



Distributed and dynamic intracellular organization of extracellular information

Alejandro A. Granados^{a,b,1}, Julian M. J. Pietsch^{b,c,1}, Sarah A. Cepeda-Humerez^d, Iseabail L. Farquhar^{b,c}, Gašper Tkačič^d, and Peter S. Swain^{b,c,2}

^aDepartment of Bioengineering, Imperial College London, London SW7 2AZ, United Kingdom; ^bSynthSys–Synthetic and Systems Biology, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom; ^cSchool of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom; and ^dInstitute of Science and Technology Austria, 3400 Klosterneuburg, Austria

Edited by Eric D. Siggia, The Rockefeller University, New York, NY, and approved April 27, 2018 (received for review September 21, 2017)

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.

cell signaling | mutual information | time series | transcription factors | stress

All organisms sense their environment and internally represent the information gained to elicit a change in behavior (1). Much is understood about such representations in neural systems (2), but single cells must perform an analogous task (1, 3), encoding intracellularly the information about extracellular environments, and yet little is known about the nature of their encoding.

The activation of transcription factors is thought to provide an internal representation of a cell's environment (4–10), but how information is encoded dynamically, whether information is spread across multiple factors, and how information is read downstream all remain unclear (Fig. 1A). We do know that the biochemical implementation of such representations is likely to be stochastic (11) and that the same biochemistry may be used to sense disparate environments. Furthermore, cells typically have just “one shot” at mounting the appropriate response from these internal representations, with competition being unforgiving for those that delay, at least among microbes (12–14). Here, we use information theory to investigate how eukaryotic cells answer these challenges.

To do so, we turn to budding yeast and to environmental changes for which we expect information encoding to be key: stresses that compromise growth and evoke adaptive gene expression (15). In yeast, extracellular changes are sensed by signaling networks that regulate the activity of transcription factors, often by their translocation either into or out of the nucleus (16), analogous to p53 and NF- κ B in mammalian cells (17, 18). We therefore consider the movement of these transcription factors

as a cell's internal representation of an environmental transition. The translocations are dynamic and stochastic, and the information available from the full time series of the response could be substantially higher than that available from any temporal snapshot (9) (Fig. 1A).

Tens of transcription factors translocate in yeast (16), and we focus on a representative subset: Msn2 and its paralog Msn4, which drive the environmental stress response; Mig1 and its paralog Mig2, which respond to low glucose; Hog1 (a kinase), which responds to hyperosmotic stress; Yap1, which responds to oxidative stress; Sfp1, which promotes, and Dot6 and its paralog Tod6, which repress the biogenesis of ribosomes; and Maf1, which represses the synthesis of tRNAs. We include Dot6 and Tod6, which are little studied, to determine if our approach can help determine their physiological importance. Some of these factors (Msn2/4, Mig1/2, and Dot6/Tod6) have pulsatile dynamics, with stochastic bursts of nuclear localization even without stress (19).

We consider environmental shifts from rich medium (2% glucose) into carbon stress (0.1% glucose), hyperosmotic, or oxidative stress. Using fluorescent tagging and microfluidics (20), we measure the degree of nuclear localization of the transcription factors in hundreds of single cells both before and after the stress is applied (Fig. 1B).

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.

Author contributions: A.A.G., J.M.J.P., and P.S.S. designed the study; S.A.C.-H. and G.T. developed the estimation algorithm; A.A.G., J.M.J.P., and I.L.F. performed the experiments; A.A.G., J.M.J.P., S.A.C.-H., and P.S.S. analyzed the data; and A.A.G., J.M.J.P., S.A.C.-H., I.L.F., G.T., and P.S.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

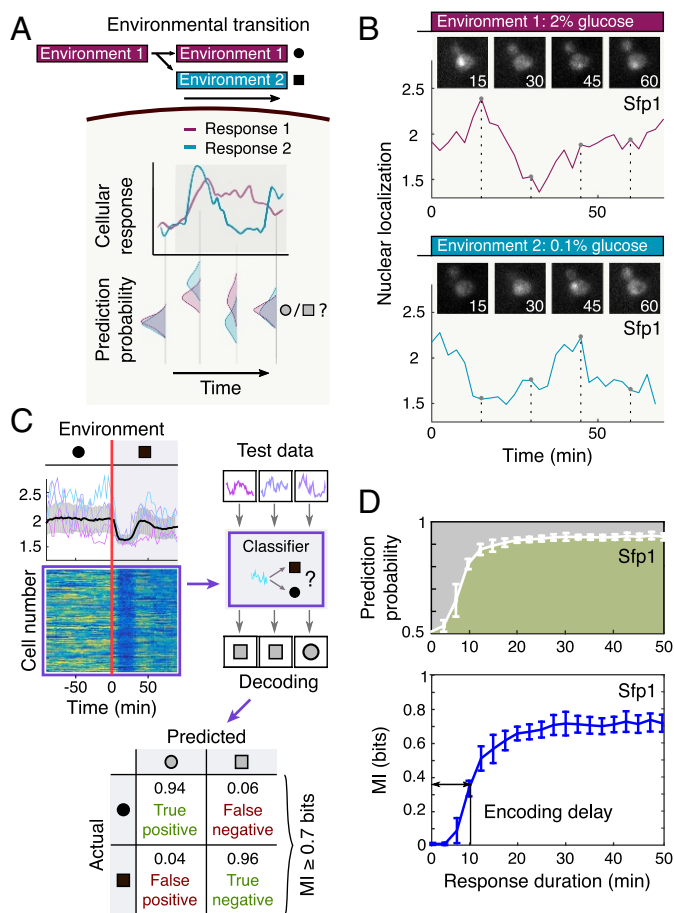
Published under the PNAS license.

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

²To 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.



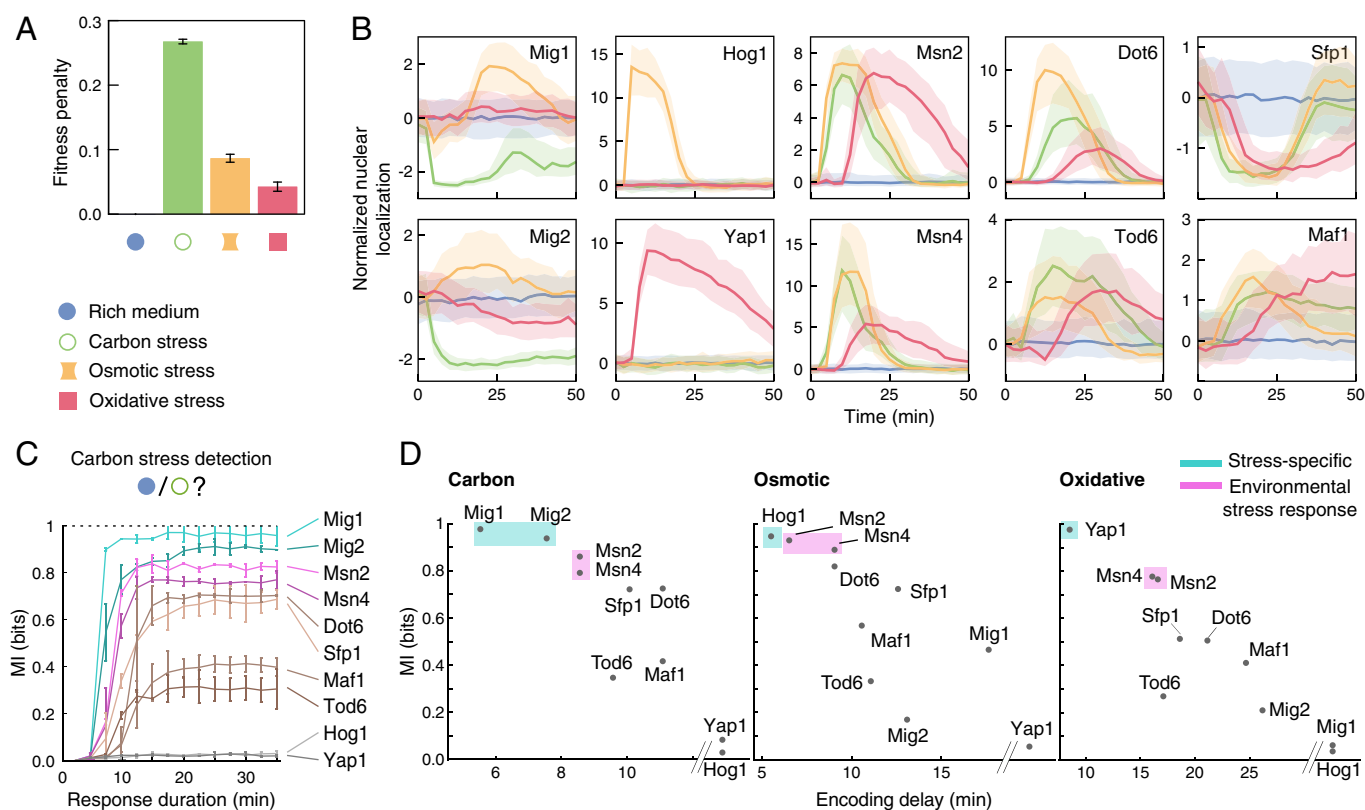


Fig. 2. A hierarchy of information encoding, both in bits and encoding delays, holds across environmental transitions. (A) From population measurements (*SI Appendix, Fig. S3*), each new environment—carbon stress (0.1% glucose), osmotic stress (0.4 M NaCl), or oxidative stress (0.5 mM H_2O_2)—reduces 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 environmental stress response (pink).

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).

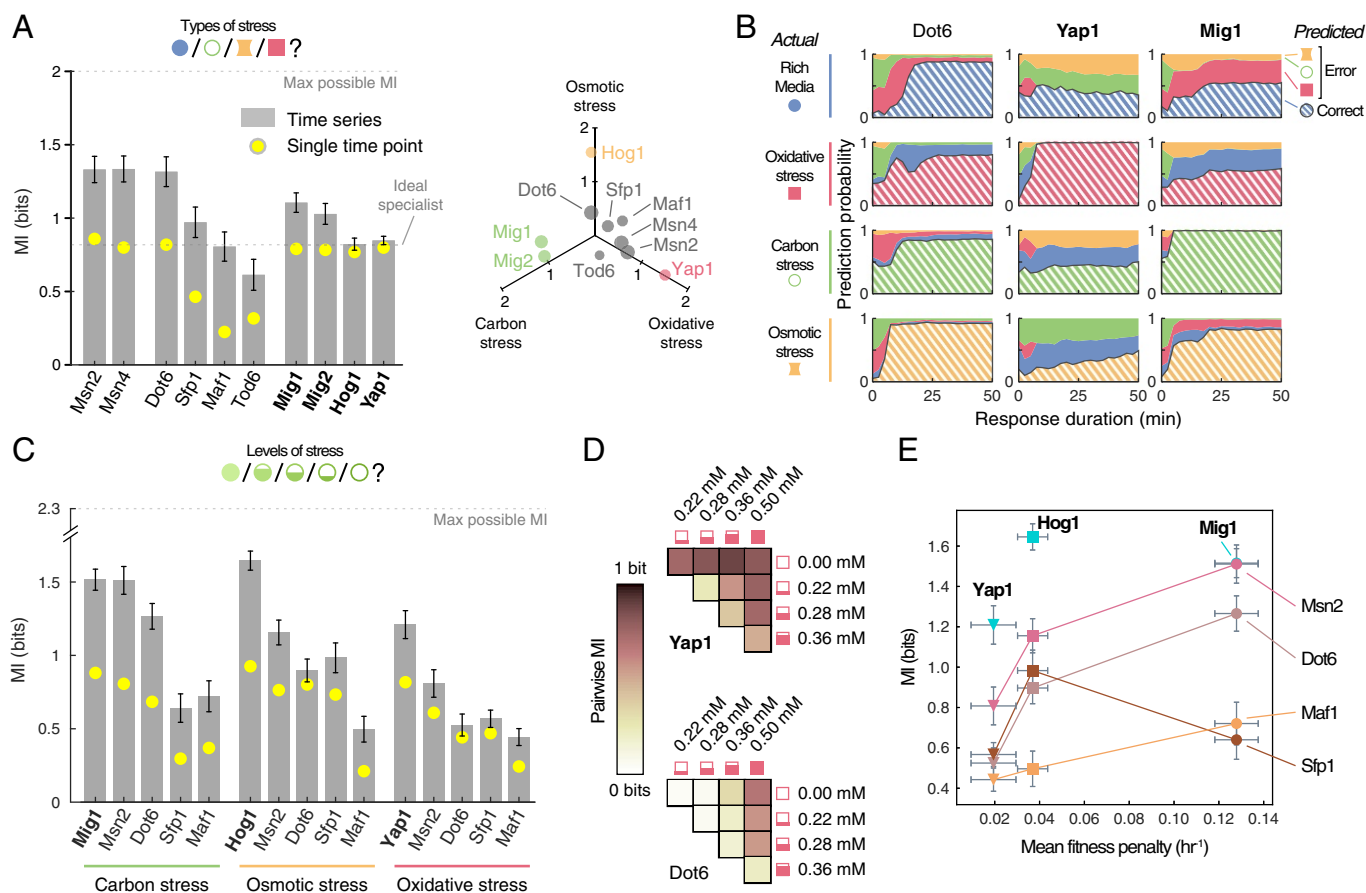
Conditioning the mutual information on the identity of the environmental states delineates specialists from the other transcription factors (Fig. 3A, *Inset*, and *SI Appendix, Fig. S15*), which we term generalists because they encode information on multiple types of stress. Nevertheless, these groups 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*).



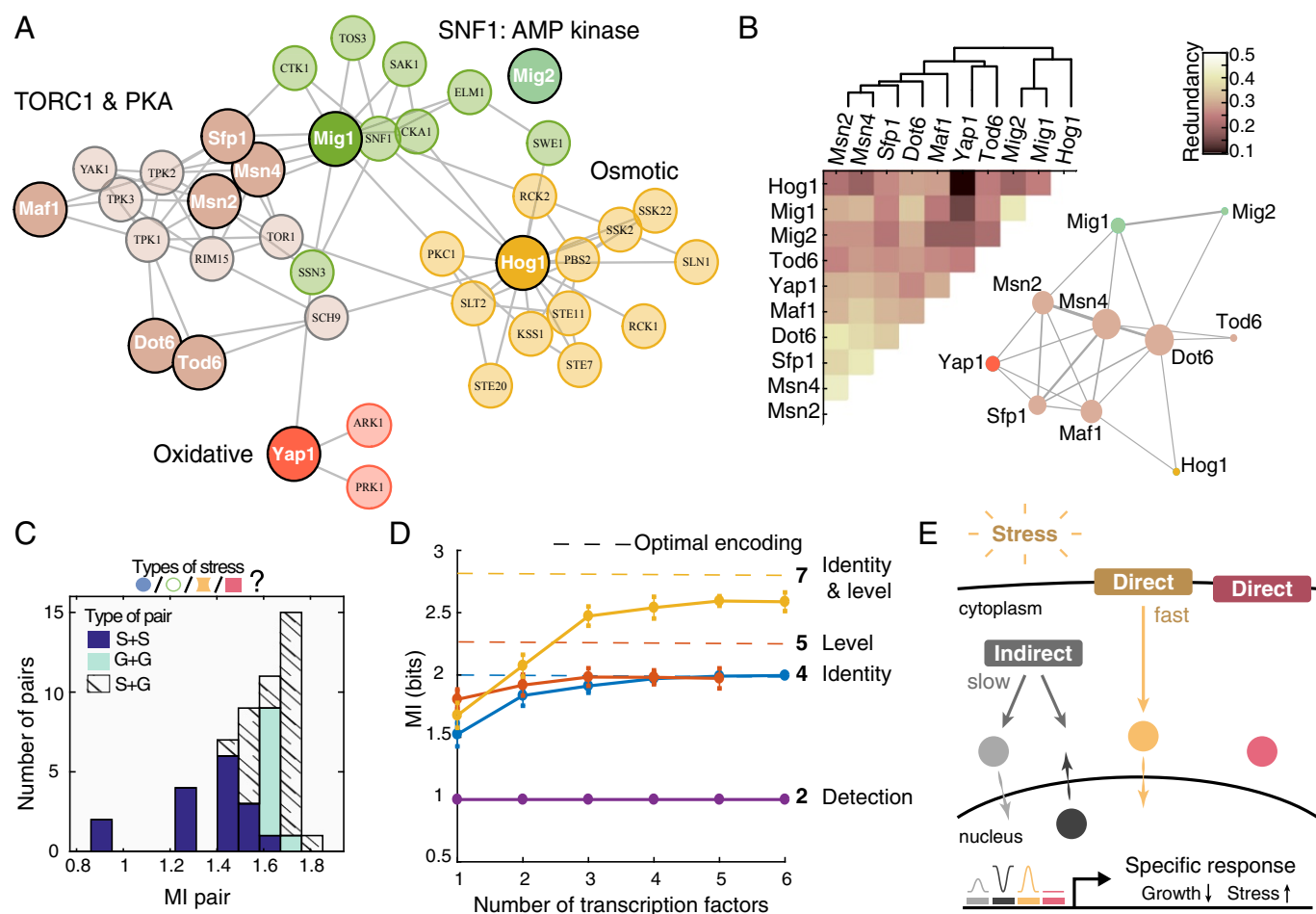


Fig. 4. Collectively, transcription factors provide an internal representation of complex environments. (A) In the intracellular signaling network, generalists are colocated, and specialists are distinct. Edges between kinase and substrate are weighted by the evidence for that interaction (22). Clusters of highly interconnected components have different colors, and we include only the transcription factors in our study. (B) The redundancy in information 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.

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 numbers of transcription factors increase, 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.

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

