Supporting Information: Distributed and dynamic intracellular organization of extracellular information

Alejandro A. Granados[1,2,⋆], Julian M. J. Pietsch[2,4,⋆], Sarah A. Cepeda-Humerez[3], Iseabail L. Farquhar[2,4], Gašper Tkačik[3], and Peter S. Swain[2,4]

[1]Department of Bioengineering, Imperial College London, London SW7 2AZ, United Kingdom
[2]SynthSys–Synthetic and Systems Biology, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom
[3]Institute of Science and Technology Austria, 3400 Klosterneuburg, Austria
[4]School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom
[⋆]These authors contributed equally to this work.

Contents

www.pnas.org/cgi/doi/10.1073/pnas.1716659115

List of Figures

1 Materials and Methods

1.1 Strains and Media

Strains containing fluorescently labeled transcription factors that are known to translocate (Hog1 [1], Yap1 [2], Msn2/4 [3], Sfp1 [4], Maf1 [5], Tod6 and Dot6 [6]) were acquired from the *S. cerevisiae* GFP collection [7] and validated by sequencing. Because of the poor growth of the GFP collection's Mig1-GFP strain in our experimental conditions, we generated an equivalent strain (also validated by sequencing) in BY4741 using PCR-based genomic integration of yEGFP-His from pKT0128 (AddGene). Strains indicated as having a nuclear marker were derived from the GFP-labeled transcription factor strains by PCR-based genomic integration of mCherry-Kan^R from pBS34 (EUROSCARF) to tag the chromatin-associated Nhp6A protein. Media used for propagation and growth was standard synthetic complete (SC) medium supplemented with 2% glucose (unless specified otherwise). For stress media, we added sodium chloride or hydrogen peroxide as indicated; carbon stress is SC medium supplemented with lower concentrations of glucose (0.1% unless specified otherwise). Cells were grown at 30°C.

1.2 Microscopy and microfluidics

1.2.1 Cell preparation and loading ALCATRAS

Overnight cultures were inoculated such that cells would reach mid-log phase by the following morning. Cells were diluted in fresh medium to OD_{600} 0.1 and incubated an additional 3–4 hours for loading into microfluidics devices at OD_{600} 0.3–0.4. To expose multiple strains to the same environmental conditions and to optimize data acquisition, we used a multi-chamber version of ALCATRAS[8,9], which allowed five different strains to be loaded into distinct chambers but still be exposed to the same extracellular media (Fig. S1). Polydimethylsiloxane (PDMS) barriers between strains ensure that there is no cross-contamination during loading. The ALCATRAS chambers were pre-filled with growth medium with added 0.05% bovine serum albumin (BSA) to facilitate cell-loading and reduce the formation of clumps of cells.

The microfluidic chamber and the syringe pumps containing the extracellular media were located inside an incubation chamber (Okolabs) that maintained a constant temperature of 30°C. We used a 60× 1.4 NA oil immersion objective (Nikon). The Nikon Perfect Focus System (PFS) ensured consistent focus over the experiment. Fluorescence imaging was performed with an OptoLED light source (Cairn Research). Images were acquired using an Evolve 512 EMCCD (Photometrics), and we correct for a delay of a few seconds in image acquisition as the microscope moves between different positions.

1.2.2 Changing the extracellular environment

Syringe pumps (Aladdin NE-1002X) were used to supply media at 4 μL/min to the device, and an external mixer (a sterile metal T-junction) allowed switching between different media. Cells were grown with imaging for 5 hours in rich medium (3 hours for transitions to oxidative stress) before setting the pumps to supply stress medium for growth for a further 5 hours. Sharp switches were achieved by programming the pumps to flush 450 μL of stress medium through the mixer at 150 μL/min. Following the switch, the first pump continued to withdraw at 0.5 μL/min with the second pump supplying stress medium at 4.5 μL/min to maintain an overall rate of 4 μL/min to the device. The first pump supplied rich media in all experiments. For the stress-type experiments, the second pump supplied medium either containing 0.1% glucose (carbon stress), supplemented with 0.4 M sodium chloride (osmotic stress), or supplemented with 0.5 mM hydrogen peroxide (oxidative stress). For

the stress-level experiments, the second pump supplied medium containing 0.8%, 0.4%, 0.2% or 0.1% glucose (carbon stress); medium supplemented with either 0.1 M, 0.2 M, 0.3 M or 0.4 M sodium chloride; or medium supplemented with either 0.22 mM, 0.28 mM, 0.36 mM or 0.50 mM hydrogen peroxide. The dye Cy5 was added to the syringe containing stress medium to monitor its arrival time and the sharpness of the switch.

1.2.3 Image segmentation and quantification of nuclear localization

During each experiment, we acquired bright-field and fluorescence images at five z-sections spaced 0.6 micron apart. The maximum projection of these images (the maximum pixel values across all z-sections) was used for quantification. The Cy5 channel was acquired in a single focal plane. Cells were segmented from bright-field images using the DISCO algorithm [10,11], which identifies cell centres with a support vector machine (SVM) and cell edges with an active-contour-based method applied to the bright-field z sections.

We quantified the nuclear accumulation of transcription factors by calculating the ratio between the average of the five brightest pixels within the cell and the median fluorescence of the whole cell. This widely-adopted measure [12–14] is robust to systematic multiplicative variations in image intensity between time-points (a small but not insignificant source of experimental error). We validated its use for transcription factors where this measure has not been applied before (Dot6, Tod6, Sfp1 and Maf1) and also for Mig1 where it has [14] (as a positive control) by adding a nuclear marker (Nhp6A-mCherry) (Fig. S2). The ratio of nuclear and cytoplasmic fluorescence correlated ($R^2 = 0.92$ over combined data) with the measure obtained without a nuclear marker, and the choice of measure did not affect our calculations of mutual information (Fig. S5). Therefore, we performed all other experiments without a nuclear marker to reduce the time of image acquisition and so increase the number of cells monitored.

1.2.4 Numbers of cells

With our multi-strain experimental set-up (Fig. S1), we can monitor on average 150 cells per strain (5 strains per assay) at a sampling rate of 2.5 minutes. Biological replicates (identical experimental conditions applied to different cultures of the same strain) were performed on different days. Table 1 gives the number of cells for the stress-type experiments (Fig. 2 & 3A); Table 2 gives the number of cells for the stress-level experiments (Fig. 3C); and Table 3 for the transition from rich media to rich media (Fig. S8.)

Table 1. Number of cells in each of the stress-type experiments

Replicate 1

	Msn2	Msn4	Dot6	Tod6	Sfp1	Maf1	Mig1	Mig2	Hog1	Yap1
Oxidative stress	162	183	184	163	171	159	207	219	205	237
Carbon stress	178	202	198	132	164	219	199	191	190	215
Osmotic stress	208	209	212	215	213	191	206	186	211	202

Replicate 2

	Msn2	Msn4	Dot6	Tod6	Sfp1	Maf1	Mig1	Mig2	Hog1	Yap1
Oxidative stress	166	166	186	172	208	183	217	209	308	131
Carbon stress	117	211	219	216	155	208	183	223	97	127
Osmotic stress	226	195	107	142	154	190	210	198	206	124

1.3 Plate reader experiments

Fitness penalties were estimated using plate readers (Infinity M200, Tecan Group Ltd., Switzerland) and an established protocol [15]. Cells were cultured for 16 hours in SC with

Table 2. Number of cells in each of the stress-level experiments

	Mig1	Msn2	Dot6	Sfp1	Maf1		Hog1	Msn2	Dot6	Sfp1	Maf1
0.8% glucose	181	177	182	161	197	0.1 M NaCl	183	160	194	159	185
0.4% glucose	209	177	162	121	233	0.2 M NaCl	196	172	169	107	200
0.2% glucose	203	195	204	212	214	0.3 M NaCl	184	198	180	185	77
0.1% glucose	204	185	195	178	120	0.4 M NaCl	210	230	224	199	143

	Yap1	Msn2	Dot6	Sfp1	Maf1
0.22 mM H_2O_2	193	171	218	194	188
0.28 mM H_2O_2	185	109	206	189	181
0.36 mM H_2O_2	149	199	150	249	191
0.50 mM H_2O_2	270	186	186	144	182

Table 3. Number of cells in each of the rich-to-rich transition experiments

Msn2	Msn4	Dot6	Tod6	Sfp1	Maf1	Mig1	Mig2	Hog1	Yap1
163	169	156	184	187	153	163	94	133	85

2% glucose at 30°C with ODs maintained below 0.6, diluted to an OD of 0.05 with SC in 2% glucose and incubated for a further 5 hours, before being spun down, washed in the appropriate stress media, spun down again, resuspended in stress conditions in a 96-well plate and placed in the plate reader.

1.4 Data availability

Data are freely available at `http://dx.doi.org/10.7488/ds/2214`

2 Estimating fitness penalties

We estimate fitness penalties for the different stresses using bulk measurements made in plate readers. OD measurements are corrected for non-linearities in the relationship between OD and numbers of cells [15,16] and by the measured OD of wells containing only media. For each well, the plate reader returns a growth curve of the OD versus time of a growing population of cells. From these growth curves, we use a Gaussian process with a Matern covariance function to estimate the growth rate [17].

The fitness penalties are defined in comparison to growth in 2% glucose (denoted c for control). We first use the growth curve and the growth rate as a function of time to write the growth rate, λ, as a function of OD: $\lambda = \lambda(o)$. Then the fitness penalty, Δw, is defined as

$$\Delta w = \int_{o_{\min}}^{o_{\max}} do \, \frac{\lambda_c(o) - \lambda(o)}{o_{\max} - o_{\min}} \tag{1}$$

where $o_{\min}$ and $o_{\max}$ define the range of ODs shared between the two strains (the control in 2% glucose and the test strain in stress) (Fig. S3A). Conditions that alter mean growth rates (averaged over ODs) substantially from the mean growth rate in 2% glucose therefore impose high fitness penalties (Fig. S3B). Integrating over OD rather than over time means that the growth curve is used to calculate fitness rather than only its end points: $\int dt \, \lambda(t) = \int dt \, \frac{d}{dt} \log o = \log(o_{\max}/o_{\min})$.

3 Estimating the mutual information

3.1 Data preparation

We normalized the single-cell time-series of nuclear localization using the population's mean (μ) and standard deviation (σ) across the 20 time-points (50 min) before the stress was introduced from all cells in a single microfluidics experiment. For each time-series, we applied a z-normalization: for cell i: $x_i(t) \rightarrow (x_i(t) - \mu)/\sigma$. After normalization, the steady-state (pre-shock) nuclear localization across the population of cells then has a mean of approximately zero and a standard deviation of approximately one.

This normalization scales the time-series, but conserves cell-to-cell variation. The normalization was applied for consistency in calculations where classification of time-series comprised multiple experimental conditions (i.e. for stress-type, Fig. 3A, and stress-level, Fig. 3C, classifications).

The exact time-point at which the transition takes place ($t = 0$) is detected using the signal from the Cy5 dye. For each experiment, we partition the time-series at $t = 0$ to generate two environmental states: the 20 time-points before the transition are labeled as *rich media* data and the 20 time-points after the transition are labeled as *stress* data. For experiments addressing the presence of environmental change (Fig. 2C & D), we train and test the classifier using these two environmental states. For environmental transitions comprising more than two new environmental states (Figs. 3 & 4), we merged the rich media segments of each individual experiment into a single matrix of cells that is then randomly sampled.

3.2 Estimation of mutual information through decoding

Let data $\mathcal{D} = \{(\mathbf{x}'_1, y_1), (\mathbf{x}'_2, y_2), \ldots, (\mathbf{x}'_n, y_n)\}$ consist of $i = 1, \ldots, n$ pairs of time-series of nuclear localization in individual cells, $\mathbf{x}'_i$, and their corresponding environmental labels, y_i. Time-series are sampled at uniform time intervals and are represented as $t = 20$ dimensional vectors, normalized as described above by subtracting the pre-stress mean and dividing by the pre-stress standard deviation. The labels represent the state of the external environment (the input signal) and can take on q discrete values: $y_i \in \mathcal{Y} = \{s_0, s_1, \ldots, s_{q-1}\}$. For Fig. 1C, for example, a total of $n = 210$ time-series of nuclear localization for Sfp1 are collected and labeled either with s_0, denoting rich medium (2% glucose), or with s_1, denoting low (0.1%) glucose.

Our goal is to estimate the mutual information, $\mathrm{MI}(\mathbf{x}'; y)$, between the complete, discretely-sampled time-series, $\mathbf{x}'$, and a discrete environment, y. Given the environment, the t-dimensional conditional distribution $P(\mathbf{x}'|y)$ describes the probability of observing a particular time-series given the environmental condition, and our individual cell time-series, $\mathbf{x}'_i$, are independent and identically distributed samples from that distribution.

For an input distribution $P(y)$, recall that the mutual information between random variables y and $\mathbf{x}'$ is defined as follows [18]:

$$\mathrm{MI}(\mathbf{x}'; y) = \sum_{y \in \mathcal{Y}} P(y) \int d\mathbf{x}' P(\mathbf{x}'|y) \log_2 \frac{P(\mathbf{x}'|y)}{P_x(\mathbf{x}')}, \tag{2}$$

where $P(\mathbf{x}') = \sum_y P(y)P(\mathbf{x}'|y)$ follows by marginalization from the known $P(y)$ and $P(\mathbf{x}'|y)$.

A direct evaluation of Eq. 2 is intractable because the high dimensionality of $P(\mathbf{x}'|y)$ prevents us from robustly sampling this distribution despite the hundreds of time-series samples at our disposal. To make the problem tractable, one can choose a parametric form for $P(\mathbf{x}'|y)$ (e.g., by assuming it is a multivariate normal distribution, which would require only sampling the mean time-series and the covariance matrix), use a binless estimator

(e.g., kNN [19] as adopted by Voliotis et al. [20] and Selimkhanov et al. [21], which, however, makes assumptions on the metric of the space of time-series and still typically requires potentially prohibitive numbers of time-series), or use the decoding methodology outlined below, which traces its roots to early applications in neuroscience [22,23].

The central idea of the decoding-based estimator of mutual information is to construct a classifier function that operates on individual time-series of nuclear localization and outputs an estimate of the environmental state, i.e., to look for a function $f : \mathbf{x}' \to \hat{y}$ that 'decodes' the responses. Then, the random variables y, $\mathbf{x}'$ and the state estimated from the nuclear localization time-series, $\hat{y}$, form a Markov chain:

$$y \xrightarrow{P(\mathbf{x}'|y)} \mathbf{x}' \xrightarrow{f} \hat{y}, \tag{3}$$

where the first mapping is stochastic, given by $P(\mathbf{x}'|y)$, and the last mapping through the classifier function f is deterministic. The data processing inequality [18] then guarantees that

$$\mathrm{MI}(\mathbf{x}'; y) \geq \mathrm{MI}(\hat{y}; y). \tag{4}$$

If we can tractably estimate the mutual information between the true and decoded environmental states, $\mathrm{MI}(\hat{y}; y)$, at the right-hand side of Eq. 4, from data, we will find a lower bound on the mutual information of interest: $\mathrm{MI}(\mathbf{x}'; y)$. This relationship is true for any f, but classifiers with higher performance make the bound tighter. The decoding estimate also has a biological interpretation as the information that can be extracted on a single-shot basis (i.e., after receiving only one time-series) by the explicit decoding procedure given by f. This decoding stands in contrast to $\mathrm{MI}(\mathbf{x}'; y)$, which offers no explicit procedure for extracting the identity of the environment from a time-series and even allows a biologically implausible block code (several environmental changes and localization responses are observed in succession before the cells would jointly decode a sequence of appropriate responses).

Note that the input distribution over possible environmental states puts a theoretical upper bound on the possible information of $I^* = H(y)$, where $H(y)$ is the entropy of y and $I^* = \log_2 q$ bits for a uniform distribution with q environmental states. This maximum can only be achieved when the classifier function f can perfectly (i.e., without any misclassification errors on the testing data, $\hat{y} = y$) identify each environmental state solely from the time-series $\mathbf{x}'$.

An advantage of using the decoding estimator is its tractability. The classifier function maps responses back from the $t = 20$ dimensional space to the discrete space of environmental states, $\mathcal{Y}$. To obtain $\mathrm{MI}(\hat{y}; y)$ we should then compute, in analogy to Eq. 2:

$$\mathrm{MI}(\hat{y}; y) = \sum_{y \in \mathcal{Y}} P_y(y) \sum_{\hat{y} \in \mathcal{Y}} P(\hat{y}|y) \log_2 \frac{P(\hat{y}|y)}{P_{\hat{y}}(\hat{y})}. \tag{5}$$

All the distributions involved are discrete and low dimensional and thus can be empirically well estimated from the data $\mathcal{D}$ on $n \sim O(100)$ individual time-series (e.g., the distribution $P(\hat{y}|y)$ is a $q \times q$ table, with q typically being small: $2 \leq q \leq 7$ for our experiments).

To construct an estimator, we replace in Eq. 5 the distribution $P(\hat{y}|y)$ with its empirical estimate, $M(\hat{y}, y)$ that we evaluate (over test data, as explained below in detail) from data $\mathcal{D}$. The matrix M is simply the confusion matrix of the classifier, i.e., the empirical probability that the classifier outputs environmental state estimate $\hat{y}$ given that the true state was y. Formally, the fraction of time-series classified as state i given that the actual state j is:

$$M(\hat{y} = s_i, y = s_j) = P(y = s_j) \times \sum_{(\mathbf{x}_r, y_r) \in \mathcal{D}_{\mathrm{test}}} \frac{\mathbb{I}\left[f(\mathbf{x}_r) = s_i\right]}{\mathcal{D}_{\mathrm{test}}} \times \mathbb{I}\left[y_r = s_j\right], \tag{6}$$

where: f is the classifier function; $\mathcal{D}_{\text{test}} \subset \mathcal{D}$ are the samples of test data and $|\mathcal{D}_{\text{test}}|$ is their number (these samples have not been used to train the classifier f); and $\mathbb{I}$ is an indicator function, which is 1 when its argument is true and 0 otherwise.

Given the data $\mathcal{D}$ and the trained classifier f, which can assign the estimated state $\hat{y}$ for each time-series, we can compute the confusion matrix in Eq. 6 by iterating across testing data and then use that matrix to evaluate MI in Eq. 5 by direct summation.

In detail, our algorithm proceeds as follows:

1. Split the complete dataset $\mathcal{D}$ into disjoint training data, $\mathcal{D}_{\text{train}} \subset \mathcal{D}$, comprising 70-80% of the samples, and testing data, $\mathcal{D}_{\text{test}} \subset \mathcal{D}$ containing the rest of the samples.

2. Train classifier on training data, $\mathcal{D}_{\text{train}} = (\mathbf{x}'_i, y_i)$:

   (a) Perform principle component analysis on the time-series $\mathbf{x}'$ in $\mathcal{D}_{\text{train}}$ across all conditions to identify the top k principal vectors that span most of the variance. In general, we use $k = 5$ to preserve $\geq 80\%$ of the variance, which holds across all transcription factors probed in the stress-type experiments (Fig S4A). As necessary, however, we choose k such that mutual information is maximized on test data on a case-by-case basis.

   (b) Project time-series $\mathbf{x}'$ into the k-dimensional space, $\mathbf{x}$, for all time-series in $\mathcal{D}$.

   (c) Train a linear Support Vector Machine (SVM) classifier f on $\mathbf{x}$ for the time series in $\mathcal{D}_{\text{train}}$. We used the LIBSVM implementation with the 'one-against-one' approach to extend to more than two classes [24].

3. Apply f on projected testing data $\mathbf{x}$ for the time series in $\mathcal{D}_{\text{test}}$, to evaluate the confusion matrix in Eq (6) and so estimate MI via Eq (5). For our data ($n \sim 100$), the estimates of information are stable with respect to the number of observed time-series (Fig. S4C). In general, however, the dependence of the estimate with the number of samples should be determined and de-biasing performed if needed [25].

4. Use bootstrapping by repeating the procedure from step 1 to 3 after randomly re-sampling the test and training sets. The procedure is repeated 100 times (unless otherwise specified) to get the mean of the estimator and its standard deviation, which we use a measure of error.

We note that the PCA step was motivated by data with time-series of different length for different transcription factors and is not in principal necessary for our data when all time-series have equal numbers of data points measured with the same time interval. Nevertheless, for cases where we explore different encoding delays through changing the duration (dimensionality t) of the time-series, $\mathbf{x}'$, and for cases where we decode from multiple transcription factors by concatenating their time-series, the PCA approach with constant k ensures that the input to the classifier always has the same dimension. This constraint avoids potential bias because of over-fitting, which can happen if the input to the classifiers has variable dimensions, and so avoids any related problems when comparing the estimated information across different decoding setups.

Fig. S4B shows the dependence of the estimated information on k; Fig. S4C shows its dependence on the number of observed time-series. Fig. S5 demonstrates that the estimates of mutual information are not sensitive to the measure used to quantify nuclear localization.

We note further that common choices for ensemble classifiers and nonlinear (kernelized) SVMs do not outperform the linear classifier on our data. Specifically, we tried the Random Forest [26] and XGBoost [27] ensemble classifiers and the radial basis function (RBF) kernel in the SVM. For each calculation, we optimize over a large grid of input parameters (for

Random Forest, the number of trees to grow $N_{\text{tree}} \in [50-500]$ and the number of variables randomly sampled as candidates at each split $S \in [1-20]$; for XGBoost, the number of iterations $N_{\text{iter}} \in [20-200]$ and maximum tree depth $D \in [3-10]$; and for SVM with the RBF kernel, the cost $C \in [2^{-3}-2^{9}]$ and the kernel parameter $\gamma \in [2^{-15}-2^{3}]$) by performing four rounds of 4-fold cross-validation for each parameter set and then for the five best-performing parameter sets, taking 100 bootstraps as described above and reporting MI for the highest of these bootstraps. For the ensemble classifiers and SVM classifiers with nonlinear kernels, training and testing was performed on the time series $\mathbf{x}'$ rather than on the projection onto PCA components. There is little change in the estimated mutual information for both the $q = 4$ stress-type and our $q = 5$ stress-level experiments (Fig. S6).

3.3 Decoding error and mutual information

We can summarize the classifier's performance as the fraction of correctly and incorrectly classified cell time-series. For example, for a transition in which two environmental states S are possible (s_0 before the transition and s_1 after the transition) the confusion matrix M is:

$$M = \begin{pmatrix} P(\hat{S}=s_0|S=s_0) & P(\hat{S}=s_1|S=s_0) \\ P(\hat{S}=s_0|S=s_1) & P(\hat{S}=s_1|S=s_1) \end{pmatrix}, \tag{7}$$

where S represents the actual environment and $\hat{S}$ represents the environment predicted by the classifier.

More generally, the confusion matrix M defines an empirical estimate of the conditional probability distribution for the predicted environment $\hat{S}$ given the actual environment S, and its elements M_{ij} represent the probability that an element, whose actual state is i, was classified as j. Therefore the elements in the diagonal represent correct classification; off–diagonal elements represent error.

The fraction of correctly classified cells is the weighted average over the diagonal elements: $f_{\text{correct}} = \sum_{i=1}^{q} P(S=s_i)P(\hat{S}=s_i|S=s_i)$. The decoding error is $1 - f_{\text{correct}}$.

3.4 Mutual information is robust across experimental repeats

We calculated the mutual information and decoding error for all 60 experiments in the screening (30 experiments shown in Fig. 2B plus an additional biological replicate for each experiment).

The calculations of decoding error and mutual information are robust across biological replicates and show only small fractional differences despite experimental variation and differing numbers of cells in each experiment (Figs. S7A & B). There is a monotonically decreasing relation between the decoding error and mutual information for the $q = 2$ transition (Fig. S7C). As a control, we calculated the information for the same experiments but with randomized data (random permutation of all time-points in each time-series) which yielded no information (Fig. S7D), as expected for random classification ($P = 0.5$ and MI $= 0$). Finally, environmental transitions in which transcription factors showed no visible response (e.g., Hog1 in either carbon or oxidative stress and Yap1 in either carbon or osmotic stress) also yielded mutual information close to zero bits (Fig. S7D).

For the stress-type ($q = 4$) experiments, the percentage difference in mutual information between Fig. 3A and the biological replicate was under 15% and rank order was preserved.

We note too that estimating the mutual information with an explicit transition from rich media to rich media rather than using the time-series in rich media before the environmental transition produces no substantial changes (Fig. S8).

3.5 Estimating the encoding delay

The long-time mutual information reports the maximum amount of bits encoded by a given transcription factor, i.e. the value at which information plateaus as a function of the duration of the response, and which we refer to as $\mathrm{MI}_{\mathrm{lt}}$. The encoding delay reports the speed at which information is encoded by the transcription factor and is defined as the time taken for the mutual information to reach 50% of its plateau value. For each transcription factor, a logistic fit was applied to the mutual information as a function of the response duration to estimate $\mathrm{MI}_{\mathrm{lt}}$ and t_{encoding} using the formula: $\mathrm{MI}_{\mathrm{lt}}/(1+\exp(-\alpha(t-t_{\mathrm{encoding}})))$ with α parameterizing the steepness of the curve. The values for $\mathrm{MI}_{\mathrm{lt}}$ and t_{encoding} reported are averages and standard deviations across 2 independent biological replicates.

3.6 Mutual information for multi-state transitions with $q > 2$

3.6.1 Behaviour of the ideal specialist

For the $q = 4$ transition, an ideal specialist (indicated by a dotted line in Fig. 3A) would correctly identify one environmental state, but randomly classify all remaining states. Assuming a specialist for environmental state s_2, the confusion matrix for this hypothetical case would be:

$$M = \begin{pmatrix} 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \end{pmatrix} \tag{8}$$

so that, for example, $P(\hat{S} = s_2|S = s_2) = 1$ and $P(\hat{S} = s_3|S = s_0) = 1/3$. Each row of M adds up to 1. An ideal specialist always correctly senses its associate stress and never reports another environmental state as this stress.

Direct calculation of the mutual information from Eq. 8 yields 0.81 bits. We also know that the mutual information is bounded above by the entropy of the distribution of input (the distribution of the environmental states) [18]. Although this distribution has 4 equally probably states, the distribution should appear for an ideal specialist as its associated stress, which has a probability of occurrence of $1/4$, and just one other state, which is not its associated stress and has a probability of occurrence $3/4$. The entropy of this apparent distribution is also approximately 0.8 bits, and the mutual information for an ideal specialist is equal to this upper bound. Similar calculations can be applied for the hypothetical situation of a 2-state specialist (one that distinguishes 2 states perfectly but lumps together the remaining states into a single category) and the mutual information is 1.5 bits.

Fig. 3A shows that both Hog1 and Yap1 perform as ideal specialists (MI $\approx$ 0.8 bits), decoding nearly optimally their own environmental state but classifying randomly, and therefore lumping together, all other states.

3.6.2 Estimating the probability of predicting each environmental state

For all transcription factors, we can use the confusion matrix as an estimate of the conditional probability of predicting the environmental states for the data of Fig. 3A (Fig. S11) and of predicting the levels of stress for the data of Fig. 3C (Fig. S12).

3.6.3 Distinguishing one level of stress from another

For each transcription factor and each type of stress, we calculated the mutual information between all possible pairs of levels of stress and their respective time-series of nuclear localization (the stress-level experiments of Fig. 3C with $q = 5$). This pairwise mutual information indicates whether the two levels of stress can be distinguished from each other by observing a typical single-cell time-series and is presented using a heat map (Fig. S13). Although the specialists show a similar change in information between neighbouring levels of stress for all levels, the generalists have more of a threshold response, particularly for oxidative stress where only high levels of stress can be distinguished.

In general, the pairwise mutual information matches differences in the fitness penalty (Fig. S14). Transcription factors that indirectly affect growth (Dot6, Sfp1, and Maf1) have the highest rank correlation.

3.6.4 Condition-specific information

The condition-specific information [23] is the mutual information obtained by considering responses from only one environment and quantifies how different that environment's responses are from the expected range of responses. Expressed in terms of the decoding estimator, the condition-specific information can be written as:

$$I\left(\hat{y}, y = s_j\right) = \sum_{i=1}^{q} P\left(\hat{y} = s_i | y = s_j\right) \log_2 \left(\frac{P\left(\hat{y} = s_i | y = s_j\right)}{P\left(\hat{y} = s_i\right)} \right). \tag{9}$$

The mutual information is an average over j of $I(\hat{y}, y = s_j)$ weighted by $P(y = s_j)$ (Eq. 5).

The condition-specific information between the single-cell nuclear localization time-series and each type of stress is plotted in Fig. S15. High values of condition-specific information for a given environment show that it is easily distinguished from the other environments. For example, all the specialist transcription factors (Mig1 and Mig2 for carbon stress; Hog1 for osmotic stress; Yap1 for oxidative stress) show high condition-specific information for their associated stress. Indeed, for Hog1 and Yap1, the condition-specific information in the non-responsive conditions is close to $\log_2(4/3) \approx 0.4$ bits (the value expected for an ideal specialist given four environmental states for which $P(\hat{y} = s_{i \neq k} | y = s_{j \neq k}) = 1/3$ where k denotes the specialist's associated stress) because there is information in identifying the absence of the stress. The generalist transcription factors do not show strong preferences for any environment. Nevertheless, Msn2 and Msn4 appear to favour both oxidative stress and identifying the absence of stress (rich medium), and Dot6 appears to favour osmotic stress and is complementary to Msn2/4.

3.7 Features of the time-series of localization that encode mutual information

The mutual information between the environmental states and the single-cell time-series determines the maximum information available to downstream processes, but we do not know how that information is either encoded or potentially decoded by the cell. Nevertheless, our estimates of mutual information show that at least that much information could be recovered using a general linear mechanism of decoding, given our use of a linear SVM.

We can gain additional insight by calculating the mutual information for suitable features derived from the single-cell time-series. For example, if almost all of the mutual information between a set of time-series and the environmental transitions could be captured at a single time-point, then a transition could be detected simply by monitoring the amplitude of the response at that time-point and monitoring the time-series is unnecessary.

We therefore calculated the mutual information at each single time-point and also for a selection of other features of the time-series (Fig. S16A) from both the stress-type ($q = 4$) and stress-level ($q = 5$) experiments.

Irrespective of the chosen feature, transcription factor, or environments, the mutual information of the full time-series is always greatest and a response duration of 25 minutes is sufficient to reach the long-time mutual information for most transcription factors (Figs. S9 & S10). Whether for stress-type or stress-level experiments, the mutual information contained in any single time-point never exceeds 1 bit (Fig. S17). This observation highlights the importance of dynamic responses for encoding complex inputs.

We find that for distinguishing the type of stress, generalists encode with their dynamics, whereas the specialists need just a single time-point to identify their associated stress if the stress is sufficiently high (Fig. S17A). In contrast, for distinguishing the level of stress (Fig. S17B), the specialists then use dynamics. For levels of either osmotic or oxidative stress, the generalists Dot6 and Msn2 encode little more in their full time-series than they do in a single time-point.

If each single-cell time-series is ordered by rank, then information about timing is lost, but common summary statistics are retained, such as the median and interquartile range (Fig. S16B). This 'rank ordered' feature captures most but not all of the mutual information available in the full time-series. There are notable exceptions: Sfp1 and Maf1 still require time-ordering to distinguish stress type (Fig. S17A), and Msn2 and Dot6 require time-ordering to distinguish levels of carbon stress (Fig. S17B).

Features used to describe the shape of the adaptive pulse [21] provide information, but each by itself provides less than 1 bit (Fig. S17) If these features are considered in combination, however, they can match and in some cases exceed the information of the 'rank ordered' feature. This highlights the importance of timing for Msn2, Msn4 and Dot6 in distinguishing stress type (Fig. S17A) and for Msn2 and Dot6 in distinguishing levels of carbon stress (Fig. S17B). In any case, the combined information of these peak features only matches that of the full time-series if all the available information is contained in a single time-point (by the amplitude) such as for the specialists Yap1 and Hog1 in distinguishing stress type (Fig. S17A).

3.8 Estimating channel capacity and optimal input distributions

For the estimations of mutual information in the main text we assume a uniform distribution of inputs because our focus is on the signal-to-noise ratio in the single-cell time-series, but the results do not change substantially if we use the input distribution that maximizes the mutual information (Fig. S18). We used the Blahut-Arimoto algorithm [18] to estimate this maximal mutual information (the capacity). Based on the confusion matrix, the algorithm finds the channel capacity through iterative optimization and outputs both the capacity and the optimal input distribution. For the environment with $q = 4$ (rich media and three types of high stress), we note that generalists perform close to capacity, but not specialists, implying that generalists favour a uniform distribution of environmental states.

3.9 Mutual information for pairs of transcription factors as outputs

The mutual information for pairs of transcription factors was calculated using concatenated time-series of nuclear localization. A concatenated time-series for two transcription factors TF_1 and TF_2 comprises t time-points of the response of TF_1 followed by t time-points of the response of TF_2. The new time-series is treated as the response of a single cell. To account for variability and possible bias in the synthetic data, we calculated the mutual information for 50 bootstrap replicates, each time generating different datasets of concatenated time-series by sampling randomly time-series of TF_1 and TF_2.

Our data only comes from strains in which only one transcription factor is fluorescently tagged, and our estimation of mutual information by concatenating time-series we shall show assumes conditional independence between the activation of different transcription factors given the environmental state. To estimate the mutual information for a distribution of time-series of a single transcription factor (TF), we must estimate $P(\mathrm{TF}|S)$ where S is the state of the environment; to estimate the mutual information for the distribution of a pair of time-series from two tagged transcription factors, we must estimate $P(\mathrm{TF}_1, \mathrm{TF}_2|S)$. For two concatenated time-series, we can estimate $P(\mathrm{TF}_1|S)P(\mathrm{TF}_2|S)$. Therefore we require $P(\mathrm{TF}_1, \mathrm{TF}_2|S) = P(\mathrm{TF}_1|S)P(\mathrm{TF}_2|S)$, which is conditional independence, if the mutual information estimated from the concatenated time-series is to equal the mutual information from doubly tagged strains.

To confirm that the assumption of conditional independence holds, we created two-colour strains in which we tagged two transcription factors using GFP and mCherry and applied three different medium: 0.4M NaCl, 0.1% glucose, and rich media (2% glucose). We calculate the mutual information between this 3-state environment and the dynamics of the transcription factors by concatenating the GFP and mCherry time-series from each cell and by concatenating GFP and mCherry time-series randomly selected from different cells (mimicking two separate experiments). There is no substantial difference between the mutual information calculated from time-series from the same cell and the mutual information calculated from time-series from different cells (Fig. S19).

3.9.1 Information redundancy between pairs of transcription factors

The degree of redundancy between pairs of transcription factors was computed for the stress-type experiments ($q = 4$) using the formula

$$r = 1 - \frac{\mathrm{MI}_{12}}{\mathrm{MI}_1 + \mathrm{MI}_2} \tag{10}$$

where MI_{12} is the mutual information calculated from the pair (the concatenated data) and MI_i is the mutual information calculated individually. If the two transcription factors are completely independent ($\mathrm{MI}_{12} = \mathrm{MI}_1 + \mathrm{MI}_2$), then $r = 0$; if the transcription factors are completely redundant ($\mathrm{MI}_{12} = \mathrm{MI}_1 = \mathrm{MI}_2$), then $r = 1/2$. We observed positive redundancy values for all the pairs investigated and never synergistic encoding where $\mathrm{MI}_{12} > \mathrm{MI}_1 + \mathrm{MI}_2$ and $r < 0$.

We report the average redundancy calculated from 2 independent datasets (a total of 6 experiments per transcription factor) and generated the dendrogram in Fig. 4B by applying a clustering algorithm (`heatmap` function in `R` [28]) to the matrix of redundancy.

From the redundancy matrix, we generated a network in which transcription factors are represented as nodes and the length of an edge connecting two nodes is inversely proportional to the redundancy between the corresponding transcription factors (more redundant transcription factors are therefore closer to each other in the network). We only display an edge between two transcription factors if the redundancy is greater than a threshold value equal to the minimum across all pairs, such that all transcription factors have at least one edge. We plotted the network using `igraph` [29] (with layout parameter `layout_with_fr`) and an alternative version in which a less stringent cut-off value was used to define the edges is shown in Fig. S20.

3.9.2 Relative timing of responses as an informative feature

We note that transcription factors translocate either into or out of the nucleus with different speeds. We define the response time with respect to $t = 0$ as the time at which nuclear localization reaches 50% of either its maximum value for transcription factors that

enter the nucleus or its minimum value for transcription factors that exit the nucleus. The response time can be different for different transcription factors in the same condition and for the same transcription factors across conditions. When considering pairs of transcription factors, their individual response times also define a relative timing: the difference between their response times.

We can estimate how much relative timing contributes to the mutual information by calculating the difference between the mutual information of the pair and the mutual information of the pair after their time-series have been aligning to reduce differences in response times. To artificially aligned the time-series, we calculate the response time for the mean time-series of each transcription factor and then align by shifting time for all the individual time-series for that transcription factor by this response time. On average, all transcription factors then reach 50% activation at $t = 0$.

Pairs of generalists are more likely to encode information in their relative timing than pairs of specialists (the aligned time-series of generalists have a mutual information reduced by up to 40% but aligning hardly alters the mutual information for pairs of specialists) (Fig. S21; $p < 0.001$). Pairs that include one generalist and one specialist also encode information in the relative timing although to a lesser extent than pairs of two generalists.

3.10 Mutual information for multiple transcription factors as outputs

Using an analogous approach to the one used for estimating the mutual information for two transcription factors, we generated time-series in which the time-series for more than two transcription factors were concatenated. For a given condition, we systematically concatenate randomly selected time-series of the first transcription factor with randomly selected time-series from a second transcription factor, and the process repeated iteratively to create time-series from n concatenated factors. The concatenated time-series then have length $n \times t$, where t is the length of the individual time-series. We conserve the order of the transcription factors.

Mutual information is calculated using the concatenated data (Fig. S22): for a $q = 2$ and $q = 4$ environment, we used the data from all transitions from rich media to high stress (Figs. 2D & 3A); for a $q = 5$ environment, we used the data with multiple levels of osmotic stress (Fig. 3C); for a $q = 7$ environment, we used rich media (before the transition) and 2 levels of carbon stress (0.1% and 0.4% glucose), 2 levels of oxidative stress (0.28 and 0.5 mM H_2O_2), and 2 levels of osmotic stress (0.2 and 0.4 M NaCl) (after the transition). We exhaustively calculated the mutual information for all combinations of transcription factors. To show the performance of the best combinations of transcription factors, we report the maximum over all possible combinations in Fig. 4E.

For the environment with $q = 7$ states, the confusion matrix implies that the generalists respond to categories of environments (Fig. S23)

3.11 Software availability

Both Matlab and R versions of the scripts used for estimating mutual information are freely available at `https://github.com/swainlab/mi-by-decoding`

4 Construction of the yeast signalling network

The yeast kinase interaction (KID) database [30] was used to construct the signaling networks that directly regulate the transcription factors. For each transcription factor, we retrieved all upstream regulatory interactions (kinase→TF) reported in the database along with interactions among the regulatory kinases (kinase→kinase). We then filtered for *bona fide*

interactions using the KID score [30] — a measure of the literature evidence for kinase-substrate interactions calibrated to maximize true positives — keeping only interactions with a score greater than 3. Edges in the network were also weighted by the KID score. Kinases that had only low-score interactions were removed. Although the interactions reported in KID are directional, we show undirected interactions for simplicity.

We applied a hierarchical algorithm to find communities in the network (`fastgreedy.community` from package `igraph` [29] in `R`-Bioconductor [28]). The algorithm merges nodes iteratively in communities while optimizing a modularity function. The resulting communities (modules) of the network are shown in Fig. 4A. When performing this neighbour analysis, the ten transcription factors and kinases were treated equally as nodes in the network.

For plotting the network, we used the plotting function of `igraph` with layout parameter `layout_with_fr`. The algorithm uses no *a priori* knowledge of the biological signaling pathways, and so the communities found (corresponding to the different colors in Fig. 4A) are based solely on the network's architecture.

References

1. Saito H, Posas F (2012) Response to hyperosmotic stress. *Genetics* 192(2):289–318.

2. Kuge S, Jones N, Nomoto A (1997) Regulation of yAP-1 nuclear localization in response to oxidative stress. *EMBO J* 16(7):1710–1720.

3. Jacquet MM, Renault GG, Goldbeter AA (2003) Oscillatory nucleocytoplasmic shuttling of the general stress response transcriptional activators Msn2 and Msn4 in Saccharomyces cerevisiae. *J Cell Biol* 161(3):497–505.

4. Jorgensen P, et al. (2004) A dynamic transcriptional network communicates growth potential to ribosome synthesis and critical cell size. *Genes Dev* 18(20):2491–2505.

5. Willis IM, Moir RD (2007) Integration of nutritional and stress signaling pathways by Maf1. *Trends Biochem Sci* 32(2):51–53.

6. Lippman SI, Broach JR (2009) Protein kinase A and TORC1 activate genes for ribosomal biogenesis by inactivating repressors encoded by Dot6 and its homolog Tod6. *Proc Natl Acad Sci USA* 106(47):19928–19933.

7. Huh WK, et al. (2003) Global analysis of protein localization in budding yeast. *Nature* 425(6959):686–691.

8. Crane MM, Clark IBN, Bakker E, Smith S, Swain PS (2014) A microfluidic system for studying ageing and dynamic single-cell responses in budding yeast. *PLoS ONE* 9(6):e100042.

9. Granados AA, et al. (2017) Distributing tasks via multiple input pathways increase cellular survival in stress. *eLife* 6:e21415.

10. Bakker E, Swain PS, Crane MM (2018) Morphologically constrained and data informed cell segmentation of budding yeast. *Bioinformatics* 34(1):88–96.

11. Bakker E, Crane MM (2017) Source code for cell-segmentation software. *Github* github.com/pswain/segmentation-software:23b5c2e.

12. Cai L, Dalal CK, Elowitz MB (2008) Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* 455(7212):485–490.

13. Hao N, O'Shea EK (2011) Signal-dependent dynamics of transcription factor translocation controls gene expression. *Nat Struct Mol Biol* 19(1):31–39.

14. Lin Y, Sohn CH, Dalal CK, Cai L, Elowitz MB (2015) Combinatorial gene regulation by modulation of relative pulse timing. *Nature* 527(7576):54–58.

15. Lichten CA, White R, Clark IBN, Swain PS (2014) Unmixing of fluorescence spectra to resolve quantitative time-series measurements of gene expression in plate readers. *BMC Biotechnol.* 14(1):11.

16. Stevenson K, McVey AF, Clark IBN, Swain PS, Pilizota T (2016) General calibration of microbial growth in microplate readers. *Sci Rep* 6:38828.

17. Swain PS, et al. (2016) Inferring time derivatives including cell growth rates using Gaussian processes. *Nat Commun* 7:13766.

18. Cover TM, Thomas JA (2006) *Elements of information theory.* (Wiley-Interscience, Hoboken, New Jersey).

19. Kraskov A, Stögbauer H, Grassberger P (2004) Estimating mutual information. *Phys Rev E* 69(6):066138.

20. Voliotis M, Perrett RM, McWilliams C, McArdle CA, Bowsher CG (2014) Information transfer by leaky, heterogeneous, protein kinase signaling systems. *Proc Natl Acad Sci USA* 111(3):E326–E333.

21. Selimkhanov J, et al. (2014) Systems biology. Accurate information transmission through dynamic biochemical signaling networks. *Science* 346(6215):1370–1373.

22. Brunel N, Nadal JP (1998) Mutual information, fisher information, and population coding. *Neural Comput* 10(7):1731–1757.

23. Borst A, Theunissen FE (1999) Information theory and neural coding. *Nat Neurosci* 2:947–957.

24. Chang CC, Lin CJ (2011) LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2(3):27:1–27:27.

25. Strong S, Koberle R, de Ruyter van Steveninck R, Bialek W (1998) Entropy and Information in Neural Spike Trains. *Phys Rev Lett* 80:197–200.

26. Breiman L (2001) Random forests. *Machine Learning* 45(1):5–32.

27. Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16. (ACM, New York, NY, USA), pp. 785–794.

28. Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. *J Comput Graph Stat* 5(3):299–314.

29. Csardi G, Nepusz T (2006) The igraph software package for complex network research. *Inter J Comp Syst* 1695:1–9.

30. Sharifpoor S, et al. (2011) A quantitative literature-curated gold standard for kinase-substrate pairs. *Genome Biol* 12(4):R39.

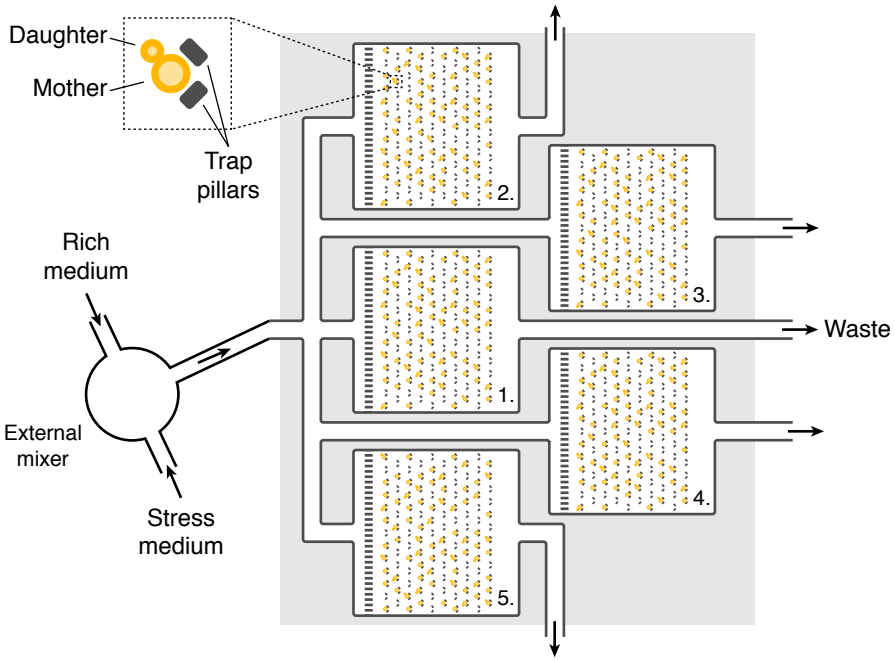

Fig. S1. Schematic showing the design of the 5-chamber ALCATRAS microfluidics device. In ALCATRAS devices, mother cells become trapped between PDMS pillars by the flow of media, which additionally carries budded daughters away and so prevents overgrowth in the device [8]. The 5-chamber version (gray square; chambers are numbered) allows for measurement of five different strains under identical environmental conditions. Cells can be loaded using the waste outlets. Cross-chamber contamination is avoided by PDMS filters (hash lines at left of each chamber). During the experiment, media is introduced to the device via an external mixer to enable switching between environments.

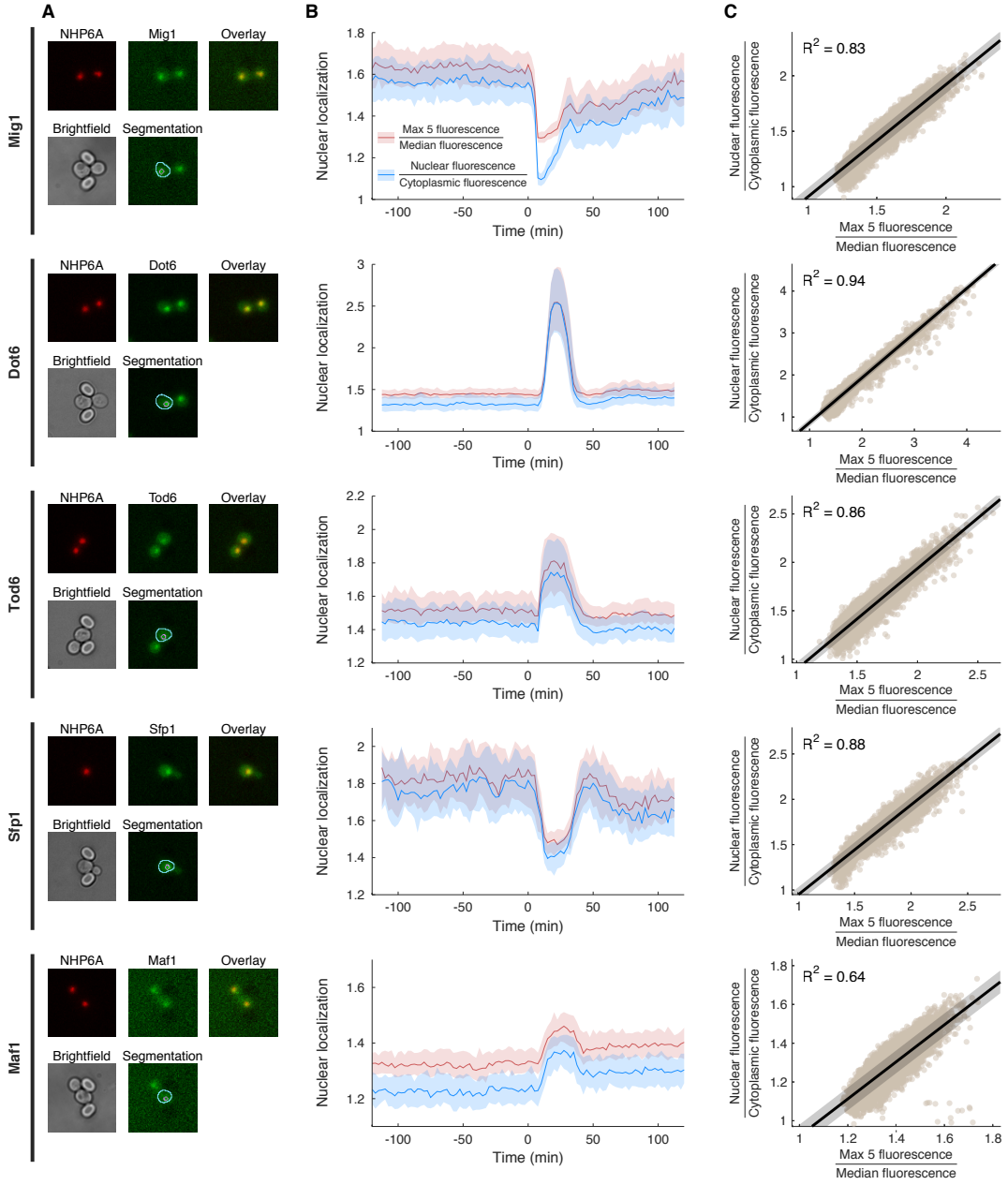

Fig. S2. Nuclear localization can be quantified without a nuclear marker. A Transcription factors are tagged with GFP, and mCherry added to NHP6A as a nuclear marker. Images are for a representative cell exhibiting nuclear localization (top row: sfluorescence of the mCherry and GFP tags; bottom row: the corresponding bright-field image and segmented cell (cyan: active contour based on bright-field image) and nucleus (pink: watershed segmentation of NHP6A tag)). **B** Nuclear localization, reported as the ratio of nuclear fluorescence (median inside segmented nucleus) to cytoplasmic fluorescence (median of cell excluding segmented nucleus), for shifts from rich media to carbon stress at time 0 either with (blue) or without (red) the nuclear marker. An approximation of nuclear localization is the ratio of the mean of the cell's 5 brightest pixels (Max 5 fluorescence) to the cell's median fluorescence (Median fluorescence) [12]. **C** This measure correlates with that using a nuclear marker (linear regression in black with the estimated standard deviation of the error in regression shaded).

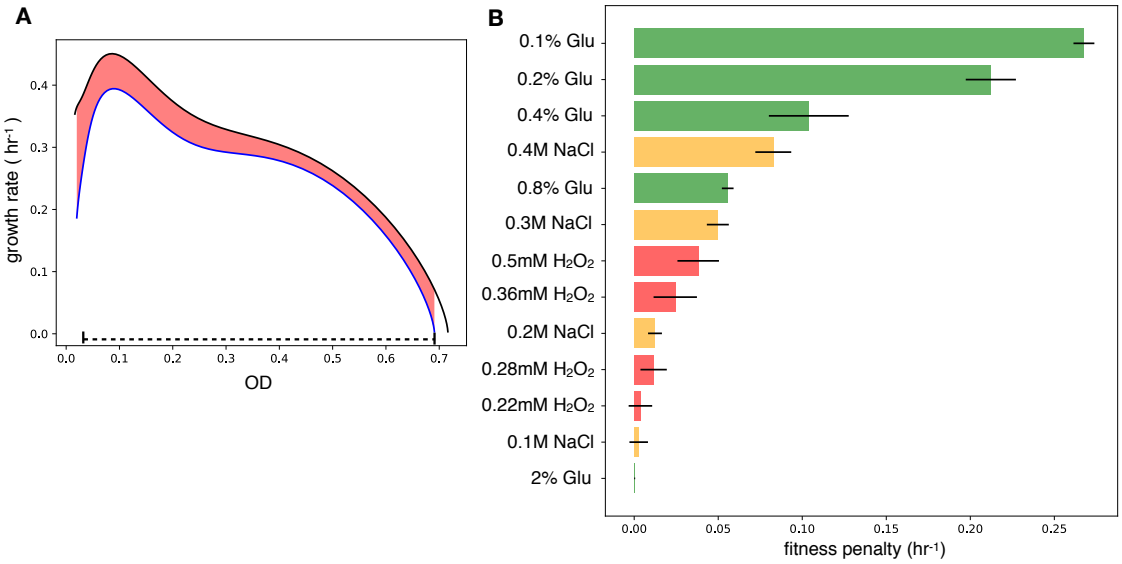

Fig. S3. The change in growth between 2% glucose and stress conditions defines a fitness penalty that typically correlates with mutual information. A From time-series data measuring the growth of a population of cells, we plot growth rate versus OD. The control for growth in 2% glucose is shown in black; growth in stress (here in 2% glucose and 0.5 mM H_2O_2) is shown in blue. We define the fitness penalty as the area between the curves (in pink) divided by the largest common range of OD (the dashed line). **B** In general, low glucose is a more severe stress than high salt which is a more severe stress than high H_2O_2 (median of $n = 4$ with SD as error).

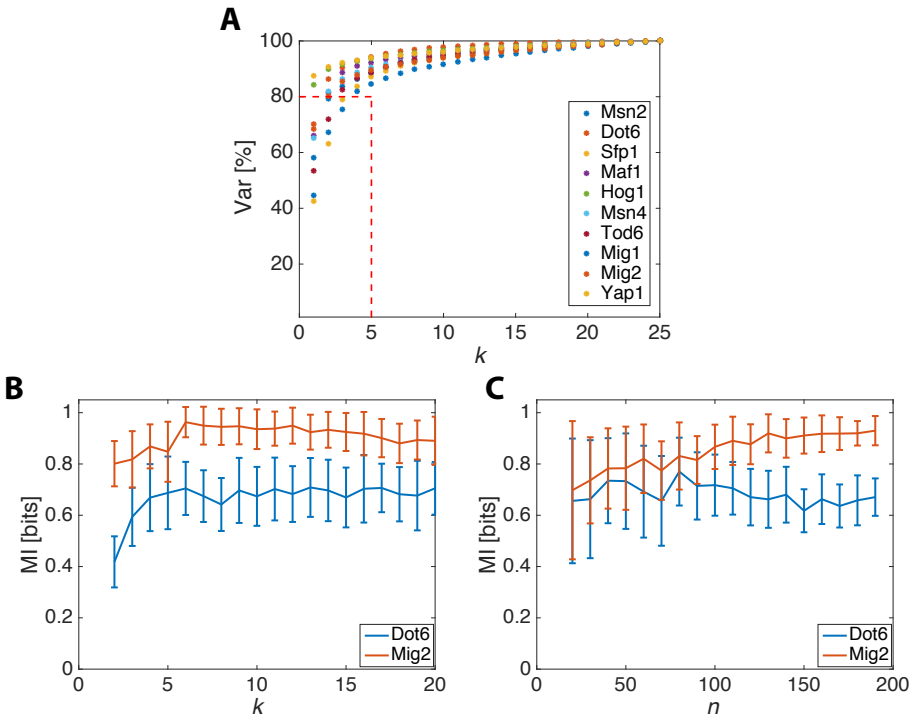

Fig. S4. Estimating mutual information requires k = 5 PCA components and n ∼ 100. **A** Five principal components are sufficient to explain at least 80% of the variance in the single-cell time-series for a $q = 4$ environment (the stress-type experiments of Fig. 3A). The cumulative percentage variance is shown on the y-axis and k is the number of principal components. The cases with lower values of q converge faster and fewer components are required. Nevertheless, we fix $k = 5$ for all the estimations for consistency. **B** Estimated mutual information versus k. Typically, having $k = 5$ PCA components yields nearly maximal estimates of information. In the limit of infinite samples, the MI curves should either monotonically increase or saturate with k; with a finite number, n, of sample time-series, however, the information can start slowly decreasing because of overfitting by the classifier. **C** Estimated mutual information for two environmental states (the state before and the state after the transition: $q = 2$) using n time-series is approximately constant above $n \sim 100$ time-series.

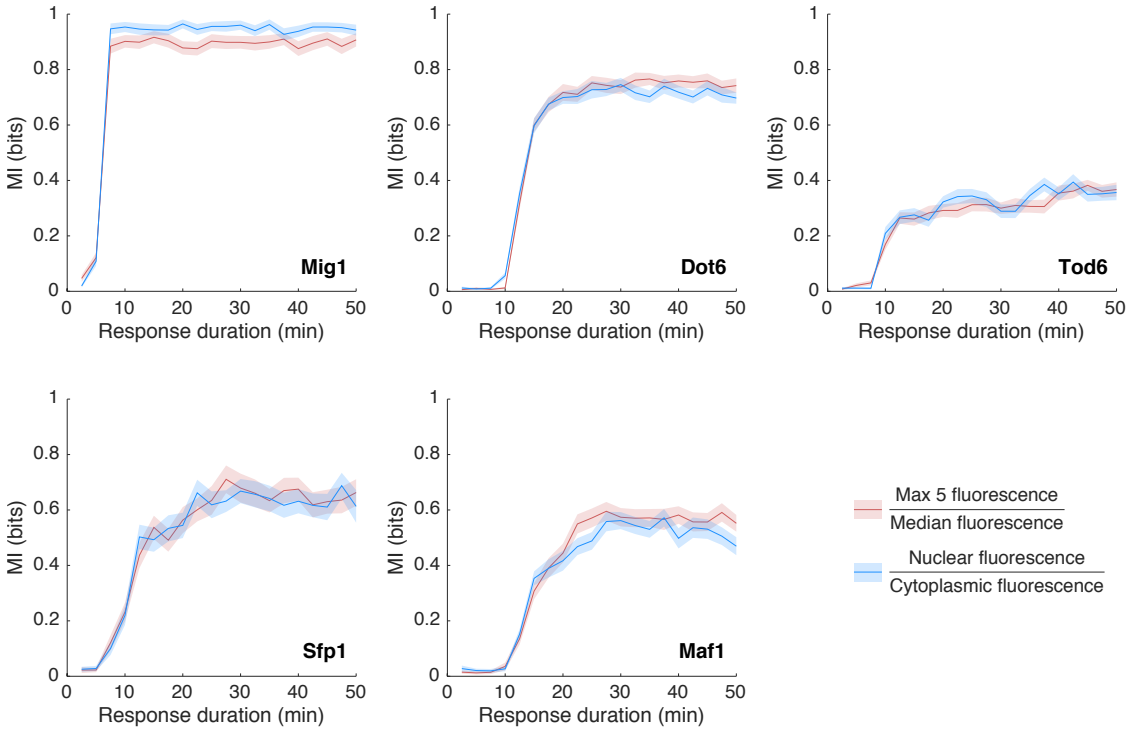

Fig. S5. Mutual information is robust to the measure of nuclear localization. The mutual information (MI) between the environment ($q = 2$: rich medium into carbon stress) and time-series of nuclear localization (data from Fig. S2) was calculated for measures of nuclear localization either requiring a nuclear marker (blue) or not (red). The shaded area shows 95% confidence limits in the mean across ($n = 52$) bootstrap replicates.

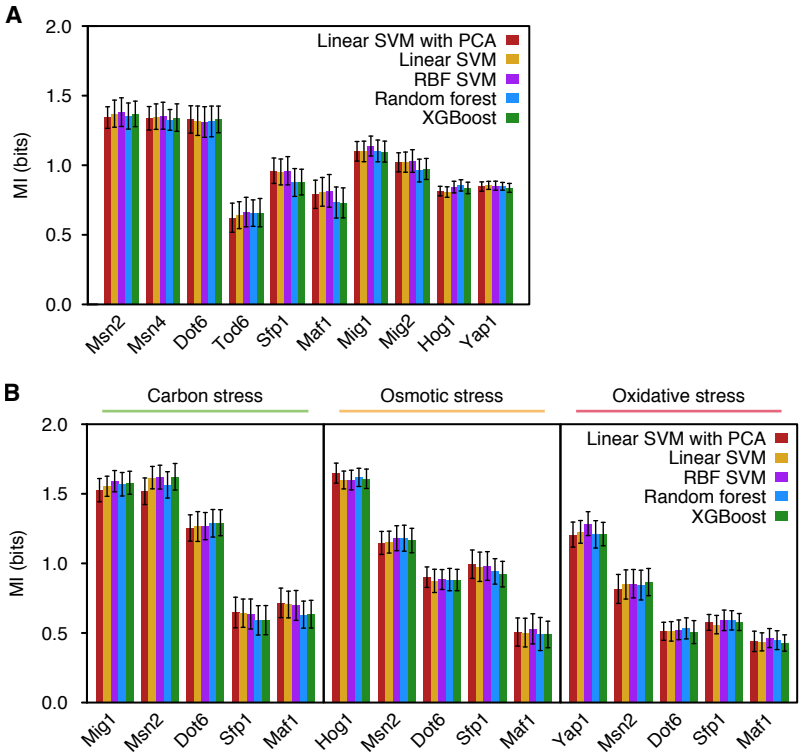

Fig. S6. Ensemble classifiers and nonlinear SVMs do not outperform a linear SVM classifier for time series of nuclear localization. Alternative classifiers were substituted into our algorithm for estimating mutual information. Error bars indicate standard deviations across ($n = 100$) bootstrap replicates. **A** Long-time mutual information calculations for the stress-type experiments with $q = 4$ of Fig. 3A. **B** Long-time mutual information calculations for the stress-level experiments with $q = 5$ of Fig. 3C.

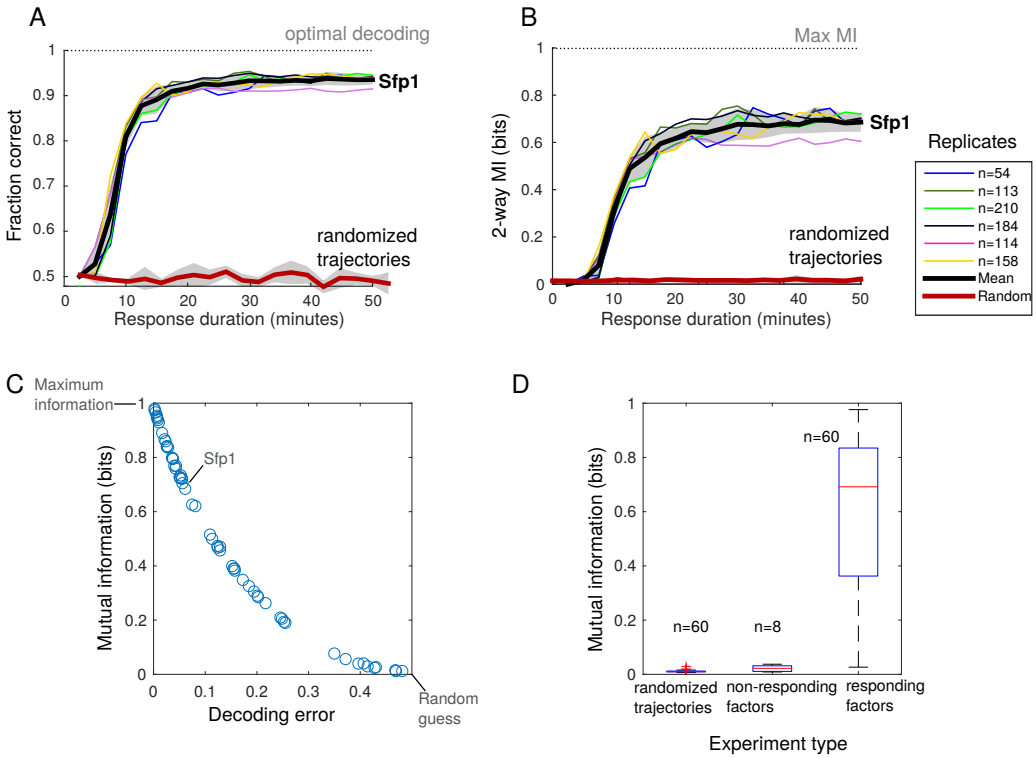

Fig. S7. Mutual information is robust across experiments. **A** The fraction of correctly classified time-series and **B** mutual information is similar for each of 6 biological replicates of a transition from rich media into carbon stress. The mean across replicates is in black with the corresponding standard deviation shaded. The mutual information when the time-series are randomly permuted is also shown. **C** Mutual information can be mapped to the decoding error. Each point is an average (averaged over all durations of the $q = 2$ environment that have maximum mutual information) and represents an independent experiment. Experiments for 10 transcription factors, 3 conditions and 2 biological replicates (60 experiments in total) are included. **D** Our algorithm consistently yields zero bits for all 60 datasets when the cell time-series are randomized (0.01 ± 0.003 bits). Experiments for transcription factors that do not respond to the environmental transition also give approximately 0 bits (0.02 ± 0.01 bits). For comparison, the third box-plot shows the mutual information for all 58 experiments in which the transcription factors do respond.

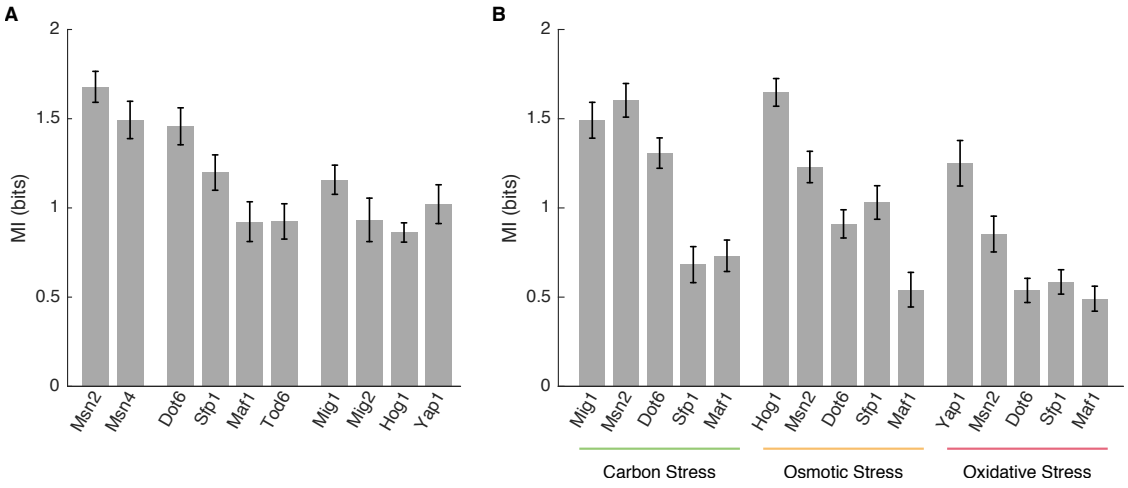

Fig. S8. An explicit rich media to rich media transition is not necessary to estimate the mutual information. Error bars indicate standard deviations across ($n = 52$) bootstrap replicates. **A** The long-time mutual information for the stress-type ($q = 4$) experiments compares favourably with Fig. 3A. **B** The long-time mutual information for the stress-level ($q = 5$) experiments compares favourably with Fig. 3C.

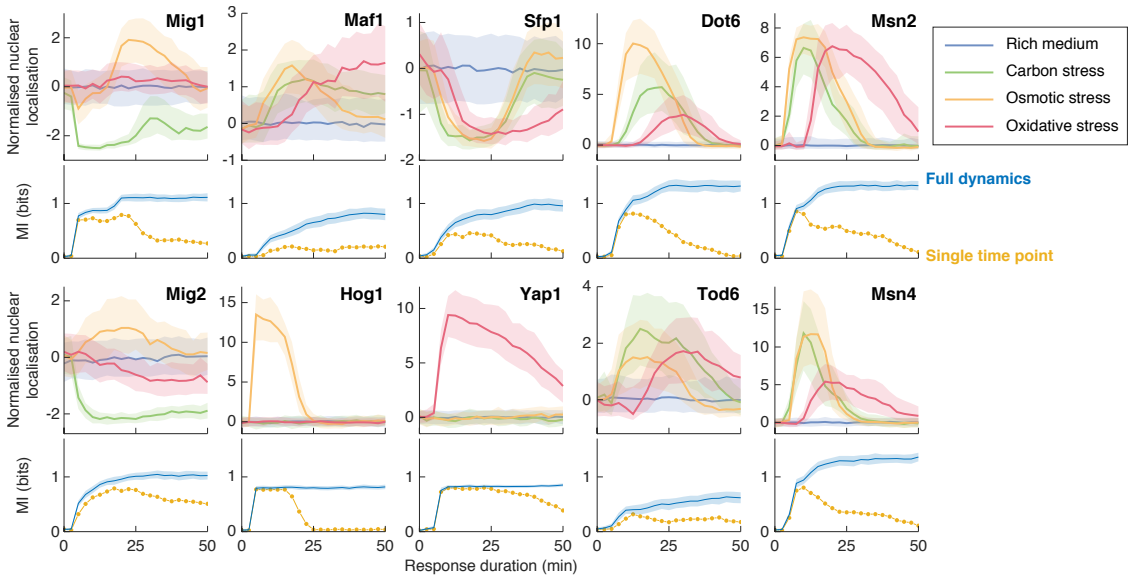

Fig. S9. Mutual information saturates in under 50 minutes for all transcription factors in the stress-type ($q = 4$) experiments. We plot the median normalized nuclear localization for each type of stress (shaded areas are the interquartile range) and the mean and standard deviation (across $n = 100$ bootstrap replicates) of the mutual information ($q = 4$).

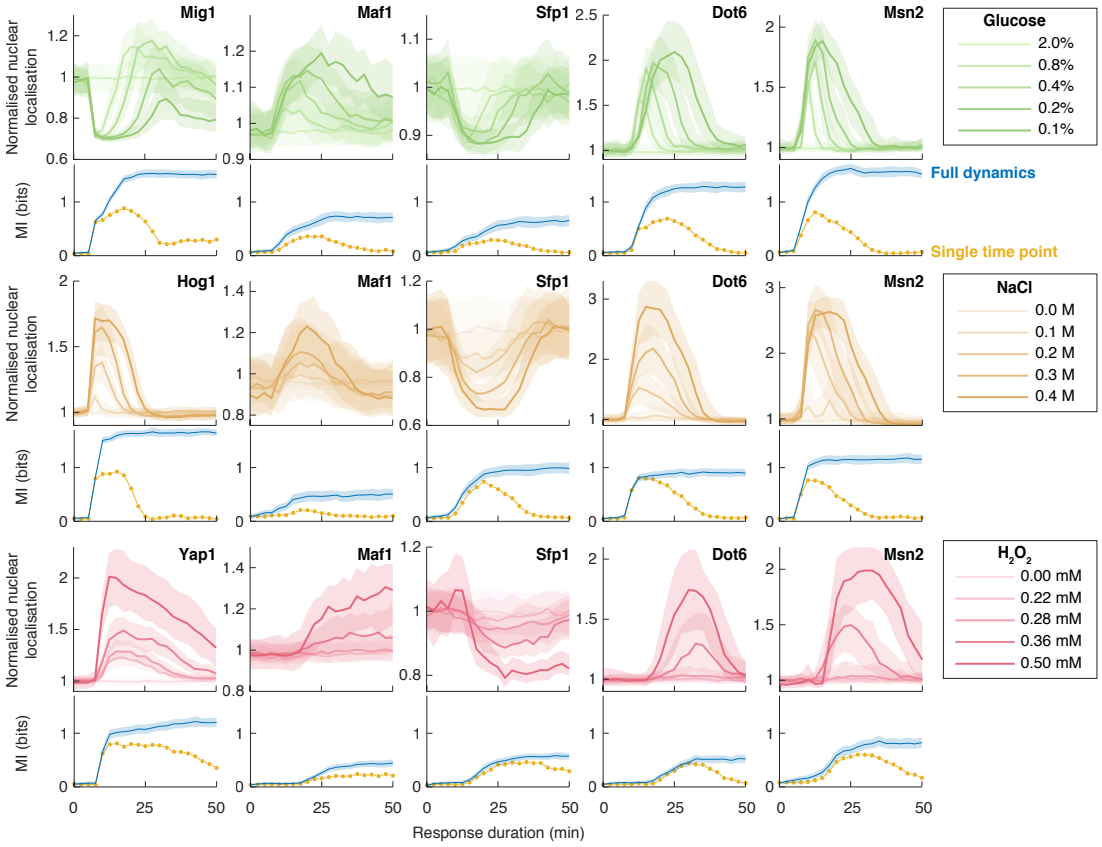

Fig. S10. Mutual information saturates in under 50 minutes for all transcription factors in the stress-level ($q = 5$) experiments. We plot the median normalized nuclear localization for each level of stress (shaded areas are the interquartile range) and the mean and standard deviation (across $n = 100$ bootstrap replicates) of the mutual information ($q = 5$).

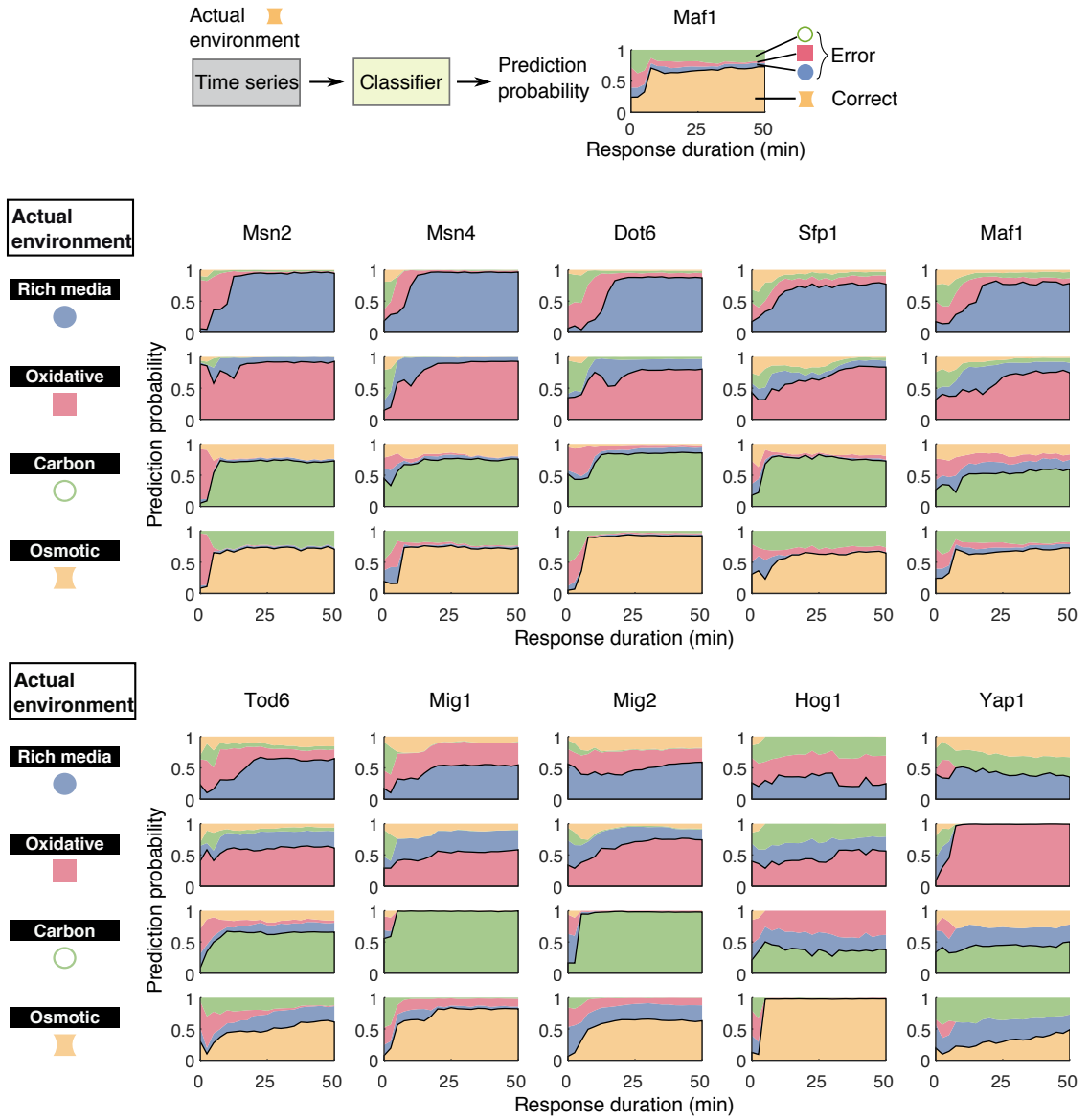

Fig. S11. The probability of predicting the state of a $q = 4$ environment from a time-series of nuclear localization. From the classifier, we can estimate the probability that a single-cell time-series is predicted to be from one of the four possible environmental states (the stress-type experiments with $q = 4$ in Fig. 3A). In the schematic, the time-series come from an environment with osmotic stress, and the total error is therefore the proportion of time-series that were not classified as being from osmotic stress. For each environment and for each transcription factor, we plot how these probabilities change as a function of the duration of the environment. The classification probabilities for Mig1, Yap1 and Dot6 are reproduced in Fig. 3B.

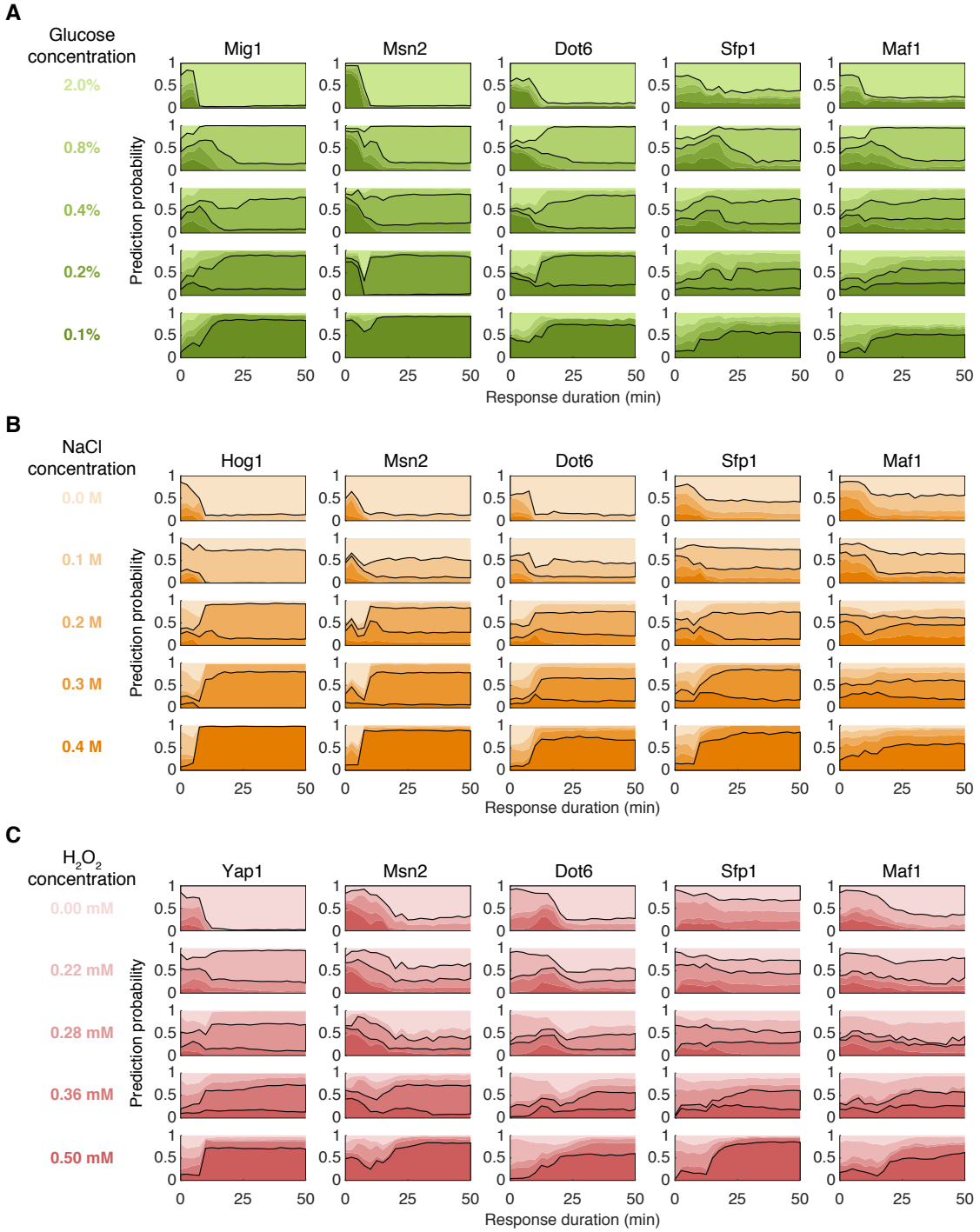

Fig. S12. The probability of predicting that state of a $q = 5$ environment from a time-series of nuclear localization. Analogous to Fig. S11, but considering environments with 5 levels of one type of stress (the stress-level experiments with $q = 5$ from Fig. 3C). The coloured numbers on the left indicate the magnitude of the stress; the area plots indicate the probability of predicting a particular level of stress from a single-cell time-series. The area corresponding to the probability of identifying the correct level of stress is bounded by black lines. **A** Carbon stress; **B** Osmotic stress; **C** Oxidative stress.

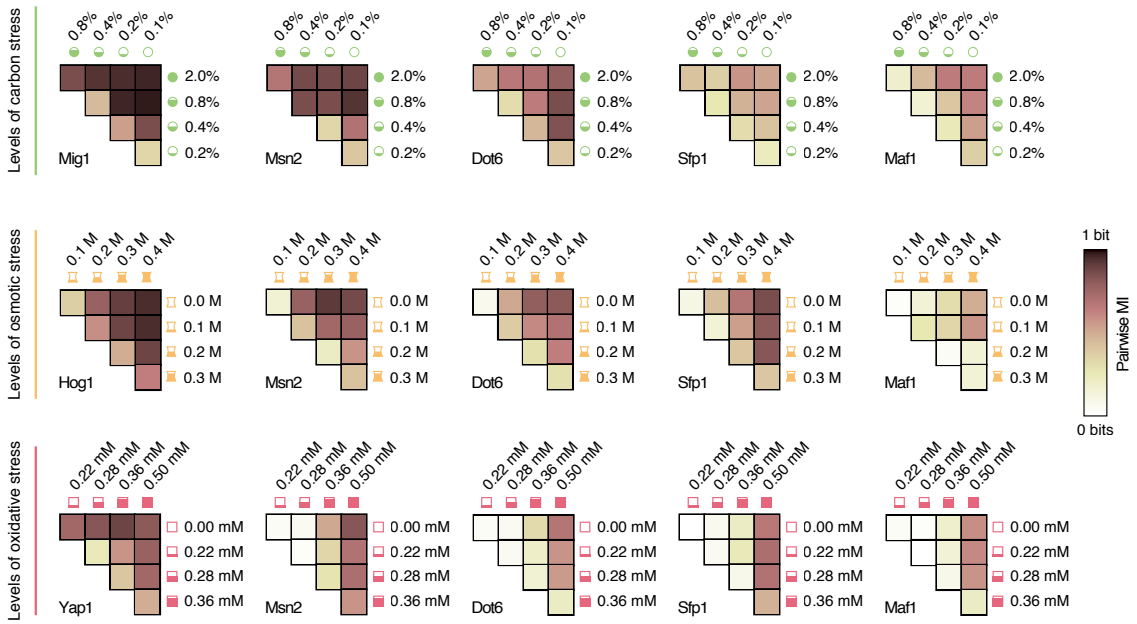

Fig. S13. Transcription factors vary in their abilities to encode increasing levels of stress in their dynamics of localization. For each transcription factor and type of stress, the mutual information was calculated between all possible pairs of levels of stress and the corresponding time-series of nuclear localization.

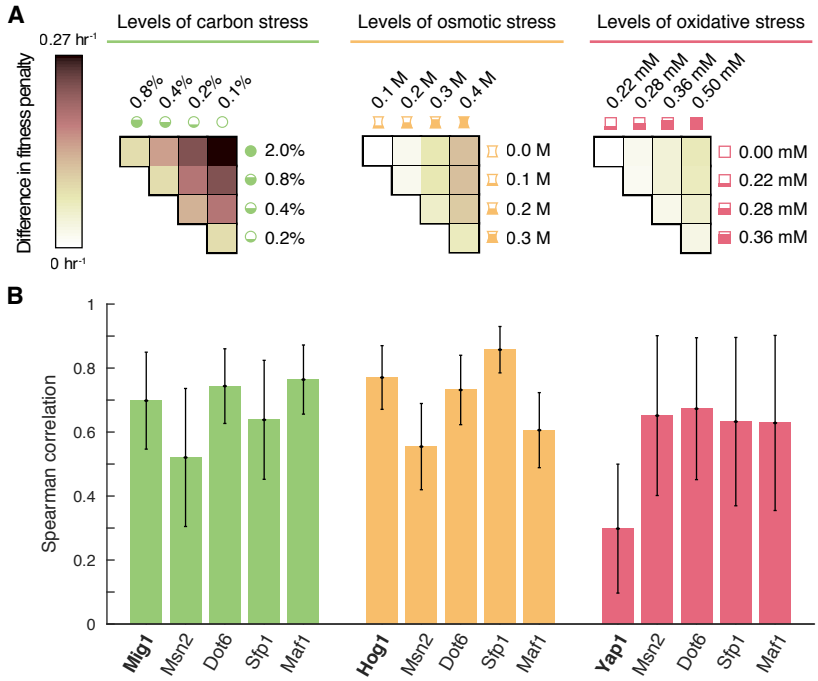

Fig. S14. Mutual information between pairs of stresses typically correlates with the difference in fitness penalty for those stresses. A A heap map showing the difference in fitness penalty between all pairs of levels of stress was calculated for the stress-level experiments of Fig. 3C (using the data from Fig. S3B). **B** The rank correlation of the pairwise difference in fitness penalties with the pairwise mutual information (shown in Figs. fig:pairwise-levels & 3D) is high for most transcription factors. We show the mean and standard deviation over ($n = 100$) bootstrap replicates drawn from the biological replicates of the fitness penalties and from the bootstrap replicates of the calculations of pairwise mutual information.

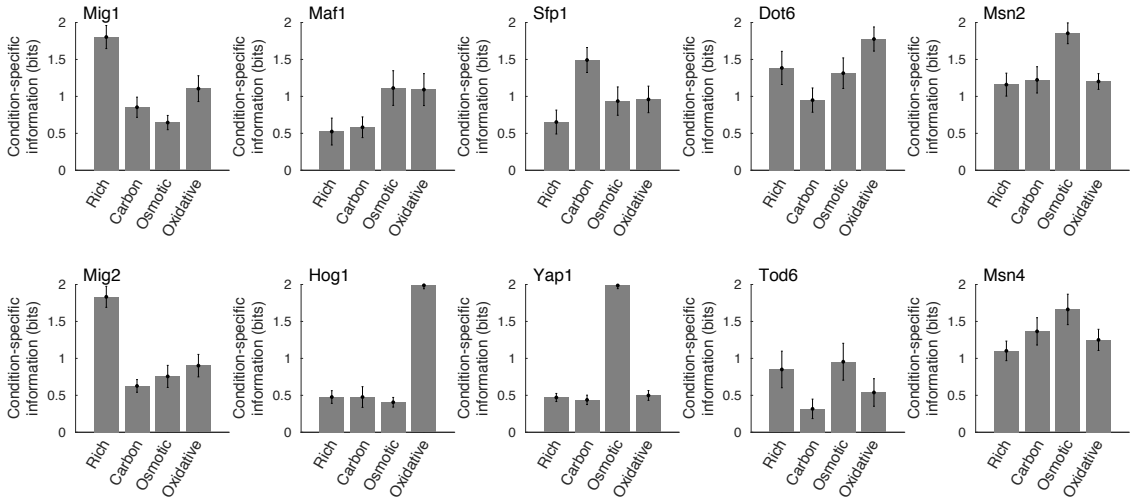

Fig. S15. Condition-specific information for transcription factors distinguishes specialists from generalists. The long-time condition-specific information (Eq. 9) between the single-cell time-series of nuclear localization and the environmental state for each of the transcription factors from the stress-type ($q = 4$) experiments. In the inset of Fig. 3A, this data is projected along three axes (all but the condition-specific information in rich media). Error bars indicate standard deviation across ($n = 24$) bootstrap replicates.

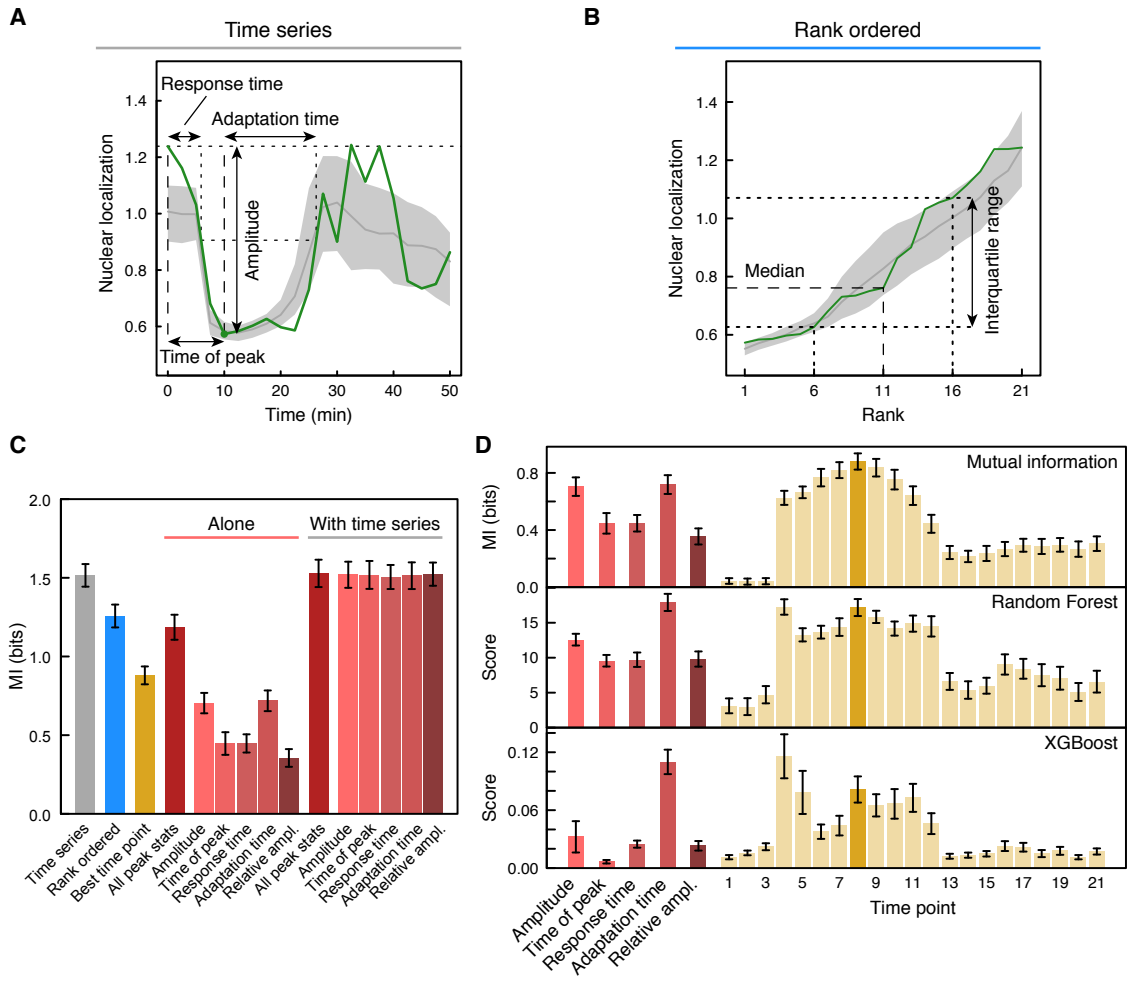

Fig. S16. The mutual information of features derived from the localization time-series indicates feature importance. **A** Features that characterize the adaptive pulse can be calculated for each cell, as demonstrated on a Mig1 nuclear localization trace (green line) for a transition from rich medium to 0.2% glucose. The median (grey line) and interquartile range (shaded area) of all cells is shown for reference. **B** By rank-ordering the localization trace of each cell in panel A, timing information is lost, but linear transformations of this data still provide statistical features. **C** Features of the traces only reduce information (as expected by the data processing inequality) and our algorithm is optimal since no additional information is gained by concatenating hand-crafted features onto the time-series. Shown is the mutual information derived from the Mig1 stress-level ($q = 5$) experiments. 'All peak stats' is the concatenation of five features: the pulse features illustrated in panel A and the relative amplitude (ampl.), the ratio of amplitude and initial localization value. For each feature, we show the mean and standard deviation (across $n = 100$ bootstrap replicates) of the mutual information. **D** Mutual information of the features compares favourably with estimates of feature importance from ensemble classifiers. In the top panel is the mutual information of each feature; in the lower panels is the importance score of each feature, obtained by training the specified ensemble classifier on a concatenated data set consisting of the time-series and all peak statistics. The mean and standard deviation (across $n = 100$) bootstrap replicates is shown.

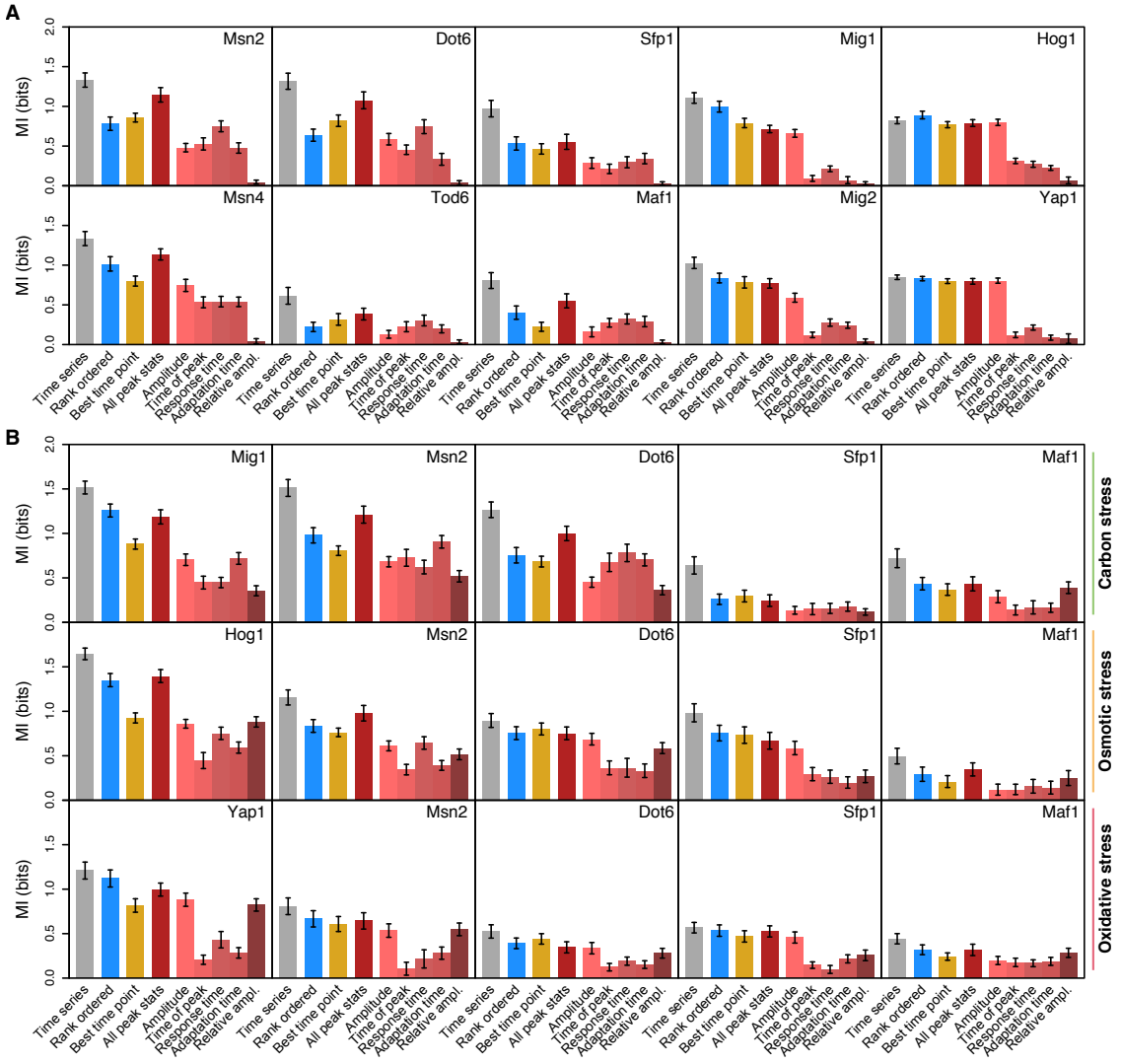

Fig. S17. Features of the localization time-series that encode mutual information. Statistical features were derived from the normalized time-series of nuclear localization as described in Fig. S16, and the mutual information calculated between these features and the state of the environment. For each transcription factor, we show the mean and standard deviation (across $n = 100$ bootstrap replicates) of the mutual information. **A** The mutual information of features in the stress-type ($q = 4$) experiments. **B** The mutual information of features in the stress-level ($q = 5$) experiments.

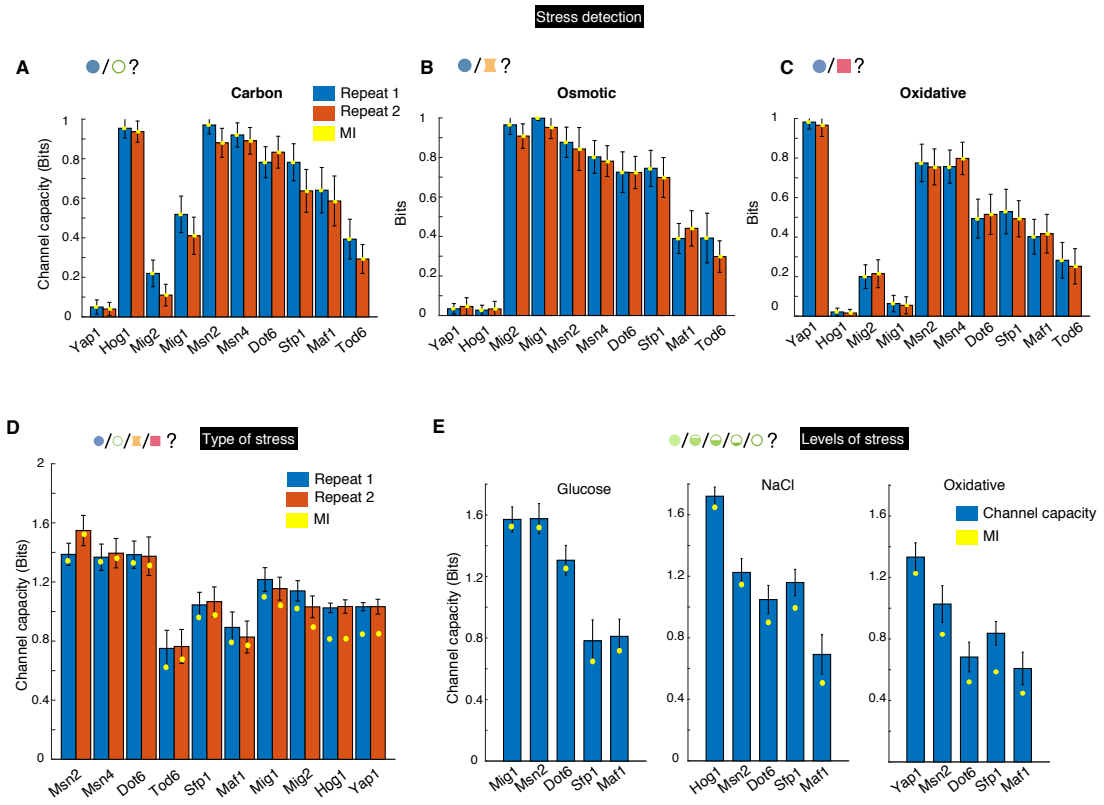

Fig. S18. The capacity is similar to the mutual information calculated with a uniform distribution of inputs. **A**, **B**, **C** The capacity calculated for the $q = 2$ data from Fig. 2 for transitions from rich media into high stress is almost identical to the mutual information calculated with a uniform distribution (yellow dots). **D** For specialists, but not most generalists, the capacity for the $q = 4$ data from Fig. 3A is higher than the mutual information calculated with a uniform distribution (yellow dots). **E** For different levels of stress ($q = 5$ data from Fig. 3C), the difference between the capacity and the mutual information calculated with a uniform distribution is highest for oxidative stress.

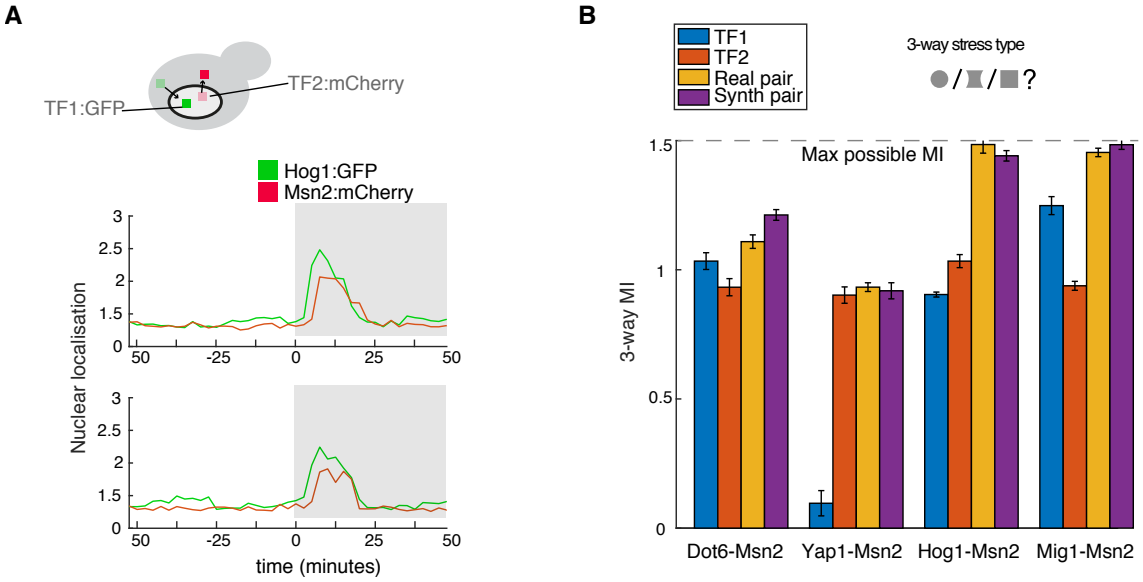

Fig. S19. Transcription factors have dynamics that are approximately independent given a particular environment. **A** We created strains where two transcription factors are simultaneously monitored using two fluorescent tags. An example (Hog1:GFP and Msn2:mCherry) is shown for two cells in osmotic stress (0.4M NaCl). **B** Estimates of mutual information are consistent with the transcription factor dynamics being conditionally independent given the state of the environment. For a 3-state environment (rich media, 0.4M NaCl, and 0.1% glucose), we show the mutual information calculated for each transcription factor independently, by concatenating the time-series for the pair of transcription factors monitored in the same cell, and by the concatenating the time-series for the pair when the trajectories are from different cells (created by randomly shuffling the time-series between cells and labelled 'synthetic pair').

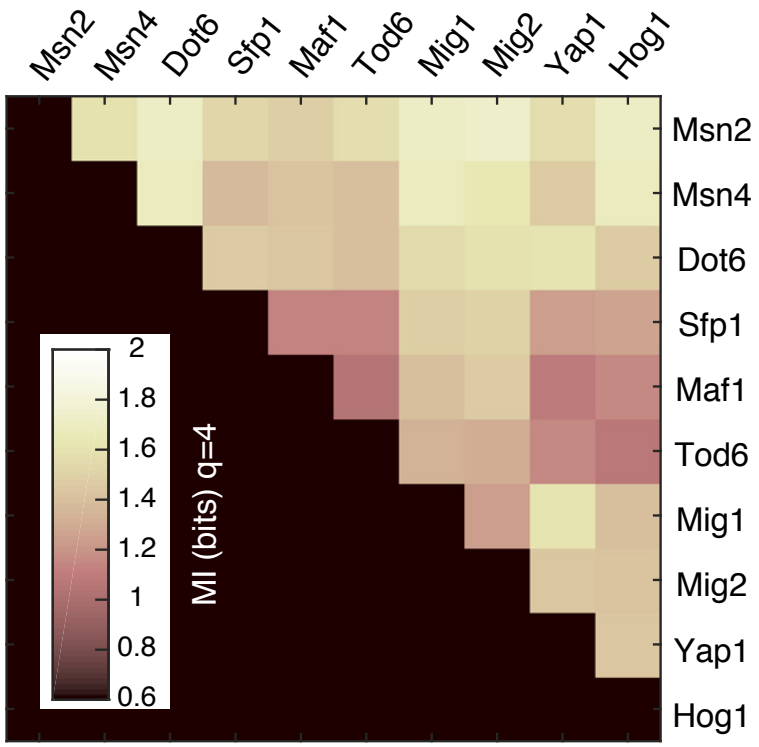

Fig. S20. Mutual information for pairs of transcription factors. The mutual information for the stress-type experiments ($q = 4$) from Fig. 3A for all combinations of two transcription factors. For each pair, the mutual information is calculated from concatenated time-series. These values are an average of 2 independent datasets (a total of 6 experiments per transcription factor). Highly informative pairs are Msn2/4 or Dot6 with a specialist, Msn2/4 with Dot6, and two specialists.

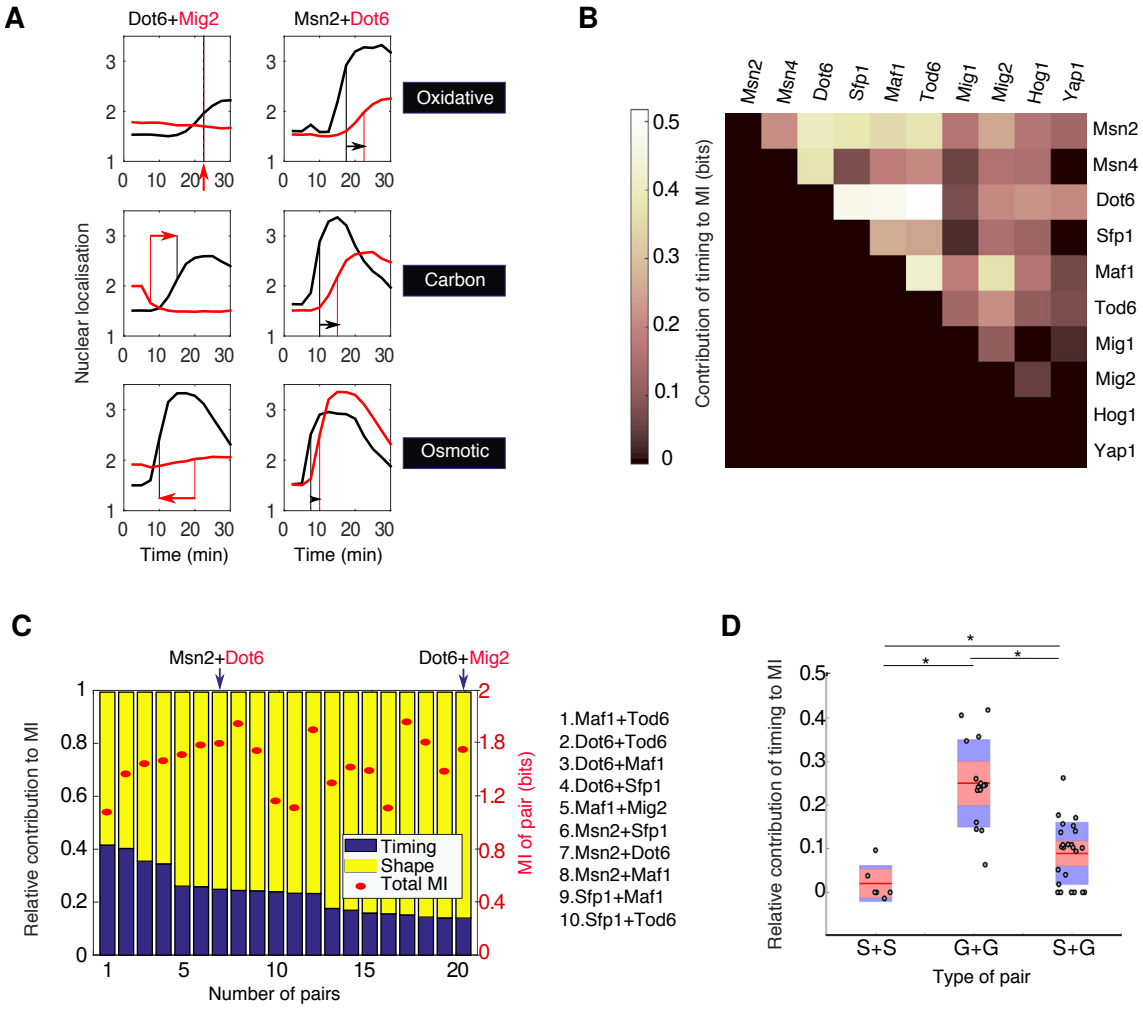

Fig. S21. Generalists often use relative timing between the responses of a pair of transcription factors to encode information. **A** Two examples illustrating differences in relative timing between a pair of transcription factors. The vertical lines indicate the time at which 50% of the maximum nuclear localization is reached and differences in these response times are shown by arrows. Although there is no difference in oxidative stress, Mig2 precedes Dot6 in carbon stress but lags in osmotic stress. **B** By removing the relative timing (by aligning the mean responses of both transcription factors), we can estimate the contribution of the relative timing to the mutual information (for the $q = 4$ environment). We show the difference between the mutual information of the pairs and the mutual information of the pairs after alignment (an average of 2 datasets with 6 experiments per transcription factor). **C** After removing the relative timing, the remaining information can be attributed to the shape of the responses. We show the top 20 pairs (the names of the top 10 are given) for which timing is most important. The total mutual information of the pair is shown as a red dot (right axis). Arrows indicate the two pairs shown in **A**. **D** Pairs of transcription factors involving generalists (G) encode more information in the relative timing of their responses compared to pairs of specialists (S). The mean for each group is shown with a red line (the 95% confidence interval is shaded red and the standard deviation is shaded blue). The star indicates a statistical significance of $p < 0.001$ using a Kolmogorov-Smirnov test.

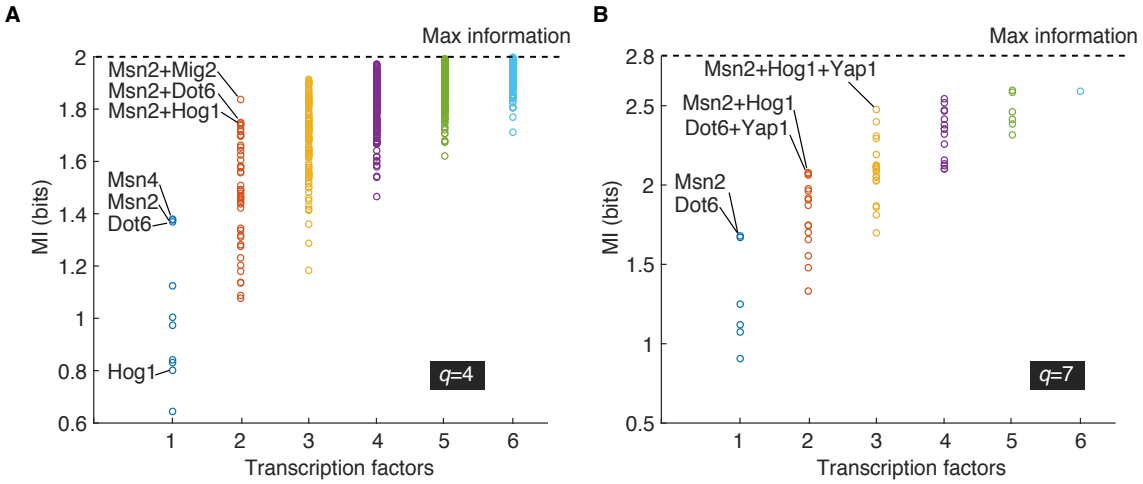

Fig. S22. Four transcription factors can encode almost all of the available information. **A** Mutual information for the stress-type experiments ($q = 4$) from Fig. 3A for all combinations of 10 transcription factors. **B** Mutual information for the $q = 7$ environment for all combinations of 6 transcription factors: Msn2, Dot6, Sfp1, Maf1, Hog1, and Yap1. This data is used for Fig. 4E.

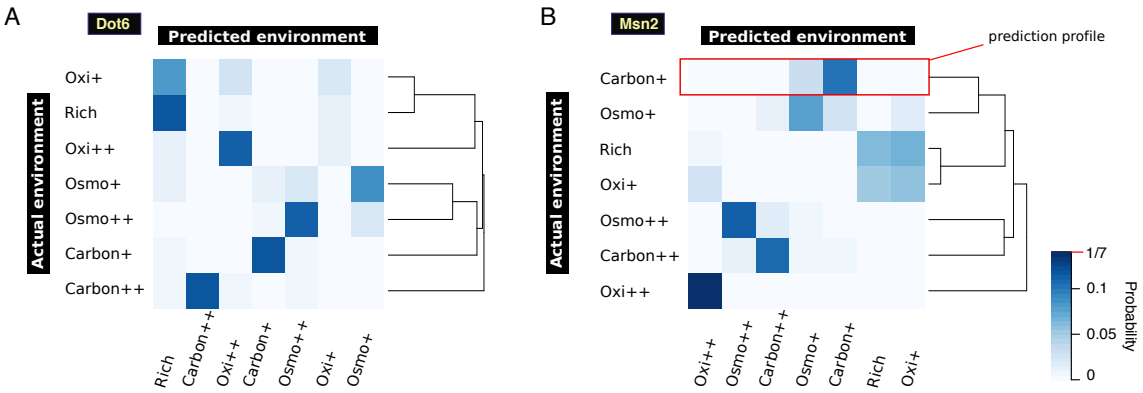

Fig. S23. Similarities between the prediction probabilities for specific environments indicate that generalists respond to categories of environments. The confusion matrix is shown for a $q = 7$ environment (rich media and two levels of each stress, where low is denoted + and high is denoted ++), where each row corresponds to one environmental state and the columns indicate the probability of predicting each of the possible environmental states from a single-cell time-series. Rows with similar prediction profiles (red square) indicate environmental states that the transcription factor treats similarly, and the dendrogram shows how environmental states are grouped into categories according to these prediction profiles. For this analysis, we combined datasets from Fig. 3A & 3C by selecting two levels of stress for each type of stress (the highest and the second lowest because the lowest levels showed no response for some transcription factors). **A** For Dot6, a highly-informative generalist, low oxidative stress (Oxi+) and rich media (Rich) have similar prediction profiles and are grouped together in the dendrogram. **B** Msn2 classifies low osmotic stress (Osmo+) and low carbon stress (Carbon+) in the same environmental category, and high oxidative stress (Oxi++) is in its own category.