Distributed and dynamic intracellular organization of extracellular information

Alejandro A. Granados1b,c, Julian M. J. Pietsch1b,c, Sarah A. Cepeda-Humerez2, Isabelail L. Farquhar1a,c, Gašper Tkačik4, and Peter S. Swain1b,c

*Department of Bioengineering, Imperial College London, London SW7 2AZ, United Kingdom; 1SynthSys–Synthetic and Systems Biology, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom; 2School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom; and 4Institute of Science and Technology Austria, 3400 Klosterneuburg, Austria

Although cells respond specifically to environments, how environmental identity is encoded intracellularly is not understood. Here, we study this organization of information in budding yeast by estimating the mutual information between environmental transitions and the dynamics of nuclear translocation for 10 transcription factors. Our method of estimation is general, scalable, and based on decoding from single cells. The dynamics of the transcription factors are necessary to encode the highest amounts of extracellular information, and we show that information is transduced through two channels: Generalists (Msn2/4, Tod6 and Dot6, Maf1, and Sfp1) can encode the nature of multiple stresses, but only if stress is high; specialists (Hog1, Yap1, and Mig1/2) encode one particular stress, but do so more quickly and for a wider range of magnitudes. In particular, Dot6 encodes almost as much information as Msn2, the master regulator of the environmental stress response. Each transcription factor reports differently, and it is only their collective behavior that distinguishes between multiple environmental states. Changes in the dynamics of the localization of transcription factors thus constitute a precise, distributed internal representation of extracellular change. We predict that such multidimensional representations are common in cellular decision-making.

Significance

To thrive in diverse environments, cells must represent extracellular change intracellularly despite stochastic biochemistry. Here, we introduce a quantitative framework for investigating the organization of information within a cell. Combining single-cell measurements of intracellular dynamics with a scalable methodology for estimating mutual information between time series and a discrete input, we demonstrate that extracellular information is encoded in the dynamics of the nuclear localization of transcription factors and that information is lost with alternative static statistics. Any one transcription factor is usually insufficient, but the collective dynamics of multiple transcription factors can represent complex extracellular change. We therefore show that a cell's internal representation of its environment can be both distributed across diverse proteins and dynamically encoded.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

1 A.A.G. and J.M.J.P. contributed equally to this work.

2 To whom correspondence should be addressed. Email: peter.swain@ed.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1716659115/-/DCSupplemental.

Published online May 21, 2018.
The average, and to avoid a priori assumptions about which
features of the response are relevant, such as its magnitude or duration.

Our method involves training a classifier to predict the state of the environment from the time series of as few as 100 cells. The classifier's output on the test data can then be used to estimate the mutual information (Fig. 1C). Formally, this approach provides a lower bound on the true information (SI Appendix), but, biologically, we are quantifying the information that a cell could plausibly recover and act upon after observing a single time series of its response. By varying the duration of the time series used by the classifier, we can determine how quickly cells accumulate information (Fig. 1D). In addition, errors made by the classifier indicate environments that are likely to be confused, giving insight into tasks potentially challenging for the cell. Although the final estimate of mutual information is determined by the signal-to-noise ratio of the data, this noise may not be all biological, again making our estimates a lower bound.

Mutual information is also determined by the choice of input distribution. Our focus is to use mutual information to quantify the signal-to-noise ratios in the single-cell time series, and therefore we choose a uniform distribution, which does not favor any one input. When we repeat the estimations using the input distribution that maximizes the mutual information, we see few changes (SI Appendix, Fig. S18). Indeed, in the true (natural) input distribution, stresses may not even occur at a time as we have assumed.

Detecting Environmental Change. Considering transitions into one of three environments, all of which reduce growth (Fig. 2A), the 10 transcription factors have diverse dynamics (Fig. 2B). The mutual information we calculate addresses whether an environmental transition can be detected from a typical time series of nuclear localization and is a number between 0 bits (indicating no detectable statistical differences between the dynamics of localization before and after the transition) and 1 bit (the dynamics of localization before and after the transition are distinct).

The glucose specialists Mig1/2 perform almost optimally in carbon stress with almost the maximum possible information of 1 bit (Fig. 2C). We note that different transcription factors encode information in different ways. For example, Dot6 and Sfp1 have dissimilar dynamics (Fig. 2B), yet encode the same amount of bits (Fig. 2C). Paralogs, however, do not necessarily carry equal information (compare Tod6 and Dot6).

Information can be encoded within minutes of the environmental transition with no trade-off. The speed of encoding typically increases the more information is encoded. Defining the encoding delay as the time for the mutual information to reach 50% of its maximum, information plateaus earliest in the time series of Mig1/2 (Fig. 2D). Fast responders are therefore typically more accurate, at least for such large stresses.

These general observations hold for transitions into osmotic and oxidative stress (Fig. 2D), establishing a hierarchy: In terms of the information and encoding delays, specialists (in blue) are followed by the environmental stress response (in pink), which in turn is followed by the others. The details of this hierarchy, however, are stress-specific, indicating that the dynamics of the transcription factors may encode not only the presence but also the nature of the environmental transition.

Detecting the Nature of Environmental Change. We therefore extend our method to calculate the mutual information between a single-cell time series and the four environmental states: rich medium (before the environmental transition) and the three stresses (after the transition). Not only do we estimate the mutual information (Fig. 3A), but we can also predict how a typical time series is likely to be classified as a function of
the duration of the environment (Fig. 3B and SI Appendix, Figs. S11 and S12).

Although no single transcription factor reaches the maximum of 2 bits, the time series of Msn2, Msn4, and, unexpectedly, Dot6, carry sufficient information to identify three environmental categories (for example, two environmental states and the remaining two states lumped together) (Fig. 3A). We observe, however, that the information is instead “spread” so that all environmental states are eventually classified with a >80% accuracy (Fig. 3B, Dot6). In contrast, an ideal specialist should perfectly discriminate one environmental state and lump together the remaining states to encode 0.8 bits (SI Appendix). Indeed, Hog1 and Yap1 do encode this much information (Fig. 3A), and their signaling networks operate nearly optimally in these high stresses. After only a few minutes, both specialists unequivocally identify their associated stress and never report false positives (Fig. 3B, Yap1).

Conditioning the mutual information on the identity of the environmental states delineates specialists from the other transcription factors (Fig. 3A, Inset, and SI Appendix, Fig. S15), which we term generalists because they encode information on multiple types of stress. Nevertheless, these groups are not mutually exclusive: Mig1 is not only a specialist for carbon stress, but also carries information on the other environmental states at late times, particularly osmotic stress, for which the probability of correctly identifying the environment is more than twice the 25% probability of a random choice (Fig. 3B).

**Detecting the Magnitude of Environmental Change.** In such high stresses, specialists appear unnecessary because the generalists identify stress so well, but this situation changes if we consider transitions into stresses of lower magnitude (Fig. 3C and SI Appendix, Fig. S10). From rich medium, we apply four different levels of the same type of stress and estimate the mutual information between the time series of translocation and the five environmental states.

Specialists now outperform generalists. Considering the mutual information between the time series and all pairs of the different levels of stress (Fig. 3D and SI Appendix, Fig. S13), we see that distinguishing between adjacent levels is most challenging, and generalists, but not specialists, can often only identify high stress.

Generalists and specialists also encode information differently: Generalists often use their entire time series, whereas specialists only do so to distinguish stresses of lower magnitude. By calculating the mutual information between summary features of the single-cell time series and the environmental state (SI Appendix, Figs. S16 and S17), we find that the amplitude of a specialist’s initial translocation can identify its associated stress if that stress is sufficiently severe (yellow dots in Fig. 3A), explaining specialists’ short encoding delays. For transitions into stresses of lower magnitude, however, information is encoded in the dynamics of the specialists’ response (yellow dots in Fig. 3C). Generalists can encode twice the amount of information in their time series compared with the highest information encoded by any single time point, and both the timing of their initial translocation, particularly for Msn2 and Dot6, and the shape of the times series can be important (SI Appendix).

---

**Fig. 2.** A hierarchy of information encoding, both in bits and encoding delays, holds across environmental transitions. (A) From population measurements (SI Appendix, Fig. S3), each new environment—carbon stress (0.1% glucose), osmotic stress (0.4 M NaCl), or oxidative stress (0.5 mM H$_2$O$_2$)—reduces growth compared with growth in rich medium (2% glucose). (B) For 10 transcription factors, we quantify nuclear localization across four environments using a step change from rich medium to stress at $t = 0$. The median and interquartile range are shown. (C) In response to carbon stress, the mutual information (MI) shows a hierarchy. The maximum possible information is 1 bit (dotted line). Mean and SD of two experiments per transcription factor are shown. (D) The hierarchy’s order is maintained for transitions into other stresses: Specialists (blue) encode the most information and are fastest, followed by the intermediate types of stress. Nevertheless, these groups are not mutually exclusive: Mig1 types of stress. Nevertheless, these groups are not mutually exclusive: Mig1 types of stress. Nevertheless, these groups are not mutually exclusive: Mig1 types of stress.
ATP through its regulation by AMP kinase (16). In contrast, the nitrogen (16). Similarly, Mig1 is sensitive to the levels of cellular cose (16), and TORC1 controls the response to the availability of glucose (16). Sch9 orchestrates the cell’s response to the availability of glucose (16), and TORC1 controls the response to the availability of nitrogen (16). Similarly, Mig1 is sensitive to the levels of cellular ATP through its regulation by AMP kinase (16). In contrast, the downstream kinase Sch9 (S6 kinase) (Fig. 4A). We quantify the redundancy between pairs of transcription factors to determine if the cell’s organization of information reflects the signaling network. If regulated by the same upstream signaling, two transcription factors may be completely redundant, so that when paired together, the amount of information does not increase above that from any one factor alone. By concatenating two time series (SI Appendix, Fig. S19), we can estimate the mutual information from simultaneously observing two transcription factors and so their redundancy (Fig. 4B).

For three of the four generalists considered, there is a substantial correlation between the amount of mutual information encoded and the severity of the stress (estimated by its reduction of growth compared with growth in rich medium) (Fig. 3E). This correlation may reflect that generalists are typically involved in regulating growth, through, for example, affecting translation (16). Specialists, in contrast, do not encode more information if their cognate stress is more severe (Fig. 3E).

**Generalists vs. Specialists.** To better understand the differences between generalists and specialists, we ask how the transcription factors are organized within the cell’s network of signal transduction. Using data on the substrates of kinases (22), we confirm (16) that the generalists are either directly or indirectly targets of protein kinase A (which has isoforms Tpk1-3) and TORC1 or its downstream kinase Sch9 (S6 kinase) (Fig. 4A). The generalists therefore do respond to the cell’s potential for growth; Protein kinase A orchestrates the cell’s response to the availability of glucose (16), and TORC1 controls the response to the availability of nitrogen (16). Similarly, Mig1 is sensitive to the levels of cellular ATP through its regulation by AMP kinase (16). In contrast, the specialists Hog1 and Yap1 are mostly embedded in their own signaling networks.
Collectively, transcription factors provide an internal representation of complex environments. (A) In the intracellular signaling network, generalists are colocated, and specialists are distinct. Edges between kinase and substrate are weighted by the evidence for that interaction (22). Clusters of highly interconnected components have different colors, and we include only the transcription factors in our study. (B) The redundancy in information between pairs of transcription factors reflects the structure of the signaling network in A. Edges are proportional to a pair’s redundancy, 1 – MI_{12}/(MI_{1} + MI_{2}) (23), and the size of each node is determined by the number of its edges. Redundancy was calculated from an average of two datasets (six experiments per transcription factor). (C) A generalist and a specialist typically encode the most information out of all possible pairs of transcription factors. For each pair, time series from one transcription factor were concatenated randomly with the time series of another to calculate the mutual information (MI). G, generalist; S, specialist. (D) Only through combinations of transcription factors can information on complex environments be encoded. Different colors represent different numbers of environmental states (SI Appendix): orange, seven states; red, five states; blue, four states; and purple, two states. Plotted lines are maxima calculated from all combinations of concatenated time series for a given number of transcription factors. (E) Cells transduce information through two types of channels—specialists and generalists—and four transcription factors are sufficient to encode all of the available information. The direct channels (specialists) respond to extracellular change; the indirect channels (generalists) respond to intracellular change and so to broader categories of extracellular change.

Discussion

In summary, we have shown that transcription factors can encode enough information in the dynamics of their nuclear translocations to unambiguously report an environmental change if that change is sufficiently large, that the nature of the change can also be encoded although with some degree of error, that how the information is encoded alters for changes of different magnitudes, and that no single transcription factor can accurately encode both the nature and magnitude of environmental change.

Information is transduced through two channels of specialists and generalists. Specialists are faster and can better identify a transition into their associated stress than generalists, but the variety of environments experienced by cells makes having a specialist for every environment implausible. We postulate that generalists avoid this constraint by providing an indirect channel that responds not to the extracellular signals sensed by specialists (25, 26), but to intracellular signals (27, 28), such as changes in cAMP, uncharged tRNAs, and the availability of amino acids (16) (Fig. 4E). By detecting physiological

partly redundant with Msn2/4 but not Dot6; Hog1 and Mig2 with Dot6 but not Msn2/4; and Mig1 is partly redundant with all three.

The redundancies imply that pairing a generalist with a specialist is best (Fig. 4C), and indeed such pairs typically encode the highest information (SI Appendix, Fig. S20). With its distinct signal transduction (Fig. 4A), a specialist can identify the environmental state that is most poorly distinguished by a generalist. For example, Msn2 is best paired with Mig2 (SI Appendix, Fig. S22).

As environments become more complex, multiple transcription factors are needed to generate an internal representation. Pooling the data to consider environments with different states (Fig. 4D), the maximum mutual information plateaus as the number of transcription factors increases, with four sufficing to generate ~95% of the information. This increase comes both from the distinct dynamics of the transcription factors (24), such as differences in timing (SI Appendix, Fig. S21), and from decreasing the effects of stochasticity by averaging the multiple readouts.
perturbations, generalists respond to broader ranges of stress (SI Appendix, Fig. S23) and are agnostic to the environment's precise nature. Generalists are therefore necessarily slower than specialists because they must wait for the environment to modify intracellular biochemistry. Indeed, we conjecture that the stochastic pulsing of the generalists in constant environments (19) is a response to spontaneous fluctuations in intracellular physiology.

Consistent with more recent interpretations (29), our data do not support a distinct environmental stress response controlled by Msn2/4, but show that aspects of the behavior of Msn2/4 are present in the dynamics of multiple transcription factors, such as Dot6, Sfp1, and Maf1. These latter factors act to determine rates of translation, consistent with the push–pull relationship between stress and growth (12). In particular, we have demonstrated that Dot6, although not Tod6, encodes the nature of environmental change in its dynamics to an accuracy almost comparable with Msn2/4, implying that Dot6 may play a similar role in cellular physiology. We find, too, that Mig1/2, although considered with Msn2/4, implying that Dot6 may play a similar role in cellular physiology. We find, too, that Mig1/2, although considered a glucose specialist, encodes information on osmotic stress and growth (12). In particular, we have demonstrated that Dot6, Sfp1, and Maf1. These latter factors act to determine rates of translation, consistent with the push–pull relationship between stress and growth (12). In particular, we have demonstrated that Dot6, Sfp1, and Maf1. These latter factors act to determine rates of translation, consistent with the push–pull relationship between stress and growth (12).

Finally, our results show that it is only through the collective dynamics of multiple transcription factors that cells can encode sufficient information to generate specific downstream responses (15, 30). Paralleling discoveries in neuroscience, we expect that such multidimensional internal representations are widespread within cellular biology and that their failures in encoding information, by causing dysfunctional decision-making, instigate deleterious behaviors and disease (31).

Materials and Methods

Time-Lapse Microscopy. BY strains with fluorescently tagged transcription factors (32) are grown in ALCATRAS microfluidic devices (20) in synthetic complete medium with 2% glucose for at least 3 h and then exposed to stress for 5 h (SI Appendix). Switching between media is via an external mixer and syringe pumps. Bright-field and fluorescence images are acquired every 2.5 min, and cells are segmented by using the DISCO algorithm (33).

Estimating the Mutual Information. Our algorithm involves (SI Appendix): (i) using principal component analysis to delineate a basis for the time series; (ii) training a linear support vector machine to classify time series in this basis; (iii) calculating a confusion matrix using the test data; and (iv) estimating (a lower bound on) the mutual information by interpreting the entries of the confusion matrix as conditional probabilities.

ACKNOWLEDGMENTS. We thank S. Jaramillo-Rivero, P. Thomas, M. Voliotis, E. Wallace, and the P.S.S. laboratory for critical comments and Reiko Tanaka for her support and advice (A.A.G.). This work was supported by the Biotechnology and Biological Sciences Research Council (J.M.J.P., I.F., and P.S.S.), the Engineering and Physical Sciences Research Council (EPSRC) (A.A.G.), and Austrian Science Fund Grant FWF P28844 (to G.T.).