X, \tag{1}$$

where X ($= \Sigma_{j=1}^n X_j$) denotes the sum of signal molecules, and the superscript st denotes the stable state before a signal event. The proportion after EGF stimulation is expressed as follows:

$$p_j(t) = X_j(t)/X \tag{2}$$

$X_j(t)$ and $p_j(t)$ are the functions of time t during the signal event, and $t = 0$ represents the time for which the cell is stimulated. The increases in $\Delta X_j(t) = X_j(t) - X_j^{st}$ and $\Delta p_j(t)$ ($= p_j(t) - p_j^{st}$) depend on the amount of time that has passed, and plots of the actual experimental time-course data are provided (S1 Data). The typical time course pattern of $p_j(t)$ (and X_j) increases and then decreases to p_j^{st} during τ_j (Fig. 1A)[20]. This pattern indicates that the phosphorylation of one signal molecule increases within the first few minutes after stimulation and declines to the initial level within 3 h (Fig. 1A). Subsequently, we designated the average signal duration of the jth signal molecule X_j as τ_j, which indicates the duration from the increase in the phosphorylated signal to the decline to the initial phosphorylation level of the molecule. Using τ_j, total signal event duration τ is defined as the sum of each phosphorylation duration multiplied by the molecule concentration.

$$\tau = \sum_{j=1}^n X_j \tau_j = X \sum_{j=1}^n p_j \tau_j \tag{3}$$

Information entropy is expressed as $I_j = -\log p_j$, and the information entropy gain during the signal event is denoted as $\Delta I_j = \log p_j - \log p_j^{st} = \log p_j/p_j^{st}$. The average information entropy gain in a given signal cascade is expressed as $\Sigma_j\, p_j \log p_j/p_j^{st}$[7]. This form is regarded as the signal transduction quantity (ST), which is identical to the KLD, a relative entropy.

In the current study, we evaluated the EGFR signal cascade comprising the chain of phosphorylation reactions of Ras, Raf1, MEK1, and ERK1. Based on the definition above, the signal amount of the EGFR cascade was defined using the KLD[4,5,21]. For example, when Raf1 has a signal value of $-\log p_{Raf1}^{st}$ before the signal event, $-\log p_{Raf1}$ changes in a time-dependent manner following EGFR stimulation. In this case, the signal information entropy gain for Raf1 phosphorylation was determined by calculating the $\log (p_{Raf1}/p_{Raf1}^{st})$. Therefore, ST of Raf1-MEK1-ERK1 was calculated in the form of the KLD[2,7]:

$$\Delta I = X \sum_{j=1}^{n} p_j \Delta I_j = X \sum_j p_j \log \frac{p_j}{p_j^{st}} = \sum_j X_j \log \frac{X_j}{X_j^{st}} \tag{4}$$

where j = Raf1, MEK1, and ERK1. ΔI represents the expected ST and $I_j = \log (p_j/p_j^{st})$.

Signal transduction maximization and signal transduction rate independent of signal molecule type. The equation above was obtained from Eqs. (1)–(3); the total signal transduction rate was expressed as $\Delta I/\tau$ and each signal transduction rate (STR) as $\Delta I_j/\tau_j$. Then, the STR of the j-th signal molecule was expressed as follows[7]:

$$\frac{\Delta I_j(0 \le t \le \tau_j)}{\tau_j} = \frac{1}{\tau_j}\log \frac{p_j(0 \le t \le \tau_j)}{p_j^{st}} = \frac{1}{\tau_j}\log \frac{\int_0^{\tau_j} X_j^{st} + \Delta X_j dt}{\int_0^{\tau_j} X_j^{st} dt} = \beta_j \tag{5}$$

The integrals in the third item in Eq. (5) were calculated using the integral of the plot area (Fig. 1B). The STR of each molecule, β_j, should vary for each signal molecule. However, we previously reported that signal molecules in the same cascade yield identical STR values when the ST is maximized[21]; this is discussed in the subsequent paragraph.

We assumed that the EGFR signaling cascade maximizes the ST for the duration of signal transduction. We introduced function L using the undetermined parameters α and β as follows: $L = X \sum_j p_j \Delta I_j + \alpha \sum_j p_j + \beta X \sum_j p_j \tau_j$. Subsequently, we obtained $\partial L/\partial p_j = X\left(\log p_j/p_j^{st}\right) + \beta X \tau_j + \alpha + X$, $\partial L/\partial X = \sum_j \left(p_j \log p_j/p_j^{st} + \beta\, p_j \tau_j\right)$. To determine α and β, $\partial L/\partial p_j$ and $\partial L/\partial X$ were set to zero, which yielded $-X = \alpha$[20] and

$$\frac{1}{\tau_j}\log \frac{p_j}{p_j^{st}} = -\beta \tag{6}$$

Comparing Eqs. (5) and (6), we deduced that $\beta_j = -\beta$, indicating that each molecular STR is independent of the signal molecule type number, j. Therefore, when the ST is maximized, STRs tend toward a constant value, and we predicted that the STR of each molecule in the EGFR cascade would have a similar value[21,22].

A431 stimulation with EGF. To examine whether STRs tended toward similar values, we stimulated cultured A431 cells with EGF[13]. The A431 cells were then collected at different times for protein extraction. The phosphorylation molecule ratio, p_j/p_j^{st}, in Eq. (4) was obtained from the fluorescence intensity of an antibody microarray against phosphorylated signal molecules in the extract. For example, we sampled the cultured cells at $t = 0$, 15, 20, 45, 60, and 90 min after EGF stimulation ($t = 0$) and serially measured the phosphorylation ratio, p_j/p_j^{st}. The ratio of the common time-course pattern is shown in Fig. 1A. The ratio initially exceeded 1.0 and then decreased to 1.0 during τ_j[1]. The ratio was integrated from $t = 0$ to τ_j, which was estimated from the plot (see S1 Text and S1 Fig.).

Signal transduction rate in the Src-Raf1-MEK1-ERK1 cascade. The cascade network expected to be activated following EGF stimulation is shown. We applied the Bayesian statistic approach to evaluate the similarity of the STRs. The expected a posteriori (EAP) values in Bayesian statistics correspond to the mean in classical statistics. We set μ_j representing the EAP of STR of the jth step ($n = 6$). The STR difference ($|\mu_j - \mu_{j+1}|$) between the jth and $j+1$th steps, effect size $|\delta_j| = |\mu_j - \mu_{j+1}|/\sigma_j = \Delta\mu_j/\sigma_j$ (σ_j; posterior standard deviation), and probability of dominance (π_d) were calculated[23]. (1) When the $|\delta_j|$ value was < 0.3, which satisfied Cohen's criteria for significant data similarity, or (2) when the π_d value between the two selected signal molecules (here, jth and $j+1$th molecules) was > 0.4 and < 0.6, we considered that the STRs of the two molecules to have similar values. Specifically, (3) if the probability $\Delta\pi$ that $\mu_j > \mu_{j+1}$ (or $\mu_j < \mu_{j+1}$) was > 0.4 and < 0.6, the STRs were considered similar. When two of the three conditions of (1)–(3) were met, the STR at the jth step and the STR at the $j+1$th step were considered similar.

We prepared two types of A431 cell cultures. One culture was incubated in a medium containing fetal bovine serum, whereas the other was incubated in a medium without serum to induce starvation stress. In the absence of stress, the Src-Raf1-MEK1-ERK1 (SRME) cascade was activated according to the STR values (Fig. 2A). The $|\delta_j|$ values were < 0.3 and the π_d values were < 0.5. The probabilities $\Delta\pi$ satisfied (3) (Fig. 2B, Table 1). These data indicate that the STRs throughout the SRME cascade had similar values. Although the STR difference for ASK1-MMK3 satisfied the criteria, the cascades including steps JNK and HSP27 were not activated following stimulation because their STRs were lower (Fig. 2B, S2Data).
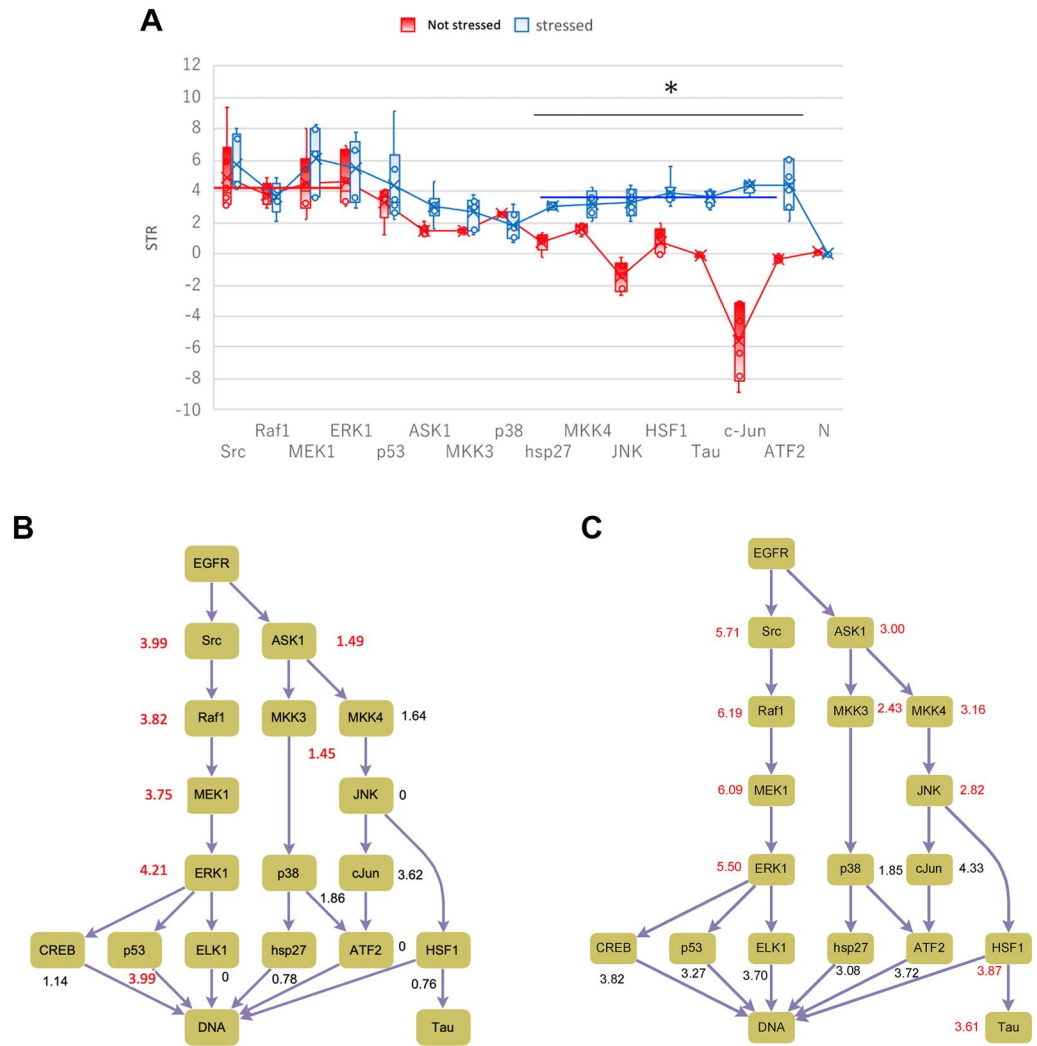
Figure 2. Signal transduction rate (STR) values in non-stressed and stressed cells. MAPK signal cascade in A431 cells. (**A**) Box plots of EGFR-related signaling molecules. The vertical axis represents the STR (min^{-1}; $n = 5$ or 6 per protein). N represents the negative control using a value of the array without antibody binding. The horizontal line with the asterisk indicates significant differences between non-stressed and stressed cells (*$p < 0.01$, in hsp27, MKK4, HSF1, Tau, c-Jun for non-stressed vs. stressed cells). (**B**) & (**C**) Schematics of the signaling cascades during stress. (**B**) represents the non-stressed state, and (**C**) represents the stressed state. The numbers in red represent the STRs in the working cascades ($|\sigma_j| \leq 0.3$). *HSF1* heat shock transcription factor 1, *hsp27* heat shock protein 27[34,35], *ATF2* activating transcription factor 2, *MKK3/4* mitogen activated protein kinase 3/4, *JNK* c-Jun N-terminal kinase.

For A431 cells under starvation stress, the stress-related ASK1-MKK4/MKK3-JNK (AMMJ) cascades were activated in addition to SRME, and the STR data met the similarity criteria, indicating that the STRs throughout the activated cascades had similar values (Fig. 2A,C, Table 1, S2 Data).

t-Distributed stochastic neighbor embedding (t-SNE) analysis. To illustrate the stress effect on the activation of signal cascades, t-distributed stochastic neighbor embedding (t-SNE) analysis was performed. t-SNE transforms high-dimensional data into two dimensions and visualizes relationships between high-dimensional data using dimensionality reduction algorithms[24]. In this research, the distance distribution of a six-dimensional vector, whose elements were the STR numerical data of six replicated measurements of phosphorylation, were used. These vector data were converted to two-dimensional data that matched the distance distribution of the vector data as well as possible.

In the absence of stress, SRME and AMMJ cascade molecules showed a linear trend distribution (Fig. 3A), indicating that the STRs of the molecules participating in SRME and AMMJ were weakly similar. Under stress conditions, the AMMJ molecule STRs clustered separately from those of the SRME molecules, indicating that the AMMJ molecule STRs were similar (Fig. 3B). The clustering can be interpreted by the signal transduction cascade

Without starvation	EAP			Probability
	STR difference	$|\sigma|$	π_d	$\Delta\pi$
SRME cascade				
Src-Raf1	0.2	**0.1**	**0.5**	**0.6**
Raf1-MEK1	0.0	**0.0**	**0.5**	**0.5**
MEK1-ERK	0.5	**0.2**	**0.4**	**0.6**
ERK-p53	0.2	0.6	0.3	0.9
AMMJ cascade				
ASK1-MKK4	0.2	0.6	0.3	0.9
ASK1-MKK3	0.0	**0.0**	**0.5**	**0.5**
With starvation	STR difference	$|\sigma|$	π_d	$\Delta\pi$
SRME cascade				
Src-Raf1	0.7	**0.1**	**0.5**	**0.6**
Raf1-MEK1	0.8	**0.1**	**0.5**	**0.6**
MEK1-ERK	0.1	**0.0**	**0.5**	**0.5**
ERK-p53	2.3	0.7	0.7	0.9
AMMJ cascade				
ASK1-MKK4	0.4	**0.3**	**0.4**	0.8
MKK4-JNK	0.4	**0.2**	**0.6**	0.7
JNK-HSF1	0.7	0.7	0.7	0.9
HSF1-Tau	0.0	**0.0**	**0.5**	**0.5**
ASK1-MKK3	0.3	**0.3**	**0.5**	**0.6**
MKK3-JNK	0.7	0.5	**0.6**	**0.6**

Table 1. Statistics of the expected a posteriori STR difference. EAP, expected a posteriori, is an alternative to the mean in classical statistics. A–B represents the difference between steps A and B in the cascade shown in the left column. "STR difference" represents the difference between STR of A (STR(A)) and STR of (STR(B)). $\sigma = \Delta\mu\,(A-B)/\delta\,(A)$ represents the effect size, where $\Delta\mu\,(A-B)$ represents the EAP difference between STR(A) and STR(B), and $\delta\,(A)$ represents the posterior standard deviation of STR(A). Values in the upper column "EAP" represent the EAP values of the STR difference, $|\sigma|$, and π_d. The value under "Probability" $\Delta\pi$ in the upper column represents the larger probabilities of $\Delta\mu\,(A-B)$ and $\Delta\mu\,(B-A)$ (the closer the two STRs, the closer the value to 0.5). Bold letters indicate the STR differences that satisfy similarity.

induced by stress. These different topologies of the plots indicate that STR-based numerical analysis effectively distinguishes cell states.

Discussion

The KLD-based method provides a theoretical framework for quantifying signal transduction. Our experimental results indicate that the STR of each signal molecule in the cascade approached similar values when the cascades were activated[4]. The results also suggest that the cascade was activated to maximize the ST per signal. ST maximization suggests that the signal cascade does not allow for signal redundancy and that signaling transduction in A431 cells adopts a strategy that sends as many signals as possible. Another strategy is prioritizing signal accuracy that allows for redundancy. For example, sensory adaptation systems prioritize signal accuracy[25].

Moreover, the STR analysis provides essential data on cell status. EGF stimulation in a stable state before stress mainly activates the SRME cascade, which is thought to be mainly involved in cell proliferation; however, stress-induced activation of response AMMJ cascades involving MKK3 and MKK4 was not evident. Under starvation stress, EGF stimulation activated the AMMJ cascade, as evidenced by the plots of STR similarity data and t-SNE.

Signal transduction pathways are composed of multiple steps, some of which may be phosphorylation reactions of unknown signaling molecules. In this case, if a molecule STR is similar to the STR of molecules in a given cascade, the molecule may be responsible for that unknown step in the cascade as a signal molecule. For example, the Tau transcription factor showed similar STR values at approximately 3.0 min^{-1}. HSF1 contributes to Tau phosphorylation[26,27] (Fig. 2A,C); for the HSF1-Tau step, $|\delta_j|$ satisfied the STR similarity criteria. Therefore, STRs throughout the ASK1-MKK4-JNK and HSF1-Tau cascade had similar values, suggesting that Tau phosphorylation occurs downstream of the ASK1-MKK4-JNK cascade[28] (Fig. 2A,C, S2 Data). MKK3 and MKK4 cooperate[29], and MKK3 can modulate JNK phosphorylation[30]. In this study, MKK4 and JNK showed similar STRs, suggesting MKK4 may participate in JNK phosphorylation in cooperation with MKK3. Thus, STR similarity aids in identifying new steps and signal molecule cooperation in signal cascades.

We used a minimal EGF dose to measure phosphorylation. Stimulation with a high EGF dose causes excessive and non-specific cell responses in non-EGFR cascades, such as the c-Met[31] and protooncogene c-kit[32] receptor-related cascades. Therefore, it remains unclear how cells react to strong in-vivo stimuli, and the ligand dose limits the applicability of our results. Further studies are needed to verify our findings and investigate the stimulation range for which the cell uses a maximization strategy.
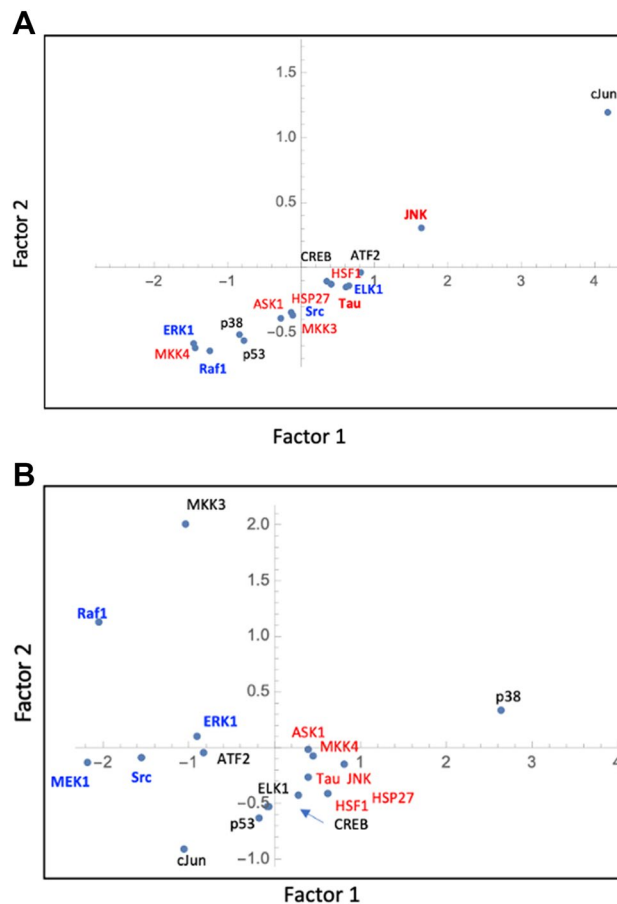
Figure 3. Dimension-reduction plot in the signal cascade in non-stressed (**A**) and stressed A431 cells (**B**). Blue letters represent the signal molecules in the SRME cascade, and MEK1 is out of the plot ($-2.73839, 3.19641$) in (**A**). Red letters represent the signal molecules in stress-related AMMJ.

Conclusions

The signal transduction system in A431 cells prioritizes the transduction of as many signals per unit time as possible, which may be an essential property of cell signal transduction. In addition, our approach based on the ST and STR is a promising method for identifying novel active cascades and investigating cellular responses.

Methods

Cell culture. The A431 human cell line derived from epidermoid carcinoma was obtained from RIKEN BioResource Research Center (Tsukuba, Japan). For the first experiment, A431 cells (0.6×10^5) in 6 mL of Dulbecco's Modified Eagle's Medium (Nacalai Tesque, Kyoto, Japan) containing 10% fetal bovine serum were cultured in a 5% CO_2 atmosphere at 37 °C for five days. EGF (100 ng/mL; Cell Signaling Technology, Danvers, MA, USA) or phosphate-buffered saline as a negative control was added to the cultures and incubated for 0 (untreated), 15, 30, 45, 60, 120, and 180 min. The total cell protein was extracted using a radioimmunoprecipitation buffer containing a protease inhibitor and phosphatase inhibitor cocktail (Nacalai Tesque). The cell extract was purified using an antibody array assay kit (Full Moon BioSystems, Inc., Sunnyvale, CA, USA).

Antibody array assay. Antibody arrays were purchased from Full Moon BioSystems (PMK185 and PEG214). Biotinylation of the proteins and conjugation and detection by Cy3-streptavidin (PA43001; GE Healthcare Life Science, Little Chalfont, UK) were performed using an antibody array assay kit (Full Moon BioSystems) according to the manufacturer's instructions. Protein samples (60 μg) were used in the antibody array assay. To calculate the STR using Eq. (5), the detection assay was performed six times independently following the addition of EGF to the cultures (50 ng/mL using the PEG214 array and 100 ng/mL using the PMK185 array). The mean fluorescent intensity (± standard deviation) was utilized to calculate the transduction characteristics. The antibody arrays were scanned using a SureScan Microarray Scanner (G2565CA Microarray Scanner System; Agilent Technologies, Santa Clara, CA, USA), after which the acquired image data were analyzed. The signal intensity was normalized by dividing the result by the negative control values from the array. Each value for phosphorylated proteins was divided by the respective value for unphosphorylated proteins at 0, 15, 30, 45, 60, 120, and 180 min. Finally, the results were divided by the value at 0 min to calculate the increase in phosphorylated molecules. The coefficient of variation for six replicates was < 0.1[33].

Phosphorylation plot. Before adding EGF, A431 cells were cultured with or without serum starvation. After adding EGF to the medium, the cells were collected at different times (0, 15, 30, 45, 60, 120, and 180 min) to extract proteins. Next, the change in the phosphorylation ratio $\int_0^{\tau j} (X_j^{st} + \Delta X_j)\, dt / \int_0^{\tau j} X_j^{st}\, dt$ of each phosphorylated protein was plotted by measuring the fluorescence levels of the antibodies against the protein in the microarray as the amount of phosphorylated X_j. The ratio was estimated by measuring the relative fluorescence intensity of the antibody microarray. The time required for the phosphorylation level to return to that observed before adding EGF was considered the signal duration, τ_j. Specifically, assuming that the increased level of phosphorylated signal molecules at measurement time t_1 was ΔX_{1j} and that the increased level of phosphorylated signal molecules at measurement time t_2 was ΔX_{j2}, then τ_j is given by the following equation (S1 Fig.):

$$\tau_j = \frac{t_2 \Delta X_{1j} - t_1 \Delta X_{2j}}{\Delta X_{1j} - \Delta X_{2j}} \qquad (7)$$

Finally, the STR during the phosphorylation of each protein was calculated by dividing the logarithm of the ratio, p_j/p_j^{st}, by τ_j (S1 Data).

Noise level evaluation. As a negative control, we measured the fluorescence intensity of the microarray spot to which the non-specific antibody was bound. Using the negative control data (S2 Data), the noise contribution to the maximal STR was estimated as 0.10 min^{-1}.

Statistics. Bayesian statistics were applied to analyze the STR at each step. For the Markov-chain Monte Carlo method, we used the R program (https://www.r-project.org/; Rstan; http://mc-stan.org/users/interfaces/rstan) to generate random numbers to obtain a posterior predictive distribution of the average entropy change rate. The R$_{hat}$ value was 1.0 (R$_{hat}$ is a Markov-chain Monte Carlo convergence index, and the series of values is generally considered as "converged" when R$_{hat} \leq 1.1$). In this study, R$_{hat} \leq 1.1$ in the calculation.

Measured STR values from five to six experiments were input to the algorithm. Outliers were excluded; however, a minimum of five replicate values was used.

Two-dimensional reduction analysis. We conducted a two-dimensional reduction analysis using the Mathematica software (Wolfram, IL, USA). The raw data are shown in S2 Data. The STR data were represented by a four-dimensional vector, excluding the minimum and maximum data as outliers. The Mathematica core was as follows: vectors = {{a_{11}, a_{12}, a_{13}, a_{14}, a_{15}}, (a_{21}, a_{22}, a_{23}, a_{24}, a_{25}),…,{a_{j1}, a_{j2}, a_{j3}, a_{j4}, a_{j5}},…}; DimensionReduce[vectors, 2]. {a_{j1}, a_{j2}, a_{j3}, a_{j4}, a_{j5}} indicates the list of the mean of the STR of the jth molecule. The vertical and horizontal axes represent the relative topological distances determined in the dimension reduction analysis. The distribution pattern indicates the relationship between the STR data, and clustering of the data in the plots indicates correlation of those data.

Data availability
The datasets generated and/or analysed during the current study are available from the corresponding author upon reasonable request.

References
1. Purutçuoğlu, V. & Wit, E. Estimating network kinetics of the MAPK/ERK pathway using biochemical data. *Math. Problems Eng.* **2012**, 1–34. https://doi.org/10.1155/2012/752631 (2012).
2. Dixit, P. D., Lyashenko, E., Niepel, M. & Vitkup, D. Maximum entropy framework for predictive inference of cell population heterogeneity and responses in signaling networks. *Cell Syst.* **10**, 204-212.e208. https://doi.org/10.1016/j.cels.2019.11.010 (2020).
3. Cheong, R., Rhee, A., Wang, C. J., Nemenman, I. & Levchenko, A. Information transduction capacity of noisy biochemical signaling networks. *Science* **334**, 354–358. https://doi.org/10.1126/science.1204553 (2011).
4. Uda, S. & Kuroda, S. Analysis of cellular signal transduction from an information theoretic approach. *Semin. Cell Dev. Biol.* **51**, 24–31. https://doi.org/10.1016/j.semcdb.2015.12.011 (2016).
5. Uda, S. *et al.* Robustness and compensation of information transmission of signaling pathways. *Science* **341**, 558–561. https://doi.org/10.1126/science.1234511 (2013).
6. Tsuruyama, T. Entropy in cell biology: Information thermodynamics of a binary code and szilard engine chain model of signal transduction. *Entropy* **20**, 617. https://doi.org/10.3390/e20080617 (2018).
7. Tsuruyama, T. Analysis of cell signal transduction based on Kullback–Leibler divergence: Channel capacity and conservation of its production rate during cascade. *Entropy*. **20** (2018).
8. Kullback, S. & Leibler, R. A. On information and sufficiency. *Ann. Math. Stat.* **22**(1), 79–86 (1951).
9. Cannon, L., Garcia, C. A. V., Piovoso, M. J. & Zurakowski, R. Prospective HIV clinical trial comparison by expected Kullback–Leibler divergence. *Proc. Am. Control Conf.* **2016**, 1295–1300. https://doi.org/10.1109/ACC.2016.7525096 (2016).
10. Wang, C. P. & Ghosh, M. A Kullback–Leibler divergence for bayesian model diagnostics. *Open J. Stat.* **1**, 172–184. https://doi.org/10.4236/ojs.2011.13021 (2011).
11. Dragalin, V., Fedorov, V., Patterson, S. & Jones, B. Kullback–Leibler divergence for evaluating bioequivalence. *Stat. Med.* **22**, 913–930. https://doi.org/10.1002/sim.1451 (2003).
12. Cai, B., Tomida, A., Mikami, K., Nagata, K. & Tsuruo, T. Down-regulation of epidermal growth factor receptor-signaling pathway by binding of GRP78/BiP to the receptor under glucose-starved stress conditions. *J. Cell Physiol.* **177**, 282–288. https://doi.org/10.1002/(SICI)1097-4652(199811)177:2%3c282::AID-JCP10%3e3.0.CO;2-C (1998).
13. Kozyulina, P. Y., Okorokova, L. S., Nikolsky, N. N. & Grudinkin, P. S. p38 MAP kinase enhances EGF-induced apoptosis in A431 carcinoma cells by promoting tyrosine phosphorylation of STAT1. *Biochem. Biophys. Res. Commun.* **430**, 331–335. https://doi.org/10.1016/j.bbrc.2012.11.041 (2013).

14. Ruiter, G. A., Verheij, M., Zerp, S. F., Moolenaar, W. H. & Van Blitterswijk, W. J. Submicromolar doses of alkyl-lysophospholipids induce rapid internalisation, but not activation, of epidermal growth factor receptor and concomitant MAPK/ERK activation in A431 cells. *Int. J. Cancer* **102**, 343–350. https://doi.org/10.1002/ijc.10741 (2002).
15. Xue, L. & Lucocq, J. M. Low extracellular pH induces activation of ERK 2, JNK, and p38 in A431 and Swiss 3T3 cells. *Biochem. Biophys. Res. Commun.* **241**, 236–242. https://doi.org/10.1006/bbrc.1997.7759 (1997).
16. Yu, Y. *et al.* The tyrphostin AG1478 augments oridonin-induced A431 cell apoptosis by blockage of JNK MAPK and enhancement of oxidative stress. *Free Radic. Res.* **46**, 1393–1405. https://doi.org/10.3109/10715762.2012.720017 (2012).
17. Zhang, X., Liu, X., Kang, S., Liu, C. & Hao, Y. Resveratrol enhances the effects of ALA-PDT on skin squamous cells A431 through p38/ MAPK signaling pathway. *Cancer Biomark.* **21**, 797–803. https://doi.org/10.3233/CBM-170495 (2018).
18. Liu, Z. M. & Huang, H. S. As2O3-induced c-Src/EGFR/ERK signaling is via Sp1 binding sites to stimulate p21WAF1/CIP1 expression in human epidermoid carcinoma A431 cells. *Cell Signal* **18**, 244–255. https://doi.org/10.1016/j.cellsig.2005.04.006 (2006).
19. Li, D., Wu, L. J., Tashiro, S., Onodera, S. & Ikejima, T. Oridonin-induced A431 cell apoptosis partially through blockage of the Ras/Raf/ERK signal pathway. *J. Pharmacol. Sci.* **103**, 56–66. https://doi.org/10.1254/jphs.fpj06016x (2007).
20. Brillouin, L. *Science and Information Theory* 2nd edn, 42 (Dover Publication Inc, 2013).
21. Tsuruyama, T. Information thermodynamics derives the entropy current of cell signal transduction as a model of a binary coding system. *Entropy* **20**, 145. https://doi.org/10.3390/e20020145 (2018).
22. Tsuruyama, T. Non-equilibrium thermodynamics of biological signal transduction predicts conservation of entropy production rate. *J. Theor. Biol.* **472**, 84–87. https://doi.org/10.1016/j.jtbi.2019.04.008 (2019).
23. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* (Routledge, 1988).
24. Sanchez, J. E. *et al.* BiasNet: A model to predict ligand bias toward GPCR signaling. *J. Chem. Inf. Model* **61**, 4190–4199. https://doi.org/10.1021/acs.jcim.1c00317 (2021).
25. Lan, G., Sartori, P., Neumann, S., Sourjik, V. & Tu, Y. The energy-speed-accuracy tradeoff in sensory adaptation. *Nat. Phys.* **8**, 422–428. https://doi.org/10.1038/nphys2276 (2012).
26. Freilich, R. *et al.* Competing protein-protein interactions regulate binding of Hsp27 to its client protein tau. *Nat. Commun.* **9**, 4563. https://doi.org/10.1038/s41467-018-07012-4 (2018).
27. Kim, E., Sakata, K. & Liao, F. F. Bidirectional interplay of HSF1 degradation and UPR activation promotes tau hyperphosphorylation. *PLoS Genet.* **13**, e1006849. https://doi.org/10.1371/journal.pgen.1006849 (2017).
28. Yu, A. *et al.* Tau protein aggregates inhibit the protein-folding and vesicular trafficking arms of the cellular proteostasis network. *J. Biol. Chem.* **294**, 7917–7930. https://doi.org/10.1074/jbc.RA119.007527 (2019).
29. Llopis, A. *et al.* The stress-activated protein kinases p38α/β and JNK1/2 cooperate with Chk1 to inhibit mitotic entry upon DNA replication arrest. *Cell Cycle* **11**, 3627–3637. https://doi.org/10.4161/cc.21917 (2012).
30. Sun, Y. *et al.* MKK3 modulates JNK-dependent cell migration and invasion. *Cell Death Dis.* **10**, 149. https://doi.org/10.1038/s41419-019-1350-6 (2019).
31. Lee, Y. J. *et al.* Expression of the c-Met proteins in malignant skin cancers. *Ann. Dermatol.* **23**, 33–38. https://doi.org/10.5021/ad.2011.23.1.33 (2011).
32. Zhang, L., Yang, X., Zhao, B. & Cai, Z. Acquired resistance to EGFR tyrosine kinase inhibitor in A431 squamous cell carcinoma xenografts is mediated by c-Kit pathway transduction. *Tumour Biol.* **36**, 2993–2999. https://doi.org/10.1007/s13277-014-2932-7 (2015).
33. Zong, Y., Zhang, S., Chen, H. T., Zong, Y. & Shi, Y. Forward-phase and reverse-phase protein microarray. *Methods Mol. Biol.* **381**, 363–374. https://doi.org/10.1007/978-1-59745-303-5_18 (2007).
34. Li, J. *et al.* NFkappaB regulates HSF1 and cJun activation in heat stress-induced intestinal epithelial cell apoptosis. *Mol. Med. Rep.* **17**, 3388–3396. https://doi.org/10.3892/mmr.2017.8199 (2018).
35. Zhu, X. N. *et al.* PP2A-AMPKalpha-HSF1 axis regulates the metal-inducible expression of HSPs and ROS clearance. *Cell Signal* **26**, 825–832. https://doi.org/10.1016/j.cellsig.2014.01.002 (2014).

Acknowledgements

Author contributions
Conceptualisation: T.T., data curation: T.T., funding acquisition: T.T., investigation: K.K.F., T.T., software: K.K.F., T.T., writing—original draft: T.T.

Funding

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-20663-0.

Correspondence and requests for materials should be addressed to T.T.

Reprints and permissions information is available at www.nature.com/reprints.

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