Evaluation of information flows in the RAS-MAPK system using transfer entropy measurements

Nobuhisa Umeki[a], Yoshiyuki Kabashima[b,c], Yasushi Sako[a,*]

[a]Cellular Informatics Laboratory, RIKEN, Cluster for Pioneering Research
2-1 Hirosawa, Wako 351-0198, Japan
[b]Institute for Physics of Intelligence, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Toko 113-0033, Japan
[c]Trans-Scale Quantum Science Institute, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Toko 113-0033, Japan

[*]To whom correspondence may be addressed. Email: sako@riken.jp

The RAS-MAPK system serves as a prominent signal processing network within animal cells, controlling various cell fates such as proliferation, differentiation, and cell death. Dysregulation of this system has been identified in genetic diseases and cancer affecting diverse tissues. To better understand the dynamics of this system, we employed information flow analysis based on transfer entropy (TE) between the activation dynamics of two key elements: SOS, a guanine nucleotide exchanger for the small GTPase RAS, and RAF, a RAS effector serine/threonine kinase. TE analysis enables a model-free assessment of the direction and strength of the control capability between components within the network. Here, we analyzed the information flows HeLa cells stimulated with EGF. Notably, we observed a switching of information flows during the progression of signal transduction. In the early phase, the dominant regulator was the forward information flow from SOS to RAF, contingent on the EGF dose. However, in the late phase, the EGF-dependency diminished, and mutual regulation between SOS and RAF took precedence. TE analyses on cells treated with a MEK inhibitor or expressing a SOS mutant associated with Noonan syndrome, a human genetic disease revealed anomalies in the information flow, leading to decreased followability of RAF inputs. Our findings demonstrate that reaction network analysis based on TE holds significant promise for applications in molecular pharmacology and pathology.

Significance

Intracellular signal processing relies on intricate regulatory networks composed of various molecular reactions. Ordinary differentiation equation (ODE) modeling, which proposes the reaction formula and parameter values, is a conventional approach to analyzing intracellular reaction networks. In contrast, statistical analyses focus on correlations between the reactions, complementing ODE modeling. Among these, TE analysis is advantageous in dealing with the reaction dynamics. The dynamics of molecular activation estimated using the ODE models do not directly mean the control capability between the molecules, a factor that TE excels at evaluating. In this study, we analyzed the dynamics of control capabilities between SOS and RAF. The analysis has revealed context-specific information flows within the cell.

Introduction

Assessing the functions of intracellular reaction networks is a central focus of contemporary molecular cell biology. Significant progress has been made through biochemical and genomic analyses, revealing the networks' elements and their interconnections. Analytical models based on coupled ordinary differential equations (ODEs) are widely used to explore the most probable network structure and determine reaction parameter values. These models offer mechanistic insights into the network and facilitate predictions of molecular activation dynamics (Omony, 2014; Travassoly et al., 2018). However, due to the networks' molecular diversity and structural complexity, the credibility of analytical models remains partially uncertain in many cases. Moreover, in networks with multiple bifurcations and loops, it becomes challenging to compare the functional importance of each reaction pathway, even with a plausible analytical model in hand. It is crucial to note that the concentration of activated molecules estimated in ODE models does not necessarily imply the ability to control downstream molecules. Control capability refers to how changes in the upstream components lead to different responses downstream. Unfortunately, this aspect is overlooked in ODE models, which usually do not account for reaction fluctuations.

Statistical models complement analytical ones (Waltermann and Klipp, 2011). Statistical analyses are fundamentally model-free, making them robust when information about the network structure and reaction formulas is limited. Results obtained from information analyses based on statistics are expressed in universal units (e.g., bits or nats), independent of the specific properties and functions of individual molecular elements within the network. Mutual information (MI) and transfer entropy (TE) are examples of measures used in information analyses (Schreiber, 2000). TE, in particular, possesses the power to quantify the information flow, which reflects the causal relationships between different reaction time courses (Granger, 1969; Barnett et al., 2009). Since TE primarily means the control capability of upstream behavior on downstream components, the TE time course provides insights into the dynamics of signal transduction from the downstream effect perspective, a crucial aspect often lacking in ODE analyses.

In practice, though mutual information (MI) analysis has found various applications in examining snapshots of cell signaling reactions (Uda et al., 2013; Miyagi et al., 2021; Fuchsberger et al., 2023), the use of transfer entropy (TE) assessment for studying the dynamics of real intracellular reaction networks has been significantly limited up to the present time. This is primarily because conventional TE calculation methods require extensive simultaneous data sequences and substantial computational power. In experimental biology, most TE analyses have been restricted to neuronal responses (Kamberaj and van der Vaart, 2009; Wibral et al., 2011; Stetter et al., 2012; Ursino et al., 2020), which typically exhibit high reproducibility and low noise. There have been only a few applications of TE to study noisy intracellular chemical reactions (Pahle et al., 2008; Ito and Sagawa, 2015). However, we recently have reported a method for TE assessment applicable to relatively small

numbers of time sequences (Imaizumi et al. 2022). This new approach involves utilizing a Gaussian approximation of the response distributions, enabling the rapid evaluation of TE using covariance matrices. Additionally, we have developed a method for result evaluation using bootstrapping. By applying these techniques, we detected significant levels of TE between SOS and RAF proteins following epidermal growth factor (EGF) stimulation, utilizing several hundred single-cell simultaneous response time courses.

SOS and RAF are essential components of the EGFR-RAF-MAPK system, which plays a crucial role in regulating cell fate (Lemmon and Schlessinger, 2010). Upon activation of the EGF receptor (EGFR), tyrosine phosphorylations occur in its cytoplasmic domain, leading to the translocation of a protein complex comprising GRB2 and SOS from the cytoplasm to the plasma membrane, where they recognize the EGFR phosphorylations. SOS functions as a guanine-nucleotide exchanger for RAS, a small GTPase located on the cytoplasmic side of the plasma membrane. This translocation of SOS induces the activation of RAS, which, in turn, results in the translocation of RAF, a serine/threonine kinase and an effector of RAS, from the cytoplasm to the plasma membrane. These intracellular translocations of SOS and RAF can be visualized using a total internal fluorescence microscope with fluorescently labeled SOS and RAF molecules, specifically observing near the basal surface of cells (Nakamura et al., 2017; Yoshizawa et al., 2021).

Given the significance of RAS as a key regulator of various cellular responses, understanding the signal transduction from SOS to RAF, from the upstream to downstream events involved with RAS activation, is of interest. Human cancer and genetic diseases, such as Noonan syndrome (Aoki and Matsubara, 2013; Aoki et al., 2016), often involve various mutations in SOS, which are expected to lead to the hyperactivation of the RAS-MAPK system. However, the specific molecular mechanisms underlying these SOS mutations remain largely unknown. In a previous study (Imaizumi et al., 2022), we observed that one of the Noonan mutations of SOS (R1131K) affected the information flows from RAF to SOS. This finding raised the possibility that TE between SOS and RAF might provide insights into the functions of the RAS-MAPK system, yet detailed analysis is lacking. In the present study, we aim to investigate the properties of TE between SOS and RAF more precisely to demonstrate the potential of TE analysis for studying intracellular reaction networks and gain deeper insights into SOS/RAF signaling dynamics.

Results

Simultaneous measurement of SOS and RAF responses in single cells

To visualize intracellular translocations of hSOS1 and CRAF subtypes, hSOS1 and CRAF, hSOS1-Halo and GFP-CRAF (hereafter SOS and RAF, respectively) were expressed in HeLa cells, in which the intrinsic wild-type hSOS1 gene was knockout (Fig. 1A). Upon EGF stimulation, the phosphorylation of both intrinsic and GFP-fused RAF at S338 increased, indicating their activation (Fig. 1B). After the tetramethylrhodamine (TMR)-ligand conjugation to Halo, cells were stimulated with EGF, and changes in fluorescent SOS and RAF densities on the basal plasma membrane were quantified simultaneously in single cells through epi and total internal reflection (TIR) fluorescence microscopies (Fig. 1C and Movies S1-S4). The intensities of translocation responses varied with the EGF dose (ranging from 0 to 100 ng/ml; Fig. 1D). To quantify SOS and RAF responses, the integration of response time profiles during 0-60 minutes of EGF stimulation was measured (Fig. 1E). The RAF response exhibited higher saturation with varying EGF doses compared to that of SOS.

Transfer entropy between SOS and RAF dynamics

TE (Fig. 2A) was calculated between the single-cell SOS and RAF dynamics in the total measurements under the 0.1–100 ng/ml EGF concentration (Fig. 2B, C). Statistically significant levels of TE (denoted by the red arrows) were detected both in the direction of SOS to RAF and RAF to SOS. The TE detections were found to be reliable, as they were not caused by the limited number of data sequences (Fig. S1) and were unlikely to be attributed to reaction noise (Fig. S2). Detections of both fw- and bw-TEs under 100 ng/ml EGF stimulation reproduced our previous result qualitatively (Imaizumi et al., 2022). This time, we also examined responses to the lower EGF doses and detected significant levels of TE (Fig. 2D). The TE values demonstrated a systematic dependence on the EGF dose (Fig. 2E), offering promising biological insights from the TE analysis. For the indication of information flow, we employed the sum of significant TEs along the time course (TEsum). Notably, fw-TEsum increased with the EGF dose, while bw-TEsum exhibited maximum values under 10 ng/ml EGF in our experiments (Fig. 2E).

The amount of TE demonstrated a dependence on the lag times (p, q) used for calculation, as expected from the definition. In general, TEsum increased with the lag time of the causal sequence (e.g., p for SOS to RAF) and decreased with the lag time of the result sequence (e.g., q for SOS to RAF) (Fig. 2C, E). However, we identified some exceptions in the lag time dependence of bw-TE, indicating complexity in the information flows. After an initial sharp spike at time 0, both fw- and bw-TE time courses exhibited a peak at around 10 minutes of EGF stimulation and a plateau at a later period (≥ 20 minutes) (Fig. 2B). Particularly in the bw-TE time courses, the early peak vanished with the elongation of the SOS lag time (p), suggesting a phase-dependent switching of the information flow.

144

145 **Transfer entropy analysis conditioned to the EGF dose**

146 It is important to emphasize that TE detection does not always imply true causality. One potential
147 reason for pseudo-TE generation is the regulation of two downstream reactions through independent
148 pathways originating from a common upstream factor (Fig. S3). Besides the SOS/RAS-mediated
149 pathway, EGFR may also regulate RAF through SOS-independent pathways. To investigate this
150 possibility, we compared the TE amounts conditioned and non-conditioned to the EGF dose (Vakorin
151 et al., 2009). We obtained the conditional TEsum as the weighted average (relative to the data number)
152 of TEsums for each EGF dose (Fig. 2D), and calculated the non-conditional TEsum for randomly
153 selected data sets from mixed data under all EGF concentrations (Fig. 3A). The direct effect of EGF
154 was mostly rejected in the conditional TE. To assess the significance of differences between
155 conditional and non-conditional TEsums, we repeated random selections for non-conditional TE
156 calculation 100 times to estimate the 5-95% confidence range. Due to the biphasic nature of TE time
157 courses, we separately obtained TEsums during the early and late phases (Fig. 3B).

158 During the early phase, most conditional fw-TEsums calculated at varying lag times were
159 smaller than the lower 5% threshold in the distributions of non-conditional TEsums. However, in other
160 cases, the conditional TEsums and non-conditional TEsums were indistinguishable. When there were
161 direct pathways from EGFR to RAF, bw-TE in the early phase should be affected by EGF dose, similar
162 to the observed fw-TE, due to the symmetry in the reaction network (Fig. S3 and S5A). Simulation of
163 simple reaction models, including a feedback loop, replicated the detection of bw-TE (Fig. S4) and
164 the findings of the conditional TE analysis for SOS and RAF during the early phase (Fig. S5B). Based
165 on these results, we propose a phase-dependent information flow as depicted in Figure 3C. A feedback
166 loop exists between SOS and RAF. In the early phase, variation in the EGF input increases the forward
167 information flow from SOS to RAF, possibly due to differences in local and global response
168 fluctuations across the fixed and wide range of EGF doses, respectively. In the late phase, direct
169 regulation by EGF becomes negligible, and the feedback loop drives autonomously (Fig. 3C).

170

171 **Response/transfer entropy relationships**

172 We measured the phase-dependent responses of SOS and RAF (Fig. 4A) in the total time courses
173 during 60 min (Fig. 1D). In the early phase, RAF exhibited a more saturable response to increasing
174 EGF doses compared to SOS. In the late phase, both proteins showed a response peak at an EGF
175 concentration of 10 ng/ml. The ratio between SOS and RAF responses (Fig. 4A insets) exhibited a
176 weak dependence on the EGF dose, as expected from the putative reaction network at this stage (Fig.
177 3C). Subsequently, we analyzed the EGF dose dependency of the TEsum for each reaction phase and
178 combination of lag times (Fig. 4B). In most TEsum profiles, we observed a minimum at 1 ng/ml EGF
179 and a maximum at 10 ng/ml EGF. Since these TEs were calculated at each fixed EGF concentration,

these EGF dose dependencies were not directly caused by the difference in EGF input itself but by the responses of SOS and RAF to the EGF input. Specific fluctuations in the SOS (RAF) response under 1 and 10 ng/ml EGF appeared to lead to strong and weak control capabilities of the RAF (SOS) response, respectively. It is worth mentioning that in the late phase, large TEsums were observed in some calculations at 0.1 ng/ml EGF. Importantly, these EGF dose dependencies of TE differed from those of the reaction response (Fig. 4A), particularly in the backward TE.

To directly examine the response/TE relationships, we sorted the time series data into four groups based on the SOS or RAF response and calculated TE for each group (Fig. 4C). In the early phase, fw-TEsum showed saturation concerning the increase of SOS response, the source of fw-TE, indicating that the SOS capability to control RAF response weakly depended on the average intensity of SOS activation. On the other hand, bw-TEsum profiles against the RAF response, the source of bw-TE, were maximal at the middle range of RAF response for $p \leq 2$ min or minimal for $p \geq 4$ min. This finding suggested the presence of a feedback loop longer than 4 min, indicating that the long past ($\geq$ 4 min) but not the short past ($\leq$ 2 min) sequence of SOS contained similar information to that in the past (1~8 min) sequence of RAF. Additionally, backward information flow was not maximal in cells with the highest RAF response. In the late phase, both fw- and bw-TEsums were small in the middle range of SOS and RAF responses, respectively. Within this range, SOS (RAF) could not regulate RAF (SOS) effectively, and the reaction noise played a significant role in determining the variation in the RAF (SOS) response. Surprisingly, the smallest SOS and RAF responses led to large fw- and bw-TE, respectively, in the late phase. We calculated net TEsum = (fw-TEsum) – (bw-TEsum) (Gencaga et al., 2015) against SOS or RAF response (Fig 4D). In the early phase, net TEsum exhibited saturation against SOS or RAF response, with slightly positive values for similar p and q lengths (blue and red lines), indicating only slightly stronger forward regulation. In the later phase, the forward and backward regulations were comparable overall.

Effects of a MEK inhibitor and a SOS mutation

The involvement of negative feedback loops in the EGFR-RAS-MAPK system has been previously reported (Corbalan-Gracia et al., 1996; Saha et al., 2012). The detection of significant levels of bw-TE suggests that these feedback loops play a role in the information flow. To investigate this possibility further, we conducted experiments in cells pretreated with a MEK inhibitor, trametinib. MEK is a protein mediating activations of RAF and ERK. While the EGF-induced ERK phosphorylation (activation) was inhibited to the basal level in cells pretreated with 10 nM trametinib (Fig. 5A), the SOS and RAF responses were not affected evidently (Fig. 5B). We measured TE between SOS and RAF after 10 ng/ml EGF stimulation under this condition and found that the MEK inhibitor significantly increased bw-TE in the early phase (Fig. 5C). This result indicates that the feedback from MEK suppresses backward information flow under normal conditions, thereby reducing the control

216 capability of RAF on the SOS response (Fig. 5F).

217 Additionally, we examined the effects of the R1131K mutation in SOS, which is observed in
218 Noonan syndrome patients, in the presence and absence of the MEK inhibitor. The R1131K mutation
219 did not cause significant changes in the EGF (10 ng/ml) response of SOS and RAF or in the sensitivity
220 to the MEK inhibitor for ERK phosphorylation (Figs. 1B and 5A, B), but it did alter some aspects of
221 TE (Fig. 5D, E). Like the MEK inhibitor, R1131K significantly increased bw-TEsum in the early phase.
222 Moreover, R1131K significantly decreased fw-TEsum in the early phase and bw-TEsum in the late
223 phase (Fig. 1F). The MEK inhibitor exhibited similar trends to the effects of R1131K, while the latter
224 two effects of the MEK inhibitor were not statistically significant, (Fig. 5C). These results suggest that
225 the R1131K mutation in SOS interferes with the feedback regulation from RAF to SOS via MEK
226 activation. Since no significant difference was observed in TEsums for R1131K with and without the
227 MEK inhibitor (Fig. 5D, E), the feedback from MEK and the effect of R1131K likely act on the same
228 pathway to reduce the control capability of RAF on SOS in the early phase (Fig. 5F). In the late phase,
229 R1131K decreased bw-TE but did not increase it, which may seem contradictory to its effect in the
230 early phase. However, we observed that in both phases, R1131K (and the MEK inhibitor) decreased
231 bw-TE relative to that under the normal condition as the SOS lag time (p) increased, suggesting related
232 mechanisms of bw-TE regulation in these two phases.

Discussion

In this study, we conducted an assessment of transfer entropy (TE) between SOS and RAF activations in living cells stimulated with EGF. By examining the TE values over the reaction time course, we aimed to understand the control relationship between SOS and RAF. Our findings revealed that the time courses of bw-TE can be subdivided into early (< 20 min) and late (> 20 min) phases, indicating a time-dependent change in information flow. In the late phase, EGF regulation was reduced (Fig. 3), possibly due to EGFR inactivation, involving dephosphorylation and/or endocytosis (Monast et al. 2012). The effect of a MEK inhibitor suggested that MEK/ERK activation plays a role in reducing bw-TE. This effect was primarily observed in the early phase, likely because of transient MEK/ERK activation following EGF signaling (Marshall, 1995; Kao, 2001; Yoshizawa et al., 2021).

The reduction of RAF's control capability over SOS in the early phase relatively strengthens forward regulation in the SOS/RAF system, forcing RAF response to precisely follow the EGF input. Conversely, the inactivation of MEK and ERK in the late phase supports backward regulation. During this later phase, the SOS/RAF feedback loop appears to operate autonomously out of the EGF control. The activation levels of SOS and RAF at the beginning of the late phase appear to determine the quasi-steady state levels of SOS and RAF. If the dominant backward regulation involves inhibiting SOS activation, the positive forward and negative backward flow of information within the network likely forms an adaptation cycle that stabilizes SOS and RAF activation levels (Ito and Sagawa, 2015). This network type may account for the observed quasi-steady states in the late phase (Fig. 1D). The detectability of bidirectional information flow in the steady state without system perturbation as shown here, provides an advantage in information analysis.

In simulations, the detectability of bw-TE depended on the nonlinearity of the feedback reaction (Fig. S4). To produce a significant amount of bw-TE, the past sequence of the downstream reaction needed to contain information different from that in the past sequence of the upstream reaction (Fig. 2 A). The detection of bw-TE in the living cell system suggests the presence of a complex feedback reaction system that appends distinctive information to estimate SOS dynamics. The EGFR-RAS-MAPK system is known to involve several feedback loops (Lake et al., 2016). For instance, ERK activation downstream of RAF induces EGFR threonine 669 phosphorylation, enhancing EGFR internalization and suppressing EGFR kinase activity (Li et al., 2008). ERK also phosphorylates several serine residues in the GRB2-binding domain of SOS, weakening its interaction with GRB2 and reducing SOS translocation to the plasma membrane upon EGF stimulation (Corbalan-Gracia et al., 1996; Saha et al., 2012). Furthermore, desensitization of RAF to RAS signal has been reported as a result of multiple serine phosphorylations on RAF by ERK (Wartmann et al., 1997). The bw-TE from RAF to SOS may reflect these negative feedback pathways. Besides the mentioned negative feedback loops, other factors would complicate the feedback pathway and contribute to the noticeable bw-TE in living cells, such as multiple membrane binding domains of RAF (Ghosh et al., 1996; Hibino

et al. 2011), RAF phosphorylation on the plasma membrane for activation (Lavoie and Therrien, 2015), and the SOS/RAS positive feedback loop (Margarit et al., 2003).

The SOS mutation R1131K exhibited similar effects to the MEK inhibitor, leading to an increase in bw-TE in the early phase and a decrease in fw-TE. These results indicate that information flows were perturbed in Noonan cells carrying the R1131K mutation. The mechanism of onset for most mutants, including R1131K found in Noonan syndrome patients, remains unknown. Our findings suggest that R1131K inhibits the feedback pathway from MEK/ERK to SOS that decreasing the backward information flow. Additionally, R1131K reduces the forward control capability from SOS to RAF. Overall, the R1131K mutation disrupts the mutual regulation between SOS and RAF and may lead to the system being out of control. Such abnormality in the information flows caused by R1131K was difficult to detect using conventional biochemical analysis.

As demonstrated in this study, TE assessment enables the analysis of intracellular reaction networks from the perspective of reaction control capability. Importantly, the control capability does not necessarily correlate with the strength of molecular activation. Thus, TE analysis provides a complementary view of the network to that obtained from conventional ODE analysis. We observed temporal switching in the information flow network within the RAS-MAPK system and identified changes in the information flow, illustrating the applicability of TE analysis to molecular pathology and pharmacology. It is essential to note that the behavior of TE is complex, and gaining more experience with TE analysis will further improve its application in understanding intracellular systems.

Materials and Methods

Sample Preparations:

Methods for the construction of SOS and RAF plasmids and the establishment of SOS1 knockout HeLa cells have been previously described (Nakamura et al., 2017). HeLa cells were cultured in D-MEM supplemented with 10% FBS at 37°C, 5% CO_2. Overnight, cells were transferred onto 25 mm--ϕ coverslips for microscopy or on a 35 mm-ϕ cell culture dish for biochemistry and then transfected with the plasmids using Lipofectamine 3000 (Thermo Fisher). The cells were cultured in MEM containing 1% BSA for one day before being used for the experiments. In some experiments, cells were pretreated with a MEK inhibitor, trametinib (ChemScene) in DMSO (final concentration 0.05%) for 1 hr before EGF stimulation.

Microscopy:

Halo-SOS in cells on a coverslip was conjugated with 300 nM TMR Halo-tag ligand for 1 hr, washed, and SOS and RAF translocations in cells were observed in MEM, 1% BSA, 10 mM HEPES (pH 7.4). in-house total internal reflection (TIR) fluorescence microscope system, based on IX83 inverted fluorescence microscope (Olympus), was used for imaging. The microscope is equipped with a multicolor laser diode (LDI, Chroma Technology) for fluorophore excitation, a 60x NA 1.5 oil-immersion objective (UPLAPO60XOHR, Olympus), a dual-color image splitting system (W-VIEW GEMINI-2C, Hamamatsu), and two cMOS cameras (ORCA-Flush4.0, Hamamatsu). Using a 2-dimensional piezo mirror scanner (S-335, PI) in the excitation beam path, this system can alternatively switch the excitation mode between TIR illumination with a circularly rotating laser beam at 40 Hz and epi-illumination, covering a field number of 19 (320 μm in diameter under a 60x objective). At each time point of the measurement, a set of TIR and epi-illumination images were acquired successively at the same field of view under the excitation laser wavelengths of 470 nm for GFP and 555 nm for tetramethyl-rhodamine (TMR) simultaneously. The exposure time was 125 ms and 25 ms for each frame in the TIR- and epi-illumination modes, respectively. Image acquisitions were repeated with 1-minute intervals. A triple-band dichroic mirror (ZT405/470/555rpc, Chroma) was used in the microscope. In the image splitting optics, the GFP and TMR signals were separated by a dichroic mirror (T550lpxr-UF2, Chroma), and emission filters for GFP (ET510/40m, Chroma) and TMR (ET575/50m, Chroma). Just before the 4th time of image acquisition (time 0), cells were stimulated by adding 550 μl of observation medium containing EGF to the 450 μl of the same medium in the observation chamber. All measurements were performed at 25°C. Representative responses of cells are shown in Supplemental Movies S1-S4.

Image processing

After the shading and background correction, the outline of single cells in the epi-illumination images

was determined using the "cellpose" algorithm (https://github.com/mouseland/cellpose), and the average intensities per pixel in the single cell areas were measured in each video frame. Data from cells with the lower 20% of signal intensities at time 0 were removed. From the single-cell time courses of the average intensity, bleed-through between the GFP and TMR fluorescence was subtracted, drift in the basal intensity with time was removed, then the TIR signal was normalized to the epi signal. The difference in the normalized TIR signal to the 3-frame average before stimulation was plotted as the single-cell response. Image processing was performed using Image J (NIH).

Transfer entropy calculations

Transfer entropy (TE) values from SOS to RAF and RAF to SOS were calculated along the time courses of the cell responses. The method used was as precisely described by Imaizumi et al. (2022). and the source code for the calculations has been deposited to https://github.com/kabashiy/transfer_entropy/.

In brief, when considering the assessment of the uncertainty of the stochastic time sequence of X at time t (X_t) by the statistical entropy $H(X_t)$, the amount of the reduction of the uncertainty after obtaining the past sequence of X during t-p to t-1, $\boldsymbol{X}_{t-1}^{(p)}$, is $I\left(X_t;\boldsymbol{X}_{t-1}^{(p)}\right) = H(X_t) - H\left(X_t\middle|\boldsymbol{X}_{t-1}^{(p)}\right)$. Here, $H\left(X_t\middle|\boldsymbol{X}_{t-1}^{(p)}\right)$ is the conditional entropy, and I is mutual information. The TE from Y to X is defined as $TE_{Y\to X}^{(q),(p)}(t) = I\left(X_t;\boldsymbol{X}_{t-1}^{(p)},\boldsymbol{Y}_{t-1}^{(q)}\right) - I\left(X_t;\boldsymbol{X}_{t-1}^{(p)}\right)$, which indicates how much uncertainty about X_t is reduced by providing $\boldsymbol{Y}_{t-1}^{(q)}$ in addition to $\boldsymbol{X}_{t-1}^{(p)}$. Usually, the calculation of TE from real experimental data is not easy because the data number is insufficient to correctly estimate the multi-dimensional distribution of the generating function and even if the data number is sufficient, a huge amount of computational power is required. However, when the joint distribution of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is multivariate Gaussian, TE can be analytically evaluated using covariance matrices for the Gaussian model as

$$TE_{Y\to X}^{(q),(p)}(t) = \frac{1}{2}\ln\left(\frac{\sum\left(X_t\middle|\boldsymbol{X}_{t-1}^{(p)}\right)}{\sum\left(X_t|\boldsymbol{X}_{t-1}^{(p)} \oplus \boldsymbol{Y}_{t-1}^{(q)}\right)}\right).$$

Here, $\sum(X)$ and $\sum(X,Y)$ denote the covariance matrix of X and the cross-covariance matrix of X and Y, respectively, and $\sum\left(X_t\middle|\boldsymbol{X}_{t-1}^{(p)}\right) = \sum(X_t) - \sum(X_t,\boldsymbol{X}_{t-1}^{(p)})\sum(x_{t-1}^{(p)})^{-1}\sum(X_t,\boldsymbol{X}_{t-1}^{(p)})^{\mathrm{T}}$. $X \oplus Y$ refers to the concatenation of the matrices X and Y.

TE calculations were evaluated first by the confidence interval determined by the bootstrapping method with 1000-time resamplings. We estimated the 95% confidence interval for each time point.

Next, to determine the significance threshold, we generate a new sample set by random sampling and calculated the TE with the null hypothesis, TE = 0, i.e., using a setting in which all elements of the cross-covariance between $X_t \oplus \boldsymbol{X}_{t-1}^{(p)}$ and $\boldsymbol{Y}_{t-1}^{(q)}$ are vanished. The lower 1% percentile point of the distribution obtained using a 1000-time resampling was adopted as the significant threshold. When the confidence interval exceeded the significance threshold, we concluded that the TE value was statistically significant.

Numerical simulations

Stochastic simulations of reaction cascade models were performed using the program described in Shindo et al. (2018), where Langevin ODE models were solved using the 4th-order Runge-Kutta method. The supplemental text provides details of these models. The purpose of analytical modeling is to assess the behaviors of TE conceptually but not to reproduce the real intracellular reaction network.

Abbreviations

bw-TE, backward TE; EGF, epidermal growth factor; EGFR, EGF receptor; fw-TE, forward TE; MI, mutual information; ODE, ordinary differentiation equation; TE, transfer entropy; TEsum, sum of the significant TE; TIR, total internal reflection; TMR, tetramethylrhodamine.

Acknowledgments

The authors thank Mutsumi Nakanishi and Hiromi Sato for technical assistance, Sinsuke Uda at Kyushu University, Atsushi Hatano at Nigata University, and Tomoyuki Obuchi at Kyoto University for helpful discussions and suggestions. The SOS knockout Hela cell was established by Mitsuhiro Abe at RIKEN. YS and YK were funded by JST CREST (JPMJCR1912). YS and YK were funded by Grants-in-Aid for Scientific Research, MEXT (19H05647) and (22H05117), respectively.

Author contributions: NU and YS designed the research; NU and YS performed the research; YK contributed analytic tools; NU, YK, and YS wrote the paper.

The authors declare no competing interest.

References

Y. Aoki, Y. Matsubara, Ras/MAPK syndromes and childhood hemato-oncological diseases. *Int J Hematol* **97**, 30–6 (2013).

Y. Aoki, T. Niihori, S. Inoue, Y. Matsubara, Recent advances in RASopathies. *J Hum Genet* **61**, 33–9 (2016).

L. Barnett, A. B. Barrett, A. K. Seth, Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.* **103,** 238701 (1-4) (2009).

D. Cai, P. S. Choi, M. Gelbard, M. Meyerson, Identification and characterization of oncogenic *SOS1* mutations in lung adenocarcinoma. *Mol. Cancer Res.* **17,** 1002-1012 (2019)

S. Corbalan-Garcia, S. S. Yang, K. R. Degenhardt, D. Bar-Sagi, Identification of the mitogen-activated protein kinase phosphorylation sites on human Sos1 that regulate interaction with Grb2. *Mol. Cell Biol.* **16**, 5674–82 (1996).

F. F. Fuchsberger, D. Kim, N. Baranova, H. Vrban, M. Kagelmacher, R. Wawrzinek, C. Rademacher, Information transfer in mammalian glycan-based communication. *eLife* **2023,** e69415 (1-32) (2023).

D. Gencaga, K. H. Knuth, W. B. Rossow, A recipe for the estimation of information flow in a dynamical system. *Entropy*, **17,** 438-470 (2015).

C. W. J. Granger, Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37,** 424–438 (1969).

S. Ghosh, J. C. Strum, V. A. Sciorra, L. Daniel, R. M. Bell, Raf-1 kinase possesses distinct binding domains for phosphatidylserine and phosphatidic acid. Phosphatidic acid regulates the translocation of Raf-1 in 12-O-tetradecanoylphorbol-13-acetate-stimulated Madin-Darby canine kidney cells. *J Biol. Chem.* **271,** 8472-8480 (1996).

K. Hibino, T. Shibata, T. Yanagida, Y. Sako, Activation kinetics of RAF protein in the ternary complex among RAF, RasGTP, and the kinases on the plasma membrane of living cells. *J. Biol. Chem.* **286,** 36460-36468 (2011).

T. Imaizumi, N. Umeki, R. Yoshizawa, T. Obuchi, Y. Sako, Y. Kabashima, Assessing transfer entropy from biochemical data. *Phys. Rev. E.* **105,** 034403 (2021).

S. Ito, T. Sagawa, Maxwell's demon in biochemical signal transduction with feedback loop. *Nat. Comm.* **6,** 7498 (2015).

H. Kamberaj, A. van der Vaart, Extracting the causality of correlated motions from molecular dynamics simulations, *Biophysical J.* **97,** 1747-1755 (200).

S. Kao, R. K. Jaiswal, W. Kolch, G. E. Landreth, Identification of the mechanisms regulating the differential activation of the mapk cascade by epidermal growth factor and nerve growth factor in PC12 cells. *J. Biol. Chem.* **276,** 18169-18177 (2001).

D. Lake, S. A. L. Corrêa, J. Müller, Negative feedback regulation of the ERK1/2 MAPK pathway.

Cell. Mol. Life Sci. **73,** 4397-4413 (2016).

H. Lavoie, M. Therrien, Regulation of RAF protein kinases in ERK signalling. *Nat. Rev. Mol. Cell Biol.* **16**, 281-298 (2015).

M. A. Lemmon, J. Schlessinger, Cell signaling by receptor tyrosine kinases. *Cell* **141,** 1117-1134 (2010).

X. Li, Y. Hung, J. Jiang, S. J. Frank, ERK-dependent threonine phosphorylation of EGF receptor modulates receptor downregulation and signaling. *Cell Signal.* **20,** 2145-2155 (2008).

S. M. Margarit, H. Sondermann, B. E. Hall, B. Nagar, A. Hoelz, M. Pirruccello, D. Bar-Sagi, J. Kuriyan, Structural evidence for feedback activation by Ras·GTP of the Ras-specific nucleotide exchange factor SOS. *Cell* **112**, 685–695 (2003).

C. J. Marshall, Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell* **80,** 179-185 (1995).

C. S. Monast, C. M. Furcht, M. J. Lazzara, Computational analysis of the regulation of EGFR by protein tyrosine phosphatases. *Biophys. J.* **102,** 2012-2021 (2012).

H. Miyagi, M. Hiroshima, Y. Sako, Cell-to-cell diversification in ERRB-RAS-MAPK signal transduction that produces cell-type specific growth factor responses. *Biosystems* **19,** 104293 (1-10) (2021).

Y. Nakamura, N. Umeki, M. Abe, Y. Sako, Mutation-specific mechanisms of hyperactivation of Noonan syndrome SOS molecules detected with single-molecule imaging in living cells. *Sci. Rep.* **7,** 14153 (2017).

J. Omony, Biological Network Inference: A Review of Methods and Assessment of Tools and Techniques. *Annu. Res. Rev. Biol.* **4**, 577–601 (2014).

J. Pahle, A. K. Green, C. J. Dixon, U. Kummer, Information transfer in signaling pathways: A study using coupled simulated and experimental data. *BMC Bioinformatics* **9,** 139 (2008).

J. M. Rojas, J. L. Oliva, E. Santos, Mammalian son of sevenless guanine nucleotide exchange factors: old concepts and new perspectives. *Genes Cancer* **2**, 298–305 (2011).

M. Saha, A. Carriere, M., Cheerathodi, X. Zhang, G. Lavoie, J. Rush, P. P. Roux, B. A. Ballif, RSK phosphorylates SOS1 creating 14-3-3-docking sites and negatively regulating MAPK activation. *Biochem. J.* **447**, 159–66 (2012).

T. Schreiber, Measuring information transfer, *Phys. Rev. Lett.* **85,** 461-464 (2000).

Y. Shindo, Y. Kondo, Y. Sako, Inferring a nonlinear biochemical network model from a heterogeneous single-cell time course data. *Sci. Rep.* **8.** 6790 (2018).

O. Stetter, D. Battaglia, J. Soriano, T. Geisel, Model-free reconstruction of excitatory neuronal connectivity from calcium imaging signals. *PLoS Comp. Biol.* **8** e1002653 (2012).

I. Travassoly, J. Goldfarb, R. Iyngar, Systems biology primer: The basic methods and approaches. *Essays Biochem.* **62,** 487-500 (2018).

S. Uda T. N. Saito, T. Kudo, T. Kokaji, T. Tsuchiya, H. Kubota, Y. Komori, Y.-I. Ozaki, S. Kuroda, Robustness and compensation of information transmission of signaling pathways. *Science* **341,** 558-561 (2013).

M. Ursino, G. Ricci, E. Magosso, Transfer entropy as a measure of brain connectivity: A critical analysis with the help of neural mass models. *Front. Compt. Neurosic,* **14,** 45 (2020).

V. A. Vakorin, O. A. Krakovska, A. R. McIntosh, Confounding effects of indirect connections on causality estimation. *J. Neurosci. Meth.* **194,** 152-160 (2009).

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

M. Wartmann, P. Hofer, P. Turowski, A. R. Sltiel, N. E. Hynes, Negative modulation of membrane localization of the Raf-1 protein kinase by hyperphosphorylation. *J. Biol. Chem.* **272,** 3915-3923 (1997).

M. Wibral B. Rahm, M. Rieder, M. Lindner, R. Vicente, J. Kaiser, Transfer entropy in magnetoencephalographic data: Quantifying information flow in cortical and cerebellar networks *Prog. Biophys. Mol. Biol.* **103,** 80 (2011).

R. Yoshizawa, N. Umeki, A. Yamamoto, M. Okada, M. Murata, Y. Sako, p52Shc regulates the sustainability of ERK activation in a RAF-independent manner. *Mol. Biol. Cell* **32,** 1838-1848 (2021).

Figures

Figure 1. SOS and RAF responses to cell stimulation

A. Expression levels of SOS and RAF probes. Halo-SOS (wild type, wt or R1131K mutant, R/K) and GFP-RAF were expressed in SOS knockout HeLa cells. **B.** RAF phosphorylation before and after 10 min of 10 ng/ml EGF stimulation of cells. Results of representative western blotting and quantification of three independent experiments. **C.** Translocations of SOS (magenta) and RAF (green). Epi-fluorescence (upper) and TIR-fluorescence (lower) images at the indicated times after 100 ng/ml EGF stimulation. Bar, 50 µm (see Supplemental Movies S1-S4.) **D.** Time course analysis of SOS and RAF translocation observed in the total 1467 cells after vehicle (black: 150 cells) or EGF (blue: 0.1 ng/ml, 300 cells; green: 1 ng/ml, 310 cells; orange: 10 ng/ml, 370 cells; red: 100 ng/ml, 337 cells) stimulation. Changes in the molecular density on the cell surface (TIR) relative to the whole cell signal (epi) are plotted. The lower panels show the average with SD of the single cell data in the upper panels. **E.** EGF dose dependence of the SOS (red) and RAF (green) responses. Response intensities were determined as the sum of the molecular translocation during 0-60 min normalized to the maximum. Inset shows the relationship between SOS and RAF responses. Cell-to-cell averages are shown with SE.

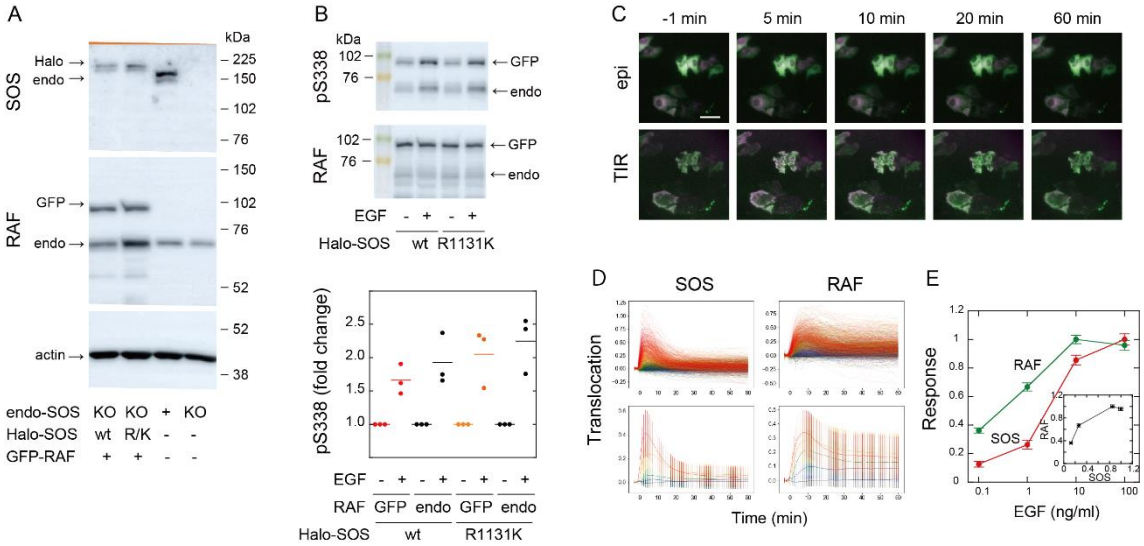


Figure 2. Transfer entropy time course analysis

A. Concept of TE between two reaction time courses. (p, q) are the time lags (past sequences) of the X (SOS) and Y (RAF) response, respectively, used for the TE calculation. **B.** TE time courses for the total 1317 pairs of translocation trajectories with 0.1-100 ng/ml of EGF stimulation. Blue line, TE. Light blue area, 95% confidence interval. Red dotted line, 1% statistical significance threshold. Red arrow, time point at which the lower limit of the confidence interval was larger than the statistical significance threshold (significant TE). **C.** Sum of the significant TEs shown in (B). **D.** EGF dose dependencies of the TE time course. EGF concentrations are indicated. p, q values were increased as 1, 2, 4, and 8 in the direction of arrows. **E.** Sum of the significant fw- (left) and bw- (right) TEs (TE_{sum}) under each EGF concentration are plotted as the functions of the lag times (p, q).

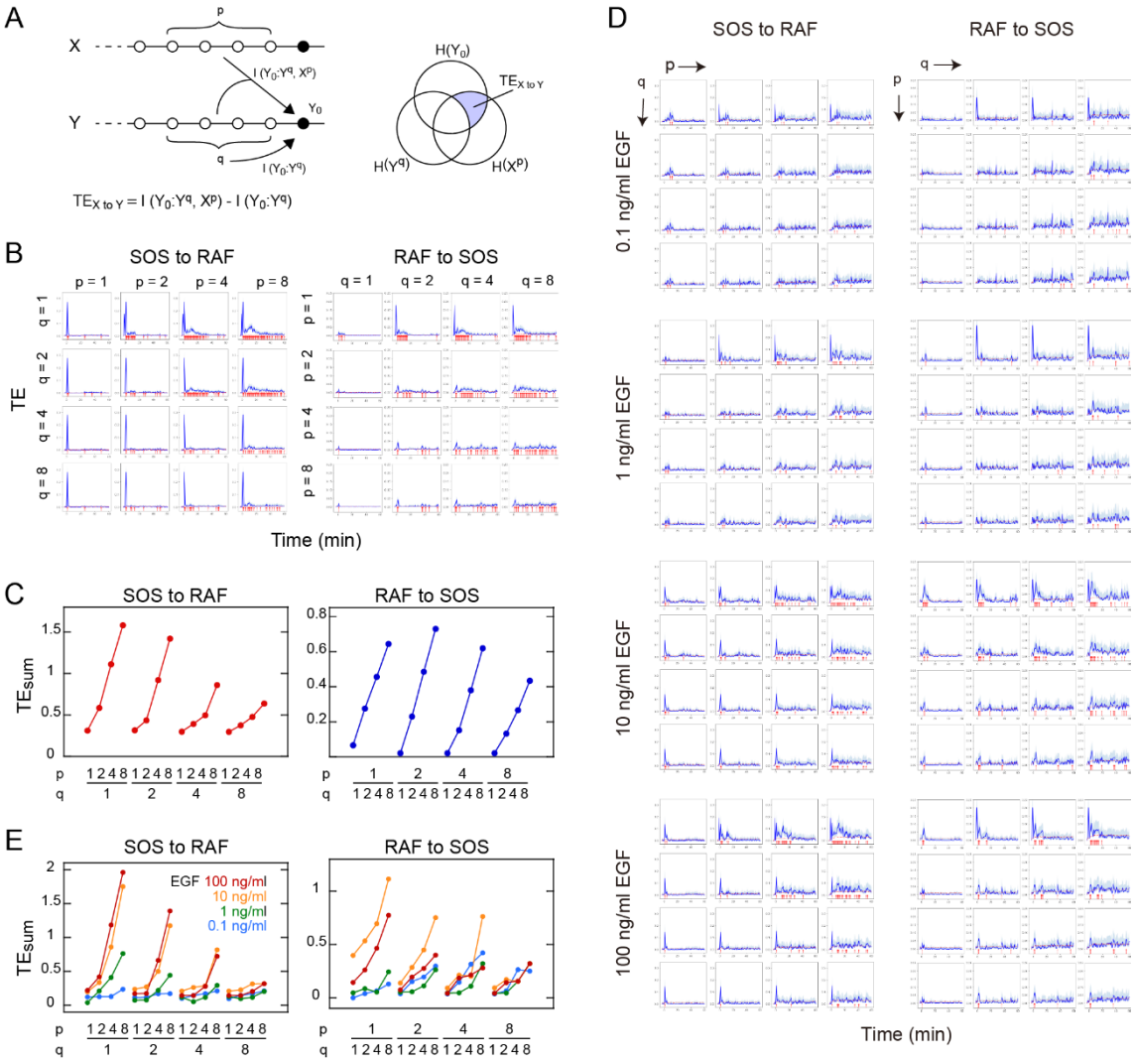


Figure 3. Conditional transfer entropy analysis with respect to the EGF dose

A. An example of the set of 329 (25%) randomly selected reaction trajectories from 1317 data points used for Figure 2B (left), from which the non-conditional fw-TE (middle) and bw-TE (right) time courses were calculated. **B.** Comparison between the non-conditional and conditional TEsum. The dotted lines indicate the lower and upper 5% thresholds in the non-conditional TEsum distributions. Solid lines indicate the average of the conditional TEsum values (Fig. 2E) weighted with the data number. **C.** Information flows suggested from the conditional TE analysis (see text).

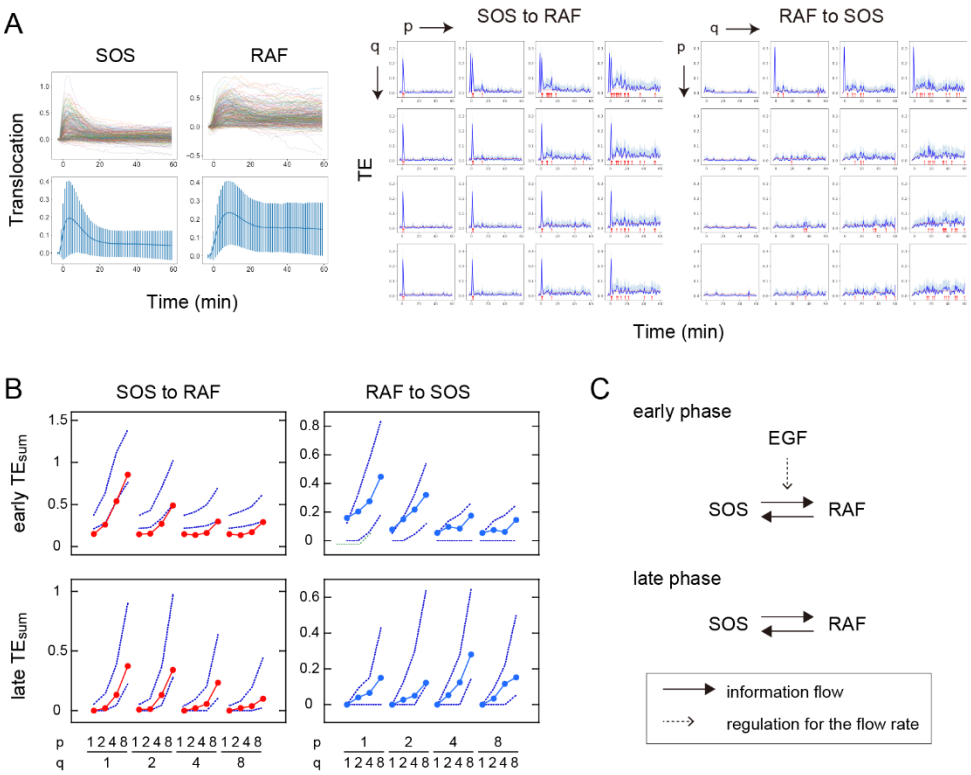
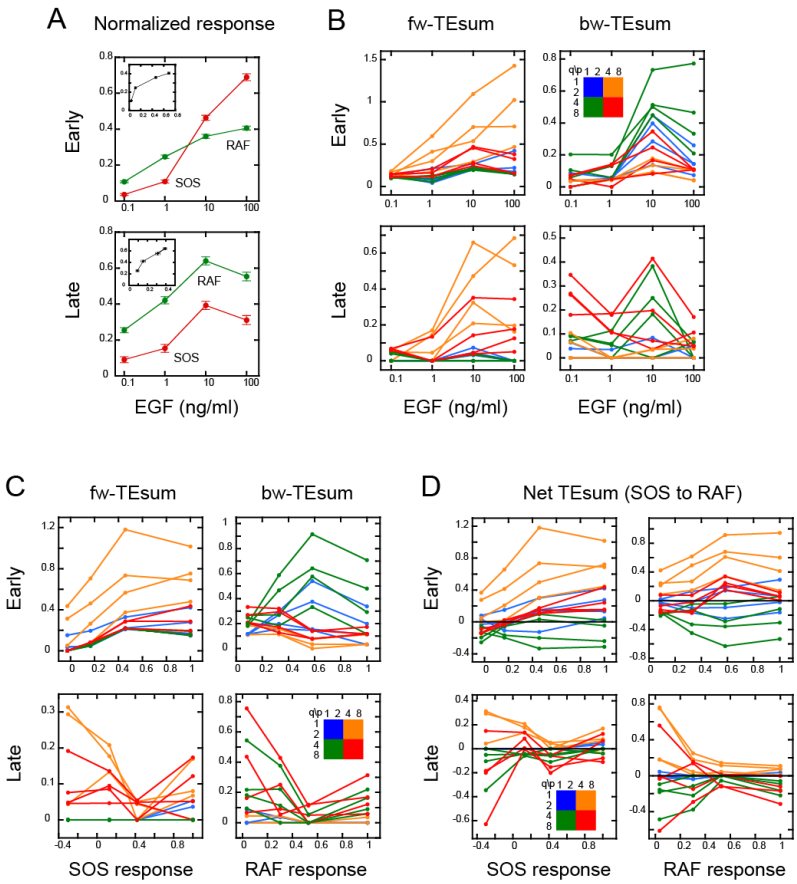


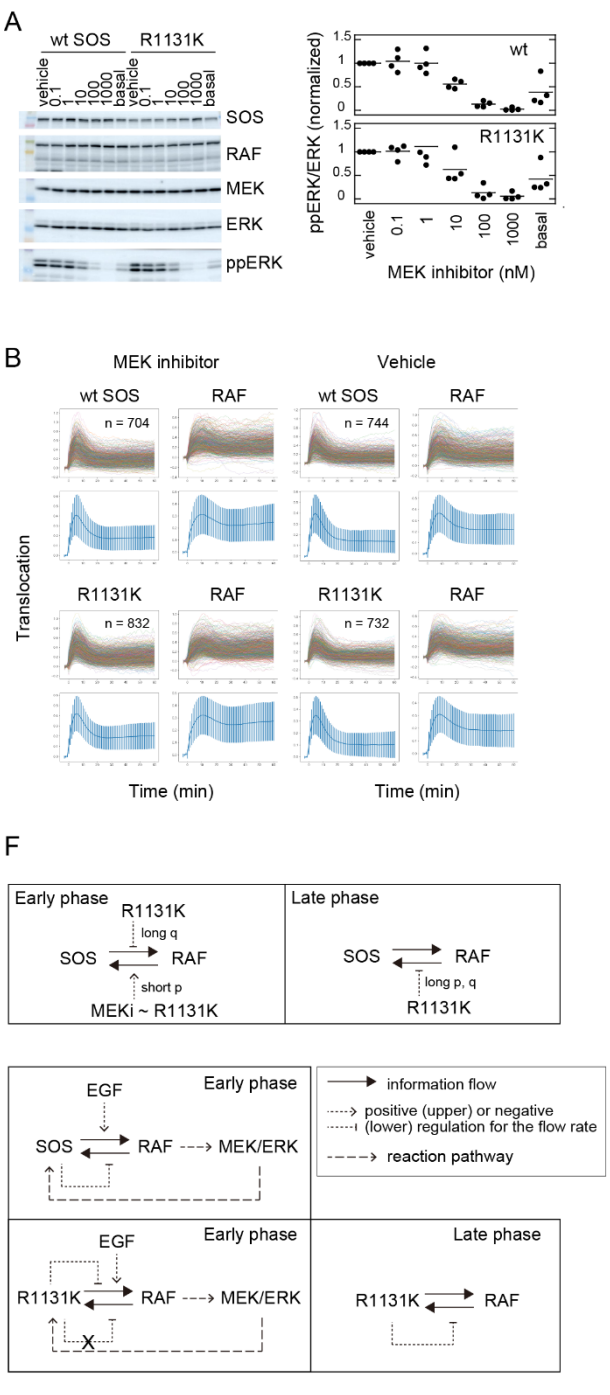
Figure 4. Response/transfer entropy relationships

A. EGF dose dependencies of the SOS (red) and RAF (green) responses. Response intensities were quantified as the sum of the cell-to-cell averaged response during 0-19 min (early) and 20-60 min (late), normalized to the maximum values of the total response during 0–60 min (Fig. 1E). Insets show the relationship between SOS and RAF responses. **B.** EGF dose dependence of the early and late fw- and bw-TEsum values. **C, D.** TEsum (C) and net TEsum (D) values are plotted as functions of the SOS and RAF responses. **B-D.** Color codes for p and q are indicated.



523 **Figure 5. MEK inhibitor and Noonan mutation effects**

524 **A**. Inhibition of ERK phosphorylation in cells treated with a MEK inhibitor, trametinib. Typical results
525 from Western blotting (left) and the inhibitor dose dependency in four independent experiments (right)
526 are shown. Bars indicate the averages. **B.** SOS and RAS responses in cells stimulated with 10 ng/ml
527 EGF. Cells were expressing wt (upper) or an R1131K mutant (lower) of SOS and pretreated with 10
528 nM MEK inhibitor (left) or vehicle solution (DMSO; right). n denotes the number of cells. **C.** TEsum
529 values in cells expressing wt SOS and exposed to MEK inhibitor (red) or vehicle solution (blue), and
530 their differences (black). **D.** TEsum values in cells expressing R1131K SOS and treated with MEK
531 inhibitor (red) or vehicle solution (blue). **E**. Differences in the TEsum values between R113K SOS
532 and wt SOS both without MEK inhibitor (upper). Differences in the TE_{sum} calculations between
533 R1131K SOS with and without MEK inhibitor (lower) are also shown. **C-E.** The average of the 100-
534 time bootstrapping analysis is shown with the 5-95% range. Bootstrapping P values of the differences
535 are *< 0.1, and **< 0.05. **F.** Effects of the MEK inhibitor and R1131K mutation (upper), and regulation
536 of the information flow networks in cells harboring wt and R1131K SOS (lower).
537


A
wt SOS
R1131K
vehicle
0.1
1
10
100
1000
basal
SOS
RAF
MEK
ERK
ppERK
ppERK/ERK (normalized)
wt
R1131K
MEK inhibitor (nM)
B
MEK inhibitor
Vehicle
wt SOS
RAF
n = 704
n = 744
R1131K
n = 832
n = 732
Translocation
Time (min)
F
Early phase
R1131K
long q
SOS
RAF
short p
MEKi ~ R1131K
Late phase
long p, q
EGF
MEK/ERK
information flow
positive (upper) or negative
(lower) regulation for the flow rate
reaction pathway
C
wt SOS to RAF
RAF to wt SOS
MEK inhibitor
vehicle
Early TEsum
Difference
Late TEsum
Lag time
D
R1131K to RAF
RAF to R1131K
E
SOS to RAF
RAF to SOS
R1131K-wt
R1131K - MEKi
Early
Late

538