Unstressed cells are alike, but stressed cells differ: Environmental and single-cell heterogeneity in yeast stress responses

Rachel Eder[1,2+], Leandra Brettner[1+], Kerry Geiler-Samerotte[1,2*]

[1]Center for Mechanisms of Evolution, Biodesign Institutes, Arizona State University, Tempe, AZ, USA

[2]School of Life Sciences, Arizona State University, Tempe, AZ, USA

[*]Corresponding Author
Email: kerry.samerotte@asu.edu (KGS)

[+]These authors contributed equally to this work.

Abstract

Cellular stress responses are central to survival, adaptation, and disease pathology, yet much of what we know comes from averages that mask heterogeneity. Using high-throughput single-cell RNA sequencing in *Saccharomyces cerevisiae*, we profiled approximately 13,000 cells exposed to diverse stresses. By annotating transposable elements (TEs) alongside coding genes, we captured expression across over 6,600 features. Leveraging these rich single-cell data, we redefined what it means to identify a “stress response,” shifting from the strongest average signals to the features that best predict whether an individual cell has experienced stress. These distinct definitions revealed different gene sets, underscoring how average versus single-cell views diverge. We found that stress-induced programs are highly condition-specific and perform poorly at identifying whether a cell has experienced stress, even within the environment where they were defined. By contrast, stress-repressed programs, dominated by ribosome-related genes, remain predictive across datasets, revealing a stable signature that distinguishes stressed from unstressed cells. This asymmetry recalls the Anna Karenina principle: Unstressed (“happy”) cells resemble one another in shared growth programs, while stressed (“unhappy”) cells diverge within and across environments. The divergence we observe does not always manifest as a continuum. We found stressed cells separating into subgroups that pursue alternative transcriptional responses, with some mounting canonical stress programs and others diverting transcription to TEs. Such mutually exclusive behaviors may highlight decision points that may influence survival, genome stability, and evolutionary trajectories. However, at the population level, canonical programs and TEs appear positively correlated since both increase under stress. At the single-cell level, they are inversely related, with cells seemingly committing to one program or the other at any given time. This false signal illustrates how averaging can be not only obscuring but also misleading and why single-cell resolution is required to understand stress responses.

Introduction

Cells encounter a wide range of environmental and physiological stresses, and decades of work have established that cells mount coordinated transcriptional programs in response. In budding yeast, these studies led to the description of the Environmental Stress Response (ESR), a stereotyped program characterized by induction of chaperones and other protective factors and repression of growth-related processes such as ribosome biogenesis [1–5]. Originally defined from bulk transcriptomics, the ESR has been an invaluable framework for stress biology. This canonical response continues to serve as a barometer of stress across diverse studies, guiding interpretation of gene expression data in yeast and other systems.

Yet, bulk measurements also have limitations due to how they average across thousands or millions of cells, emphasizing the strongest signals and potentially obscuring heterogeneity. The advent of single-cell methods has revealed that stress responses are far less uniform than once appreciated. In yeast, pioneering single-cell studies have revealed variability in the activation of stress-associated transcripts [6–14], while studies in cancer [15] and bacteria [16,17] show that heterogeneity in stress responses is a general phenomenon. These insights raise fundamental questions: How do the stress programs inferred from population averages differ from the diverse and sometimes divergent transcriptional strategies of single cells? Given this heterogeneity, to what extent can we predict whether a cell has experienced stress based on its transcriptional state?

Progress in answering these questions has been limited by scale. Historically, single-cell studies of stress in yeast have either profiled only a few hundred cells, targeted a small number of genes, or focused on a single stressor. Consequently, we lack a comprehensive view of stress responses across diverse environments for large cell populations. Yeast is uniquely well suited to fill this gap, because it combines rich prior knowledge of canonical stress biology [1,2], ultra high-throughput single-cell methods [18], and conserved eukaryotic stress phenomena such as transposable element upregulation [19–21]. These features make yeast not only an excellent reference model but also a valuable system for uncovering stress mechanisms relevant to other eukaryotes.

Here, we leverage these advantages by applying an ultra high-throughput single-cell transcriptomics platform [18,22] to profile approximately 13,000 yeast (*Saccharomyces cerevisiae*) cells exposed to diverse stresses, including protein misfolding and nutrient limitation. Each cell yielded thousands of transcripts, and by annotating transposable elements (TEs) alongside coding genes, we captured expression across more than 6,600 features. This scale exceeds earlier yeast single-cell studies and provides an exceptionally rich dataset, allowing us to use analytic tools rarely accessible in microbial single-cell work. For example, we apply machine learning to identify *de novo* stress-response programs that are obscured by population averages, and we use clustering to determine whether those programs are uniformly activated or restricted to subpopulations. In parallel, we use the ESR as a well-established barometer of stress, providing a fixed reference to track canonical induction and repression across conditions and single cells. Together, this combination of scale, analytic power, and a trusted measure of stress allows us to probe transcriptional heterogeneity both across environments and within stressed populations.

Our high-throughput, single-cell approach reveals patterns that bulk measurements miss. One is a striking asymmetry reminiscent of the Anna Karenina principle [23]. Just as Tolstoy wrote that “Happy families are alike; every unhappy family is unhappy in its own way,” unstressed cells

resemble one another in their uniform activation of growth programs, whereas stressed cells diverge in their transcriptional programs. As a result, stress-induced genes, long used as population-level markers, fail to predict stress exposure at the single-cell level. Another critical insight from our single-cell analysis is that it exposes an additional layer of divergence within stressed populations. Some cells mount canonical stress programs that include traditional components like chaperones, while others divert much of their transcriptional output to TEs. Bulk analyses not only obscure this negative relationship but can invert it, creating the illusion of a positive correlation between TEs and canonical stress response induction. This paradox highlights that averaging does not merely blur heterogeneity; it can produce misleading conclusions. If the very definition of a “stress response” has been built on averages, then single-cell studies challenge us to reconsider what stress really means, not as a singular program, but as a mosaic of strategies, discernible only at the single-cell level.

Results

All stressors surveyed elicit a canonical stress response called the iESR

The first stress we studied was the misfolding of a gratuitous, heterologous protein. Yeast responses to misfolded proteins have previously been characterized in bulk [24–27], but here we examine these responses in finer detail by using single-cell transcriptomics. To see if single cells transcriptionally respond to varying levels of misfolding stress, we utilized strains that express increasingly misfolded variants of yellow fluorescent protein (YFP) [24,28]. We used three different strains, including a version where very few of the YFPs are misfolded (YFPwt), a version where almost every YFP molecule is misfolded (YFPm4), and a version that falls in between (YFPm3) (**Fig 1A**; top row, purple). As all three strains express YFP from the same promoter and at the same levels, the only appreciable difference between them is the fraction of YFP that misfolds [24,28].

To determine if single cells respond to YFP misfolding by mounting a transcriptional response, we grew yeast strains in conditions where YFP is expressed [24], performed single-cell RNAseq [18], and measured the percent of each single cell’s transcriptome devoted to the induced Environmental Stress Response (iESR) [1]. As expected, the median proportion of transcripts per cell devoted to the iESR increases from cells expressing YFPwt to those expressing YFPm4 (**Fig 1B**; top row). However, despite the misfolded protein burden, not to mention the shared metabolic burden of overexpressing YFPs, the percent of the detected transcriptome dedicated to the iESR is ~10% or less (**Fig 1B**; top row). Further, cell-to-cell distributions overlap considerably, with some YFPwt cells mounting stronger iESR responses than some YFPm4 cells (**Fig 1B**; each gray point represents one cell).

In addition to protein misfolding, we studied stress resulting from the limitation of two different carbon sources: glucose (**Fig 1**; middle row, blue) and sucrose (**Fig 1**; bottom row, red). For both types of nutrient limitation, we compared batch cultured cells sampled early in growth, when cells are replete with nutrients, to cells sampled later in growth, when cells are becoming starved (**Fig 1A**). As with misfolded protein stress, the percentage of the transcriptome devoted to the iESR increases over time as cells deplete the nutrients in their media (**Fig 1B**). These results confirm the utility of the iESR as a barometer for stress in many different environments, as have many previous studies [1,2,29,30]. However, we again noticed that the iESR did not comprise greater than ~10% of any cell’s transcriptome (**Fig 1B**). Also similar to what we observed under misfolding stress, the distributions of stressed and unstressed cells overlap, with some nutrient-limited cells exhibiting weaker iESR induction than some nutrient-replete cells (**Fig 1B**).

Not all stressed cells mount a transcriptional response to stress

The overlapping distributions in **Fig 1B** suggest that, unlike in studies using stronger stressors [29], transcriptional profiles do not reliably distinguish stressed from unstressed cells in our study. Indeed, when we rank order cells based on the percent of their transcriptome devoted to the iESR, our stressed (**Fig 1C**; darker colors) versus unstressed (**Fig 1C**; lighter colors) cells do not separate cleanly. For example, when we attempt to classify YFP-expressing cells as either YFPwt (1165 cells) or YFPm4 (1082 cells) by ordering them based on the percent transcriptome devoted to iESR, only 61.3% of YFPwt and 58.7% of YFPm4 cells fall into the expected category (**Fig 1C** and **Fig 1D**; row 1 first set of barplots). Using the same simple rank ordering procedure to identify which cells have experienced nutrient deprivation results in a slightly higher success rate (**Fig 1C**); we correctly identify 71.7% (glucose) and 69.3% (sucrose) of nutrient stressed cells (**Fig 1D**; rows 2 and 3 first set of barplots). Still, a significant fraction of cells, especially stressed cells, are incorrectly classified as coming from the unstressed condition in all three experiments (**Fig 1D**; first set of barplots, darker bars never exceed 75% and are always lower than lighter bars).

These observations raise a central question: Why do so many cells that were grown under stressful conditions fail to mount the canonical stress response? One possibility stems from the way the iESR was originally defined from bulk data, where only the strongest signals emerge after averaging. As a result, stress-responsive genes that change modestly or only in subsets of cells may have been overlooked. Another possibility is that each stressor provokes its own transcriptional program, which may be more informative than the iESR when viewed cell by cell. To explore these possibilities, we next re-defined stress-responsive gene sets directly from our single-cell data to capture the transcriptional programs activated under each stress.

Protein misfolding and nutrient deprivation provoke distinct transcriptional responses

To identify genes significantly upregulated under each stress condition, we applied two criteria: (i) the average fold change in expression between stressed and unstressed cells, and (ii) a significance score assessing whether this difference was observed across enough cells to be unlikely by chance ($p < 0.05$ after multiple-testing correction). This approach differs from bulk studies where significance is evaluated across a small number of replicate experiments rather than across hundreds or thousands of individual cells. By leveraging single-cell resolution, we gain the power to detect modest but consistent expression differences.

In each of the three stress conditions, we compared gene expression in cells experiencing the highest versus lowest levels of stress (darkest versus lightest colors in **Fig 1**). For protein misfolding, this analysis identified 124 genes with significantly greater expression in YFPm4 compared to YFPwt, which we term the “induced misfolding stress response” (iMSR) (**Supplemental Table 1**). Strikingly, only five of these genes overlap with the canonical iESR (**Fig 1D**; top row). As expected, the iMSR accounts for a larger fraction of each cell’s transcriptome than the iESR (up to 30% versus 10%, respectively) and, on average, increases progressively from YFPwt to YFPm3 to YFPm4 (**Fig S1**). Thirty-six of the iMSR genes are transposable elements (TEs), which are often not included in microarrays such as those used to define the iESR [1]. Notably, iMSR activation is better at predicting which cells are expressing YFPm4 versus YFPwt. When we perform an analysis similar to that in **Fig 1C**, where we rank order cells based on the percent of their transcriptome devoted to the iMSR, we correctly identify 72.2% YFPm4-expressing cells (as opposed to 58.7% when using iESR activation in

Fig 1D). A similar improvement is seen when using the iMSR, rather than the iESR, to classify cells as expressing YFPm3 versus YFPwt (**Fig S3C**; top row).

For nutrient limitation, we again observed little overlap between the canonical iESR and the stress-specific programs we defined. Using the single-cell significance criteria explained above, we identified an “induced glucose stress response” (iGSR) and an “induced sucrose stress response” (iSSR) (**Fig 1D**; Venn diagrams and **Supplemental Table 1**). These gene sets are largely distinct from both the iESR and from one another. Even across the three single-cell–defined responses (iMSR, iGSR, and iSSR), only 12 genes overlap, many of which are transposable elements. Across all four induced responses, just a single gene (*PRB1*, a vacuolar endoprotease) is shared. This leads us to conclude that different types of stressors provoke largely non-overlapping sets of stress-responsive genes, at least when these responses are defined in a way that captures small but consistent changes in gene expression across single cells.

While the responses we define are better at predicting stressed from unstressed cells, they are still not reliable predictors. Like the iMSR, the iGSR and iSSR each account for larger fractions of the transcriptome than the iESR, comprising up to 60% of total expression in cells experiencing the most severe nutrient deprivation (**Fig S1**). Despite this, classification based on these responses is imperfect with only 89.8% of glucose-limited cells and 71.2% of sucrose-limited cells identified correctly, leaving 10.2% and 28.8% misclassified, respectively (**Fig 1D**). These misclassified cells are not simply “borderline” cases. Several cells that experienced nutrient stress show very low iGSR or iSSR activation (see dark red and blue lines at the left edges of the distributions in **Figs 1C** and **S2**). Classification accuracy is even worse under less extreme stress (**Fig S3C**; rows 2 and 3).

Taken together, these results suggest that some cells with a history of stress are transcriptionally indistinguishable from unstressed ones, at least when stress responses are defined by stress-induced signals that emerge on average across cells. Our findings also show that each stressor provokes a distinctive transcriptional program that is more predictive of a cell’s provenance than the canonical iESR.

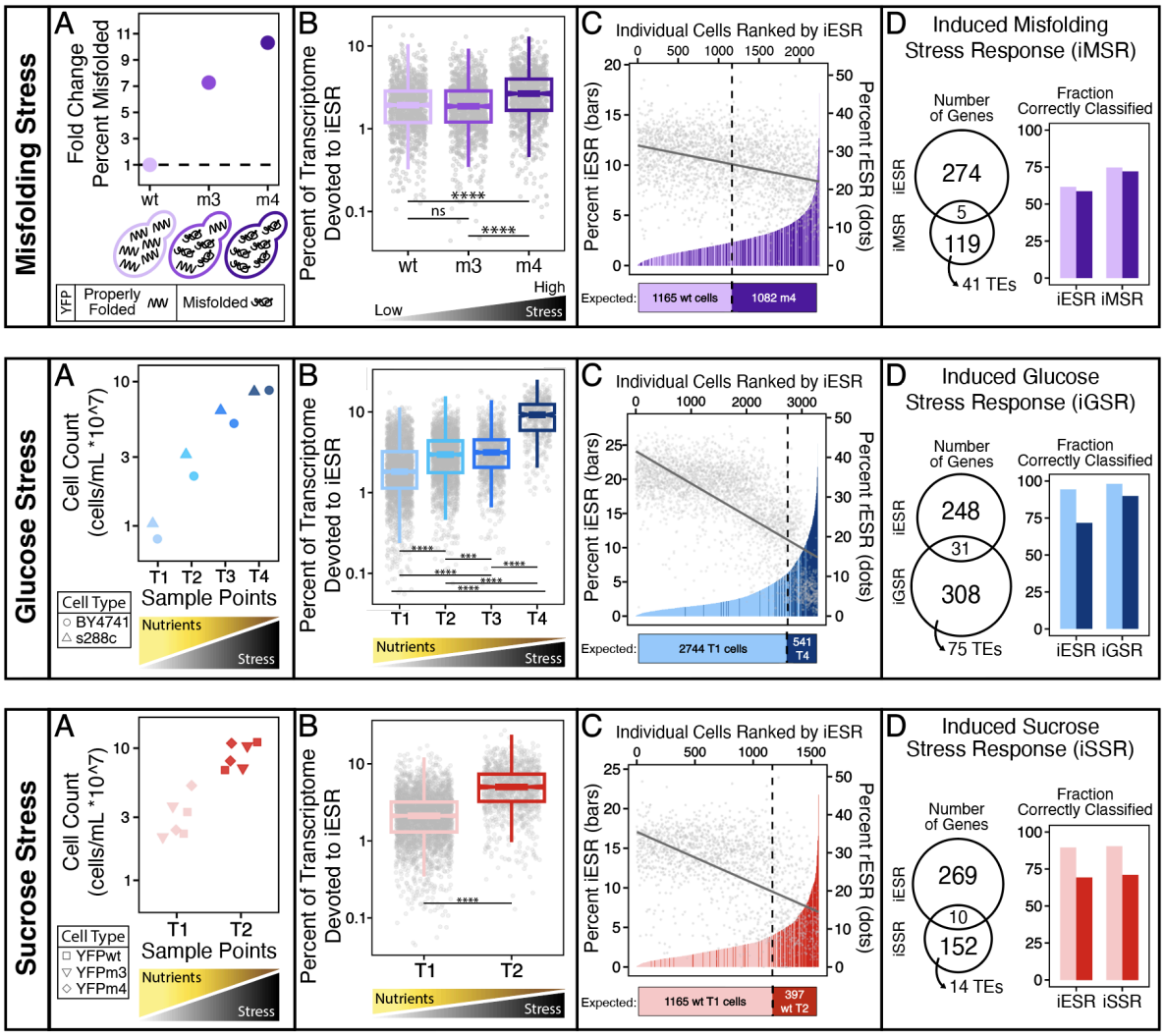

Fig 1: Transcriptional responses differ across stressors. (A, top) YFPm3 and YFPm4 expressing strains have 7-fold and 10-fold more misfolded protein than YFPwt, as measured using Western blot [24]. Darker purple indicates increasing misfolded protein stress. **(B, top)** The percent of the transcriptome devoted to the iESR increases with misfolding stress. Each dot represents a cell sampled in early log-linear growth; median and quartiles are shown across all cells from the same genetic background. Cells with no detected iESR were excluded. **** indicates $p < 0.0001$ with Wilcoxon signed-rank test; ns indicates no significance. **(C, top)** YFPwt and YFPm4 cells are ordered by the percent of the transcriptome dedicated to the iESR (bars). An inverse correlation is observed between the (induced) iESR (bars) and (repressed) rESR (points). The vertical dashed line signifies the number of unstressed cells, which we expect to fall on the left side of the line if, as in previous work [29], they induce less iESR than stressed cells. The horizontal bar below indicates the true proportion of YFPwt and YFPm4 cells. **(D, top)** Venn diagram shows only a small overlap in iESR and iMSR genes. Bars display the fraction of correctly classified YFPwt or YFPm4 cells based on the rank order of their iESR or iMSR levels. **(A, middle)** Darker blue indicates increasing glucose-limitation. Yeast strains were sampled over 4 timepoints in glucose-limited media. Nutrients are depleted as cells grow and reach higher cell counts. **(B–D, middle)** Similar to the top row with *** indicating $p < 0.001$ with Wilcoxon signed-rank test**. (A, bottom)** Darker red indicates increasing sucrose-limitation. Yeast strains were sampled over 2 timepoints in sucrose-limited media. **(B–D, bottom)** Similar to above rows with **panel C** displaying cells corresponding to a single genetic background.

Stress-induced genes vary by stressor and fail to predict cell provenance

The above results suggest that the transcriptional response to stress, specifically, the upregulation of certain genes, varies considerably across the different stressors we studied. To quantify this more directly, we measured the fraction of upregulated genes unique to each stress response. In our dataset, about 50%, and in some cases, up to 80% of the induced transcripts appear unique to a particular stress (**Fig 2A**; black bars).

So far, we have focused on these stress-induced transcripts, termed the iESR, iMSR, iGSR, and iSSR, where "i" denotes "induced." We can also examine stress-repressed transcripts (rESR, rMSR, rGSR, and rSSR), where "r" denotes "repressed." We find that genes in this category are enriched for cellular growth functions such as ribosome biogenesis, consistent with previous studies defining the rESR [1,2]. As expected, their expression anticorrelates with induced genes (**Fig 1C** and **S2**).

Unlike the induced gene sets, the repressed responses are less stress-specific (**Fig 2A**; yellow bars). Although more genes overall are significantly downregulated during stress, relatively few are unique to any one condition. In other words, stress induction tends to be diverse and condition-specific, while stress repression is comparatively conserved. Stressed cells reliably downregulate similar growth-related transcripts, regardless of the stressor. This observation recalls the Anna Karenina principle [23]: All unstressed cells "prosper" in the same way, upregulating growth-related programs, while stressed cells are "unhappy" in diverse ways, each upregulating different stress-induced programs. If this is true, then conserved repression should act as a reliable marker of unstressed cells across datasets.

Indeed, when we train an ensemble machine learning model (Random Forest [31]) on one dataset and test it on another, the model consistently identifies unstressed cells with over 90% accuracy (**Fig 2B**; yellow points are high on the vertical axis). This cross-dataset predictability held no matter which dataset was used for training versus testing. By contrast, the models often struggled to identify cells that had experienced stress, with accuracy dropping below 25% in some cases (**Fig 2B**; black points are low on the vertical axis).

The critical role of repressed genes is further highlighted when we remove either induced or repressed gene sets from every model. Excluding induced sets has little effect, while excluding repressed sets sharply reduces performance, particularly for identifying stressed cells (**Fig 2C**). Across stressors, what cells turn on under stress is relatively variable and poor for cross-condition prediction, whereas what they turn off when stressed remains conserved.

This pattern is also apparent in **Figs 1C** and **1D**, where the fraction of correctly identified unstressed cells (lighter bar) is always higher than the fraction of correctly classified stressed cells (darker bar). This is true for every dataset we studied, regardless of the type or magnitude of the stress (**Figs S4C** and **S5C**). For more extreme stressors, the fraction correctly classified for unstressed cells nears a perfect percentage (**Fig S4**; rows 2 and 3). These findings further support that stress-repressed genes are more indicative of a cell's provenance than are stress-induced genes. This conclusion supports a model in which stress-response programs comprise a conserved repression module plus flexible induction modules that reflect the nature of the stress [6].

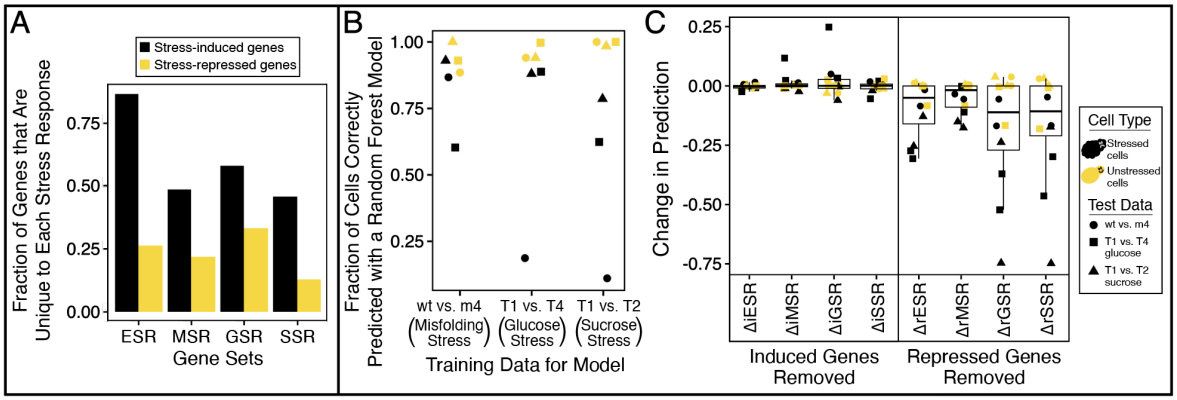

Fig 2: More variability in stress-induced genes make stressed cells harder to predict. (A) Each bar displays the fraction of stress-responsive genes that are unique to a given induced (black) or repressed (yellow) stress response. Stress-induced and stress-repressed genes were compared across the ESR, MSR, GSR, and SSR. Genes that only appeared in one of the induced/repressed gene sets are considered unique to that stress response. The number of unique genes for a gene set were divided by the total number of genes in that set to calculate the fraction of genes unique to each stress response. **(B)** A Random Forest [31] model can correctly predict which cells are unstressed (yellow) more often than which cells have experienced stress (black). Categories on the horizontal axis correspond to the training data used to build a predictive model. The fraction of the test cells correctly predicted are plotted on the vertical axis. The shape of the points correspond to the test data, which indicate either an estimated out-of-bag error rate of within model accuracy, or were cells from a different dataset, as indicated by the legend. **(C)** The change in the prediction accuracy was calculated after removing either the induced or repressed gene sets from each training model and rerunning the analysis in **panel B**. Removing stress-repressed genes (right) decreases prediction accuracy more than removing stress-induced genes (left), and predictions are again more difficult for stressed cells (black) relative to unstressed cells (yellow).

Cells experiencing the same stress can upregulate distinct transcriptional programs

Having shown that transcripts induced by stress differ across stressors, we turned our attention to differences between single cells exposed to the same stressor. Cell subpopulations are readily apparent in our dataset due to the large number of genes and cells profiled. In our most stressful condition as defined by average iESR upregulation (timepoint 4 under glucose limitation), most single cells indeed upregulate the iESR (**Fig 3A**; cluster 2). Yet a distinct subpopulation identified by Louvain clustering [32] does not upregulate the iESR, despite experiencing the same glucose limitation stress (**Fig 3A**; cluster 1). This small but interesting subpopulation would be lost in studies that rely on population averages (bulk transcriptomics) or studies limited to smaller numbers of single cells. Since these cells in cluster 1 have relatively low levels of the iESR, one might assume that they have relatively high levels of the rESR because these responses are anticorrelated in our and others' work (**Fig 1C**) [1,2,29]. However, we do not observe substantial rESR activation either (**Fig 3A**; middle panel). Instead, the iESR plus rESR together account for less than 20% of the transcriptome in most cluster 1 cells, indicating that this subpopulation is defined by an alternative transcriptional program.

Reflecting this divergence, we find that cluster 1 cells are highly enriched for transposable element (TE) expression, such that much of their transcriptome is devoted to TEs rather than canonical stress responses (**Fig 3A**; right panel). This observation, and the observation that all three stress responses we identified in **Fig 1D** (iMSR, iSSR, and iGSR) comprise TEs, led us to

study the relationship between the iESR and TEs more broadly across all cells and environments.

Similarly to the iESR, we find that the fraction of the transcriptome composed of TEs increases with stress (**Fig S6**). This observation is consistent with previous findings across various stressors [33][34,35][20,21]. Over all datasets we study, the average levels of TEs and iESR transcripts are positively correlated (**Fig 3B**). This might easily lead to the false assumption that stressed cells possess *both* more TEs and higher levels of iESR transcripts. However, this is not true. At the single-cell level, TEs and the iESR are often negatively correlated (**Fig 3C** and **Fig S7**). Cells that have high levels of TEs have low levels of iESR transcripts and vice versa. This negative correlation is true of nearly every sample of single cells that we collected regardless of strain background or stress level and persists for not only TEs versus the iESR but also TEs versus the iMSR, iGSR, and iSSR (**Fig 3D** and **Fig S8**). Oppositely, when we randomly divide the iESR or other induced stress responses in two and compare each half, we see positive correlations (**Fig 3D**; all points in the gray boxplots are above zero), supporting the idea that negative correlations between gene sets are not artifacts of our method. On the contrary, the sparsity of single-cell RNA sequencing data tends to minimize detectable correlations, thus the negative relationship between TE and iESR expression may be even stronger than we can reliably detect.

Why would some stressed cells upregulate transposable elements (TEs) instead of mounting a canonical stress response? One idea comes from the fact that the iESR includes many protein-folding chaperones, which are upregulated during stress across the tree of life [36]. Prior work suggests that at least one of these chaperones, Hsp90, contributes to TE repression. By stabilizing Sir2, a histone deacetylase, Hsp90 helps maintain heterochromatin and silence TEs [37–39]. When stress sequesters chaperones, Sir2 activity declines, euchromatin becomes more accessible, and TEs are de-repressed [19]. Analogous mechanisms are seen in other organisms, where TE silencing relies on Hsp90-stabilized piRNA pathways [40,41]. Therefore, the persistent anticorrelation we observe between stress-induced gene and TE expression (**Fig 3D**) could be related to the role of chaperones in silencing TEs.

The potential connection between TEs and stress responsive genes is intriguing, because TE activation and stress response dysregulation have both separately been implicated in aging, neurodegeneration, and protein misfolding diseases, including Alzheimer's disease [42,43]. However, our data do not provide information about the protein level and should instead be interpreted as a transcriptomic signal that motivates further investigation. A key, more general point, is that stressed cells can adopt divergent transcriptional programs, either mounting canonical stress responses or diverting expression to TE activity. This heterogeneity is obscured when transcriptional data are averaged across cells (**Fig 3E**), reinforcing a central conclusion of our work that stress responses are not necessarily uniform programs but a mosaic of transcriptional outcomes, whose diversity only becomes apparent at single-cell resolution.

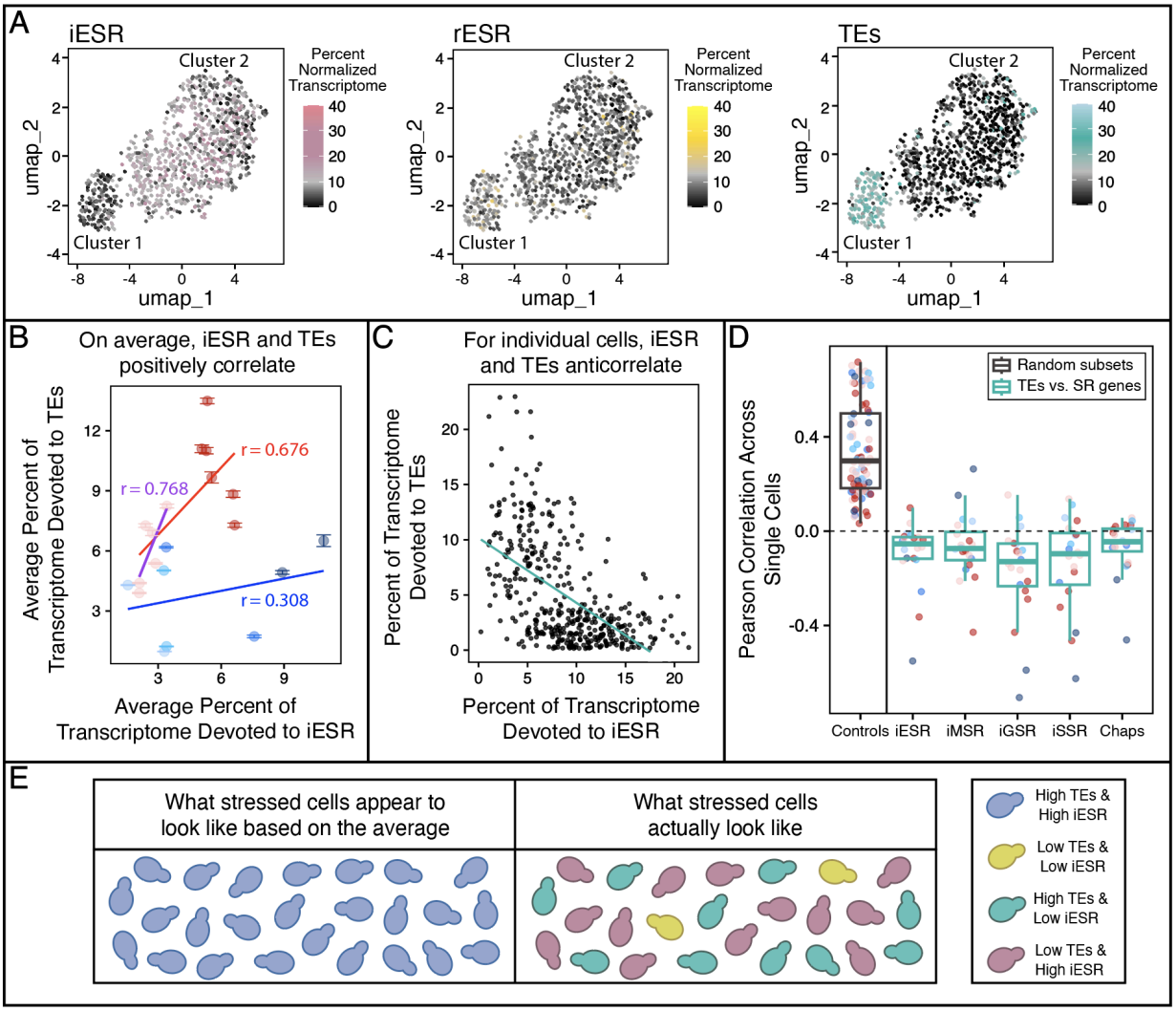

Fig 3: Anticorrelation of TEs and iESR expression at single-cell level. (A) Two distinct transcriptional clusters, or subpopulations, of cells from timepoint 4 (T4) in glucose-limited growth were found by using a Louvain clustering analysis [32]. UMAPs were generated for visualization purposes. The plots are colored by the percent normalized transcriptome for iESR, rESR, and TE gene sets. **(B)** The average percent of the transcriptome devoted to the iESR versus TEs is plotted for each strain in each condition from which we sampled transcriptomes (20 total). Points are colored the same as (**Fig 1A**; rows 2 and 3), where red corresponds to scRNAseq data from cells grown in sucrose, and blue corresponds to data from cells grown in glucose. The additional purple trend bar corresponds to the three YFP strains (wt, m3, and m4) grown in sucrose and sampled at T1 (**Fig 1A**; row 1). Error bars display the standard error of the mean across replicate flasks that were grown on the same day. Pearson correlations between the iESR and TEs are positive at the population level. Cells with no counts detected for the iESR or TEs were excluded. **(C)** For a single yeast strain (s288c) in T4 in glucose-limited growth, the correlation between the percent of the transcriptome devoted to the iESR versus TEs is displayed across individual cells. The cyan line displays a negative correlation between the iESR and TEs at the single-cell level. Cells with no counts detected for the iESR or TEs were excluded. **(D)** Expression of all TEs versus other stress-responsive (SR) genes (excluding TEs) tends to be negatively correlated across single cells. Points are colored as in **Fig 1A**. Each point represents the correlation across all single cells from a given sample (see also **Fig S8**). Cyan boxplots summarize the Pearson correlation across all 20 samples, while the gray boxplot summarizes 80 paired control experiments where we randomly split each stress response in half and quantified how expression of randomly chosen gene sets correlates across single cells. “Chaps” are chaperones as defined by [44]. **(E)** A schematic example shows a misconception of population-averaged

data where stressed cells may be assumed to have high TEs and high iESR based on the results of **Fig 3B**, when, in actuality, they have either one or the other (based on the results of **Fig 3C** and **3D**).

Population versus single-cell approaches reveal different views of stress responses

Single-cell data provide new ways to measure differences in gene expression. In traditional bulk analyses, statistical significance is calculated across replicate populations, so large fold changes that appear in the population average are most likely to be detected. By contrast, in single-cell analyses, each cell serves as an independent observation, which increases power to detect very subtle changes, as long as they occur consistently across many cells. This means the two approaches are biased in different ways and neither may adequately capture the diversity of transcriptional states in a cell population. Population-level methods tend to highlight genes with strong changes that may be driven by only a subset of cells, whereas single-cell methods preferentially detect small but consistent shifts across the entire population. As a result, each approach can emphasize different sets of stress-responsive genes. Indeed, one reason for the limited overlap between the stress-induced genes detected in our study and the classic iESR (**Fig 1D**) is likely methodological. Yet, we also observed very little overlap between the iMSR, iGSR, and iSSR, which were all detected through the same single-cell approach, suggesting these differences between induced genes in differing stressors are biologically meaningful.

To directly compare how methodological factors contribute to the identification of stress-responsive genes, we re-analyzed our data using three approaches: (i) a fold-change analysis that does not account for significance across samples or cells (**Fig 4A**), (ii) a single-cell significance test based on the fraction of cells showing gene upregulation or downregulation (**Fig 4B**), and (iii) a Random Forest model that quantifies the predictive importance of individual genes (**Fig 4C**). To implement a fold-change analysis, we pooled reads across all stressed or unstressed cells corresponding to a given condition, and we calculated the fold-change in the level of each transcript across stressed versus unstressed populations (**Fig 4A**). This analysis identified iESR genes as significantly upregulated across all stresses studied relative to all genes (**Fig 4A**; black boxplot). Although some inconsistencies emerged across our unique stressors, such as strong fold changes in non-iESR genes and occasional downregulation of iESR genes, these results largely matched previous bulk transcriptomics ([1,30]), supporting consistent iESR induction at the population level.

By contrast, the single-cell significance approach (**Fig 4B**) and Random Forest models (**Fig 4C**) did not highlight iESR genes as especially informative. Except under glucose stress in **Fig 4C**, iESR genes were statistically indistinguishable from random gene sets of equal size. In both analyses, boxplots summarizing the iESR genes were often centered around zero (**Fig 4B** and **4C**). This suggests that iESR induction is not consistent enough across individual cells to register as significant (**Fig 4B**) or to be useful for prediction (**Fig 4C**), even though it is strong enough to shift the population average (**Fig 4A**). Moving beyond the iESR, genes with the largest population-level fold changes have less predictive value at the single-cell level (pink dots are high on the vertical axis in **Fig 4A** but sometimes drop to near zero in **panels 4B** and **4C**).

Some conclusions from bulk transcriptomics may need to be revisited in light of these findings. Bulk studies inherently emphasize the strongest average signals (i.e., genes with large fold changes across the population), even if those changes come from only a subset of cells. By contrast, single-cell methods highlight genes that shift modestly but consistently across many cells. Each approach therefore captures different aspects of cellular physiology, with bulk identifying the “loudest” responders, while single-cell uncovers subtler but widespread changes.

The fact that each method detects different sets of genes (**Fig 4**), and that sets of genes defined by the same method differ across stresses (**Figs 1** and **2**), together support a conclusion in line with previous work [12,29,45], namely that stress responses are more heterogeneous than we currently know how to measure, both across single cells and across stressors. This heterogeneity itself scales with stress intensity. As average iESR expression rises, its variability decreases, while repression variability increases (**Figure S9**). Understanding these complexities will require rethinking how we define, detect, and interpret stress itself.

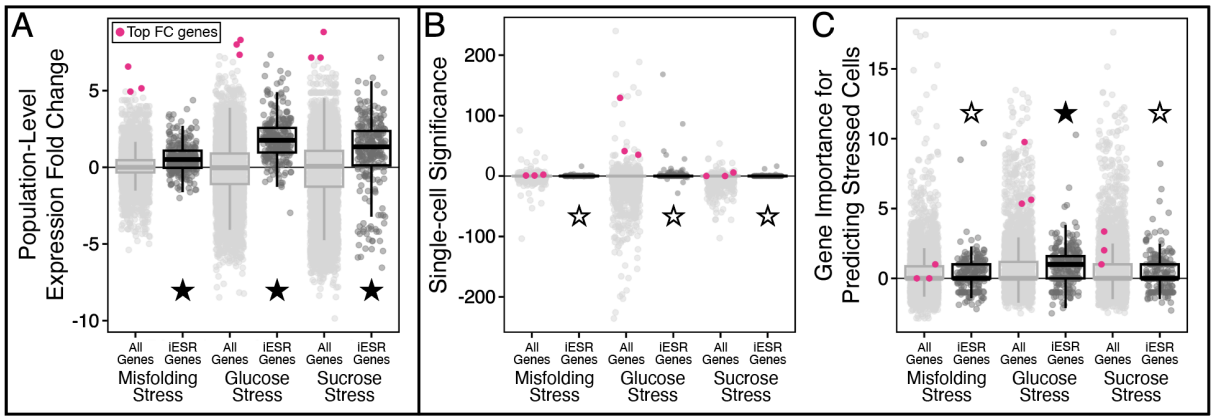

Fig 4: Different approaches detect stress-responsive genes. (A) Population-level log2 fold change (FC) measured between stressed and unstressed cells is plotted for all genes (light gray points) and for iESR genes (black points). A positive FC corresponds with genes upregulated in stressed cells. The three genes with the highest FC values in **A** are highlighted in pink across **A–C**. Symbol ★ signifies that across 10,000 iterations, the median FC of the iESR exceeded that of all randomly selected gene sets of equal size. **(B)** Single-cell significance scores (+/-log10(corrected p-value)) for gene expression differences between stressed and unstressed cells were plotted for all genes and iESR genes. Genes downregulated in stressed conditions were plotted with a negative sign change of their p-values. The farther away a corrected p-value is away from zero, the more significant it is. Because of the p-value corrections, boxplot medians fall at zero (p = 1) for all conditions. Symbol ☆ signifies that across 10,000 iterations, the median score of the iESR did not exceed that of random gene sets of equal size. **(C)** Each point represents a gene used during the construction of a Random Forest model for each stressor. The mean decrease in model accuracy upon excluding that gene is plotted. Symbol ★ signifies that across 10,000 iterations, the median FC of the iESR exceeded that of over 99% of randomly selected gene sets of equal size. Symbol ☆ signifies that across 10,000 iterations, the median score of the iESR did not exceed that of random gene sets of equal size or fell below some, meaning that some random gene sets had a higher median MDA for predicting stressed cells than did the iESR.

Discussion

Our study shows that high-throughput single-cell transcriptomics uncovers dimensions of stress-response heterogeneity that bulk methods overlook or sometimes misrepresent. By profiling over 13,000 yeast cells across protein misfolding and nutrient limitation, we reveal environmental and cell-to-cell heterogeneity in stress responses.

We found that stress-induced transcriptional programs are not reliable predictors of whether a cell has experienced a stressful environment. Even the canonical iESR, which is detectably upregulated across all stresses, does not reliably encode a cell's environmental history. By contrast, stress-repressed transcriptional programs are more reliable. Genes tied to growth and ribosome biogenesis are consistently downregulated, regardless of stress type, and this repression strongly supports prediction of cell provenance. Put simply, while cells vary widely in

the protective programs they activate, they converge on a common strategy of turning off growth when challenged. This asymmetry recalls the Anna Karenina principle, where, in our case, unstressed cells resemble each other in their growth programs, but stressed cells diverge, across stressors and within populations. Our cross-dataset classifier models support this view, reliably identifying unstressed cells but struggling to predict stressed ones.

Single-cell resolution also exposes an intriguing divergence within stressed populations. Subsets of cells mount expected stress programs, which include the upregulation of chaperones, while other cells devote large fractions of their transcriptome to transposable elements (TEs). At the population level, these responses to stress appear correlated, yet cell-level analyses reveal them to be anticorrelated such that cells are largely mounting one or the other. Bulk data not only hide this negative correlation but even invert it, underscoring how averaging can distort the biology it seeks to summarize.

Methodological comparisons further reinforce this message. Bulk fold-change analysis highlights the iESR as a reliable indicator of stress, whereas single-cell significance tests and predictive models show that iESR genes vary across individual cells and are not especially informative for prediction. As two approaches capture different signals, population-level amplitude versus cell-level consistency, stress programs defined from bulk data do not necessarily predict stress at the single-cell level.

Taken together, these results support the power of single-cell transcriptomics for understanding that stress responses are not uniform. Instead, they are diverse across environments and heterogeneous within cell populations. Recognizing this mosaic nature will be key for understanding and predicting how cells survive stress, how populations evolve, and how stress contributes to pathology.

Methods

Yeast strains

Saccharomyces cerevisiae strain *MATa ura3Δ0 his3Δ1 met17Δ0 PACT1-GAL3::SpHIS5 gal1Δ gal10Δ::LEU2 leu2Δ0::PGAL1-YFP-KanMX6* with wt, m3, and m4 YFP mutations was used for intrinsic misfolding and extrinsic nutrient deprivation stress experiments in sucrose/raffinose-limited media [24]. We also used a diploid s288c and a haploid BY4741 for glucose nutrient deprivation experiments.

Yeast cell culture

Cells were streaked on YP plus 2% dextrose agar plates from frozen -80°C glycerol stocks and grown at 30°C for 48 hours. Colonies were then picked into liquid YP plus either 2% dextrose (for glucose limitation experiments) or 2.5% sucrose, 1.25% raffinose, and 0.625% galactose (for misfolding and sucrose limitation experiments). These cultures were grown at 30°C with shaking overnight. Then they were massively diluted into fresh media to initiate a growth experiment. This was done by measuring the overnight culture density and adding approximately 15,000 total cells to 50 mL of media (~300 cells/mL) in beveled flasks. These cultures were then subsequently grown at 30°C with shaking, with cell counts measured systematically over time using a Coulter cell counter until the populations started to exhaust nutrients and enter stationary phase. At various timepoints throughout growth, cells were sampled, fixed, and prepared for single-cell RNA sequencing as previously described in [18].

The specific media used for growth, number of replicates, and the timepoints at which cells were sampled for scRNAseq differed depending on the stress being studied, as follows:

For the three YFP strains used to study stress from misfolded proteins, two independent experiments were performed on separate days. One of these experiments included one replicate flask per strain, while the other included two replicates per strain, generating a total of three biological replicates for each strain. After the overnight shaking described above, the strains were massively diluted into Synthetic Complete minus glucose (Sunrise) plus 2.5% sucrose, 1.25% raffinose, and 0.625% galactose media to induce YFP expression. Strains were allowed to grow exponentially in this media for at least 16 hours to assure YFP induction, as in previous work [24]. Cells were sampled for scRNAseq when they were in mid-log growth. Cell counts at the time of sampling are reported in **Fig 1A** for each strain in each experiment (**Fig 1A**; pink points in bottom row). When multiple (identical) replicate flasks were sampled on a given day, average cell counts are reported.

For the glucose depletion experiment, a single experiment was performed with two biological replicates (two separate flasks) of diploid s288c cells and two biological replicates of haploid BY4741 cells. After the overnight shaking in YPD described above, cells were massively diluted into Synthetic Complete (Sunrise) plus 2% glucose media. Cells were sampled for scRNAseq at four points along the growth curve. Average cell counts at the time of sampling are reported in **Fig 1A** for each strain (**Fig 1A**; blue points in middle row).

For the sucrose depletion experiment used to study stress from sucrose limitation, the same flasks used to study stress from misfolded proteins were allowed to grow to saturation. Cells were sampled for scRNAseq during early stationary phase. Cell counts at the time of sampling are reported in **Fig 1A** for each strain in each experiment (**Fig 1A**; red points in bottom row). When multiple (identical) replicate flasks were sampled on a given day, average cell counts are reported. Since the cells in this experiment experience two stressors (YFP expression and sucrose limitation), we use each strain as a control for itself to extract the effect of sucrose limitation. For example, we compare the YFPwt-expressing strain in early log phase growth (**Fig 1C**; pink) versus early stationary phase growth (**Fig 1C**; red).

Single-cell RNA sequencing

For all the cell samples described above, 3mL of liquid culture were removed from the shaking flasks at each timepoint. From that, 100uL were used to count cells (as described above). The remaining 2.9mL of liquid culture was immediately fixed with 390uL of 35% formaldehyde to preserve transcriptome information for scRNAseq. After fixation, single-cell RNA sequencing was performed using a yeast-specific application of split-pool based *in situ* transcriptomics sequencing (SPLiT-seq) as previously described in [18].

Single-cell RNAseq data processing

Transcriptome alignment, barcode assignment, and gene expression count analyses were performed as described in [18]. Briefly, sequencing reads were aligned to the R64 s288c genome from NCBI supplemented with added transposon sequences from the Saccharomyces Genome Database (SGD) [46] and sorted by barcode using STARsolo [47]. The yeast genome has significant sequence homology to itself due to a recent genome duplication [48]. Therefore, we utilized the uniform multimapping algorithm within STARsolo to keep reads that mapped to multiple locations in the genome. This outputted gene-by-barcode matrices representing the detected gene counts for every cell. rRNA genes were identified and removed so they would not distort subsequent calculations and analyses. To remove barcodes that reflect empty cells and

cells with low gene detection, we applied the “knee” detection filter previously described in [18]. These quality-thresholded gene-by-barcode matrices were then converted to Seurat Objects using the Seurat R package (version 5.1.0) for further analyses [49].

Full analysis scripts are available on the Open Science Framework: https://osf.io/pcngh/files/osfstorage.

Creating induced and repressed stress response gene lists (i.e., MSR, GSR, and SSR)

To determine the induced and repressed stress response genes, we compared gene expression of an unstressed population of cells to a stressed population of cells. For the MSR we compared log-phase cells expressing YFPwt (**Fig 1A**; light purple) to those expressing YFPm4 (**Fig 1A**; dark purple). For the GSR we compared cells sampled early in glucose limitation (**Fig 1A**; lightest blue “T1”) to cells sampled late into glucose limitation (**Fig 1A**; darkest blue “T4”). For the SSR we compared YFPwt-expressing cells sampled early in sucrose limitation (**Fig 1A**; pink “T1”) to YFPwt-expressing cells sampled late in sucrose limitation (**Fig 1A**; red “T2”).

Raw RNA count matrices for each stressed and unstressed cell population were extracted. For each population, gene expression ratios were computed by dividing the summed counts of each gene by the total read counts detected in that population. These ratios were then scaled to transcripts per million (TPM) by multiplying by 10^6. For each of the three stressed–unstressed comparisons, fold change (FC) values were calculated as the difference between log2-transformed TPM values in stressed versus unstressed populations (per-gene FC shown in **Fig 4A** and provided in **Supplemental Table 2**).

Fold change tells us about strong average differences between the two populations, not whether they are significant. To determine the latter (**Fig 4B**), often one would use replicate experiments. Here, we have the advantage of having many replicates in the sense of having sampled many cells. To fairly compare cells from one sample (e.g., stressed sample) to cells from another (e.g., unstressed cells), we must normalize to control for differences in the total number of reads per cell that were obtained in each sample. To do this, log-normalization was applied to raw counts using the standard parameters for the NormalizeData function in Seurat, which takes each gene’s expression ratio per cell (gene count/total count), multiplies it by 10,000, adds one and then takes the natural log. A Wilcoxon rank-sum test with multiple-testing correction was used to determine, for each gene, whether normalized expression per cell in the stressed cell population significantly differed from that in the unstressed population. To correct for multiple hypothesis testing (having tested many genes) we applied either the Benjamini-Hochberg method (defining the MSR) or Bonferroni correction (defining the GSR and SSR) depending on the magnitude of the differences between the cells.

Genes were classified as upregulated in the stressed cells (i.e., part of the *induced* stress response) if log2FC was greater than 0 and the adjusted p-value across all single cells was less than 0.05. Genes were classified as upregulated in unstressed cells (i.e., part of the *repressed* stress response) if log2FC was less than 0 and the adjusted p-value across single cells was less than 0.05.

Calculating the percent of transcriptome devoted to different gene sets (e.g., iESR)

After transposing the Seurat Object and converting to a data frame, each row in the data frame is a single cell defined by its barcode and each column is a gene. We took row sums to

determine the total number of raw gene counts per cell, giving us all counts for a cell. For each gene set, we subsetted the overall dataframe to only include the desired genes for one gene set at a time. Row sums for those subsets were taken. To determine the percent of the transcriptome devoted to each gene set, we then divided our subsetted counts by all counts for each individual cell. These data, for each cell, are plotted in **Fig 1B–1C** for the iESR, and in **Supplemental Figures S1–S6** for other gene sets. One note is that a cell with very low read counts may falsely appear to have a very high percent of the transcriptome devoted to a certain gene set due to sparse sampling. To avoid this type of bias, when calculating these single-cell percentages, the cells expressing the lowest 10% of total raw counts were removed from each dataset. This resulted in 12,488 cells to be used when analyzing the percent of the transcriptome throughout.

Ranking and classifying cells by their iESR (Fig 1C and 1D) or other gene sets

After calculating, for each cell, the percent of its transcriptome devoted to various gene sets (i.e., stress responses), we can rank order cells by those percentages. We did this for different pairs of stressed versus unstressed cell populations (**Figs 1C, S2, S3, S4,** and **S5**; horizontal axis). We noticed that when doing so, cells did not separate cleanly based on whether they were from a stressed versus unstressed population. In other words, some stressed cells have comparable or even lower levels of an induced stress response than other cells. This suggests we cannot always predict whether a cell has experienced stress from its degree of stress response induction. To test this we attempted to classify cells as coming from the stressed or unstressed population based on the percent of their transcriptome devoted to various gene sets. Specifically, we ranked cells by their percent of transcriptome devoted to the induced stress response (e.g., iESR or iMSR) in ascending order and placed a vertical cut at the rank representing the true number of unstressed cells (**Figs 1C, S2A, S2B, S3A,** and **S3B**). Cells left of the cut were predicted “unstressed,” and cells right of the cut were predicted “stressed.” We then reported the fraction correctly classified within each group (**Figs 1D, S2C,** and **S3C**). The same was done for repressed stress responses, where cells left of the cut were predicted “stressed,” and cells right of the cut were predicted “unstressed” (**Figs S4** and **S5**).

Building Random Forest models and predicting cell provenance (Fig 2B and 2C)

Classifier-type Random Forest models [31] were built on raw transcriptomics data (thresholded for low read counts but not normalized) in R using the package randomForest [50]. Specifically, for every stress type dataset, cells were sorted into a stress (YFPm4, gT4, or sT2) or an unstress (YFPwt, gT1, or sT1) category and given a label classification variable of “stress” or “unstress” to allow for model testing across datasets. Full models were then built for each dataset by comparing the classifier variable to the per cell gene expression for every annotated gene using largely the default parameters in the randomForest function. The exception being that importance values were called to be calculated to generate a mean decrease in accuracy score (MDA) for each gene (**Fig 4C**). Because our focus was on comparing model predictions across datasets rather than within a single dataset, we used the reported out-of-bag error as an estimate of within model accuracy, and we built each model with *all* cells from the given dataset instead of using training and testing data subsets. To calculate across model accuracy, we used the predict function from the core R stats package to generate predictions from a given model using the other datasets as tests (e.g., misfolded protein and sucrose datasets fed to the glucose model). We then used the confusionMatrix function from the R package caret [51], which calculates a cross-tabulation of observed and predicted categories to generate the accuracy statistics. These, along with the out-of-bag error rates, are the basis of the data

displayed on **Fig 2B**. For **Fig 2C**, the same process was followed, except multiple models were built for each dataset with the specified induced or repressed stress gene sets left out of the gene expression data.

Clustering cells based on their transcriptomes

For the clustering analysis in **Fig 3A**, the raw count data for cells from timepoint 4 (T4) in glucose-limited media were log-normalized using Seurat's NormalizeData function. The top 3,000 most variable genes were identified with the FindVariableFeatures function to focus the analysis on the most informative genes. These genes were then centered and scaled using the ScaleData function. Dimensionality reduction was performed using principal component analysis (RunPCA function) on the scaled, normalized expression matrix. The value of each principal component dimension was then assessed with the JackStraw and ScoreJackStraw functions to identify the principal components contributing meaningfully to the variation of the data. Nearest-neighbor relationships were calculated on the normalized and scaled data using the FindNeighbors function (dims = 1:10), and cells were grouped into discrete clusters using FindClusters (resolution = 0.1). Clusters were visualized in two-dimensional space using the RunUMAP function (dims = 1:10), allowing transcriptionally similar cells to be represented as nearby points in the embedding. To visualize the fraction of each cell's transcriptome corresponding to either the iESR, rESR, or TEs, we first calculated the percentage of normalized counts devoted to the gene sets given in **Supplemental Table 1**. Cells were then colored in the UMAP according to these percentages using the FeaturePlot function.

Plotting a population's percent transcriptome devoted to iESR versus TEs

To assess the expression of iESR and TE genes on average for each defined population, we first calculated the percent of the transcriptome devoted to these gene sets (**Supplemental Table 1**) for single cells as described above. Cells lacking detected counts for either iESR or TE genes were removed. For the remaining cells, we calculated the average expression and standard error of the mean (SEM) (represented by error bars in **Fig 3B**) for iESR and TE percentages, grouping by experiment, cell type, and sampletime. For each stressor, linear regression trend lines were fit, and the Pearson correlation coefficient (r) was reported.

Comparing expression of some genes to others across single cells

We asked whether transposable elements (TEs) behave like part of the stress-response program or whether their expression varies independently. To test this, we calculated, for each cell, the fraction of the transcriptome devoted to all TEs (**Supplemental Table 1**) and compared it to the fraction devoted to stress-response genes that are not TEs. Pearson correlations were then computed across single cells within each of the 20 samples (**Fig 3D**; teal boxplots). Because the iESR gene set does not contain any TEs, it was included in this analysis without modification. As shown in **Fig 3D**, these comparisons revealed that TE expression tends to be negatively correlated with stress-response genes.

In **Fig S8**, we refined this comparison by focusing only on the subset of TE genes that are part of each induced stress-response set (iMSR, iGSR, and iSSR). For each of these responses, we split the genes into two groups: the TE genes in the set and the non-TE genes in the set. We then asked whether, across single cells, expression of the TE subset rose and fell together with the non-TE subset. Consistent with the broader analysis, these TE subsets also tended to vary inversely with the rest of their stress-response program.

As a control, we repeated the analysis using random subsets. Each induced stress-response gene set was randomly divided into two halves, and correlations between those halves were

computed across cells (gray boxplots in **Fig 3D**). Unlike the TE analyses, these random splits showed positive correlations, confirming that the negative correlations we observed with TEs are not artifacts of the method.

Random sampling analysis to determine if iESR gene behavior substantially deviates from the norm

When identifying genes with differential expression in stressed versus unstressed cells, the choice of method often matters. For example, in **Fig 4** we show that genes comprising the iESR stand out as differentially upregulated in stressed cells in terms of their average fold change (**Fig 4A**) but not in terms of other metrics (**Fig 4B** and **4C**). To test whether the behavior of genes that comprise the iESR stand out as different from other genes, we performed a random sampling analysis. In each of the 10,000 iterations, we randomly sampled 270 genes (the size of the iESR gene set that maps to the annotated s288c genome we used for alignment) without replacement within the sample from all genes (including the iESR) in our cell-by-gene matrices. For each resampled set, we asked whether the genes comprising that set differ from those comprising the iESR in terms of the strength of their association with stressed cell populations. To address whether the iESR taken as a whole is a better barometer of stress than random samples of the same size, we tested whether the observed median score of the iESR exceeded that of random samples.

For **Figs 4A** and **4B**, we compared the randomly sampled sets to the iESR genes as follows. For each randomly sampled set of genes, we calculated the median fold-change (FC) and median +/-log10(corrected p-value) (i.e., we plotted downregulated genes in stressed conditions as negative values and the upregulated as positive) in each stressed versus unstressed comparison. This analysis showed that, across 10,000 iterations, the observed median FC of the iESR exceeded that of all randomly sampled gene sets of equal size (i.e., the observed iESR median was always higher), demonstrating that iESR genes are more strongly upregulated than expected by chance. By contrast, the median of all randomly sampled +/-log10(corrected p-values) for each stressor was 0, which is the same as the observed median +/-log10(corrected p-value) of the iESR (**Fig 4B**). This indicates that the single-cell behavior of the iESR did not differ from that of random gene sets of equal size.

Similarly, we ran 10,000 iterations of randomly sampling 270 genes without replacement for **Fig 4C** to observe whether the iESR stands out as different from other genes. The median of the mean decrease in accuracy (MDA) when each of those genes were removed from the Random Forest model were recorded for each stressor in each iteration. For comparisons involving sucrose stress and misfolding stress, the randomly chosen genes either had a median MDA value of 0, which is identical to the median of the iESR genes, or were higher than that of the iESR. We conclude that the iESR genes are not different from other gene sets in terms of their predictive power for these stressors, and, sometimes (~2% for misfolding stress and ~9% for sucrose stress), random gene sets had a higher median MDA for predicting stressed cells than the iESR. However, for glucose stress, the observed median MDA value for iESR genes was higher than 99% of random iterations. This demonstrates that, just in glucose stress, iESR genes are more important for predicting stressed cells than randomly chosen gene sets.

Supplemental Table Legends

Supplemental Table 1: A table with all gene sets used and created. iESR and rESR genes are from [1–5]. Chaperones are from [44]. Transposable elements (labeled as transposable element

gene, LTR retrotransposon, or rt pseudogene) are from SGD annotations [52]. Controls for **Fig 3D** (lists of induced stress response genes randomly split in half) are included on page 2.

Supplemental Table 2: A table with the average expression of each gene in unstressed and stressed cells across stressors. Fold change (FC), corrected p-values, and signed -log10(p-values) are listed for each gene used in **Fig 4A** and **4B**. Top FC genes from **Fig 4A** are labeled.

Data Availability Statement

Data processing and analysis scripts, as well as processed alignment data can be found on the Open Science Framework: https://osf.io/pcngh/files/osfstorage. Both raw fastq files and alignment data for the glucose stress experiment and the first misfolding/sucrose stress experiment can be accessed through NCBI GEO GSE251966 (data used in this study pertains to samples GSM7990658 and GSM7990659). Both raw fastq files and alignment data for the second misfolding/sucrose stress experiment can be accessed through NCBI GEO GSE309715. We also provide raw sequencing reads as fastq files, along with raw, unfiltered barcode (cell)-by-gene count matrices that were created after first aligning all sequenced reads to the appropriate genomes using STARsolo and assigning each read according to its corresponding barcode. All subsequent analyses performed in this publication can be reproduced directly with these data.

Acknowledgments

Funding for this work was provided by a National Institutes of Health grant R35GM133674 (to KGS), an Alfred P Sloan Research Fellowship in Computational and Molecular Evolutionary Biology grant FG-2021-15705 (to KGS), and a National Science Foundation Biological Integration Institution grant 2119963 (to KGS). We would also like to thank the members of the ASU Biodesign Center for Mechanisms of Evolution and the Geiler-Samerotte Lab for their support and feedback.

Competing Interests

The authors have declared that no competing interests exist.

Supplemental Figures

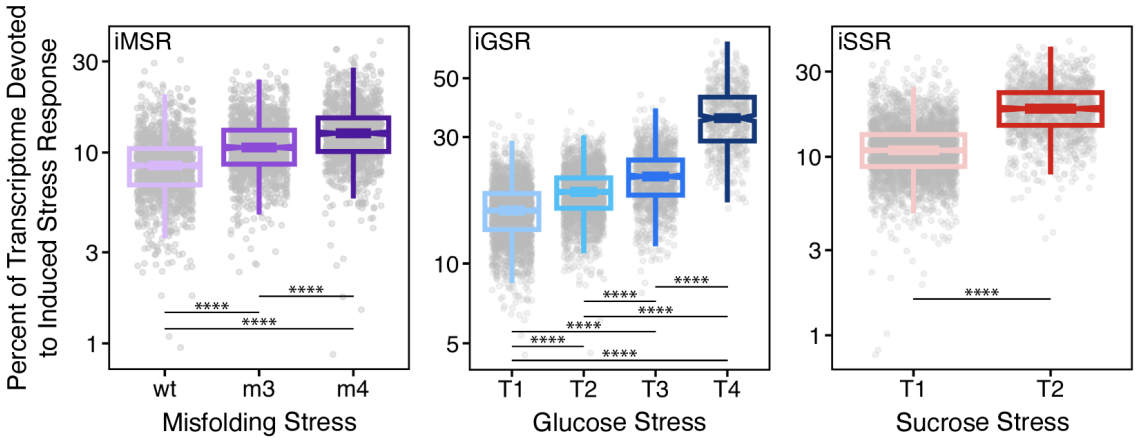

Fig S1: Percent of transcriptome devoted to induced stress responses increases as stressor increases. The percent of the transcriptome devoted to corresponding stress responses (iMSR, iGSR, or iSSR). Every dot represents a single cell; median and quartiles are shown across cells pertaining to a given strain (misfolding stress) or timepoint (nutrient stress). Cells with no detected induced stress response were excluded. **** indicates $p < 0.0001$ with Wilcoxon signed-rank test.

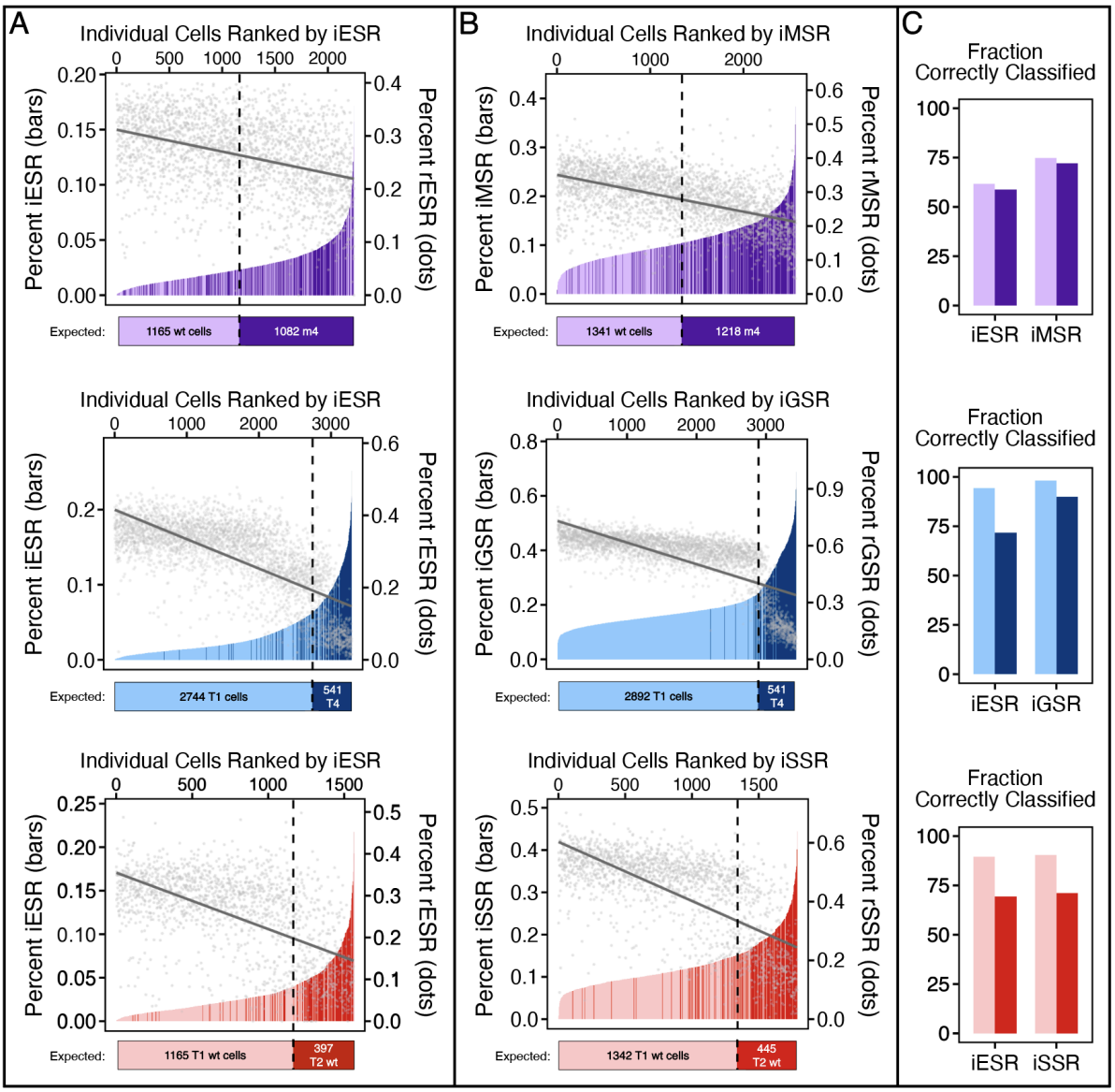

Fig S2: Single-cell heterogeneity across induced stress responses. Rows: (1) YFPwt and YFPm4, (2) T1 and T4 (glucose), and (3) T1 and T2 (sucrose). **(A)** Cells are ordered by their percent of the transcriptome dedicated to the iESR (bars). Any cells with no counts detected to the iESR were removed for plotting purposes. An inverse correlation is observed between the (induced) iESR and (repressed) rESR (points). The horizontal dashed line signifies the number of unstressed cells. Horizontal bars indicate the true number of unstressed and stressed cells in each plot. Same as shown in **Fig 1C**. **(B)** Cells are ordered by percent of their transcriptome dedicated to corresponding induced stress responses (bars). Any cells with no counts detected to the induced stress response for the corresponding graph were removed for plotting purposes. An inverse correlation is observed between the induced stress response (bars) and repressed stress response (points). **(C)** Bars display the number of correctly classified unstressed or stressed cells based on the rank order of their iESR or corresponding induced stress response. Same as shown in **Fig 1D**.

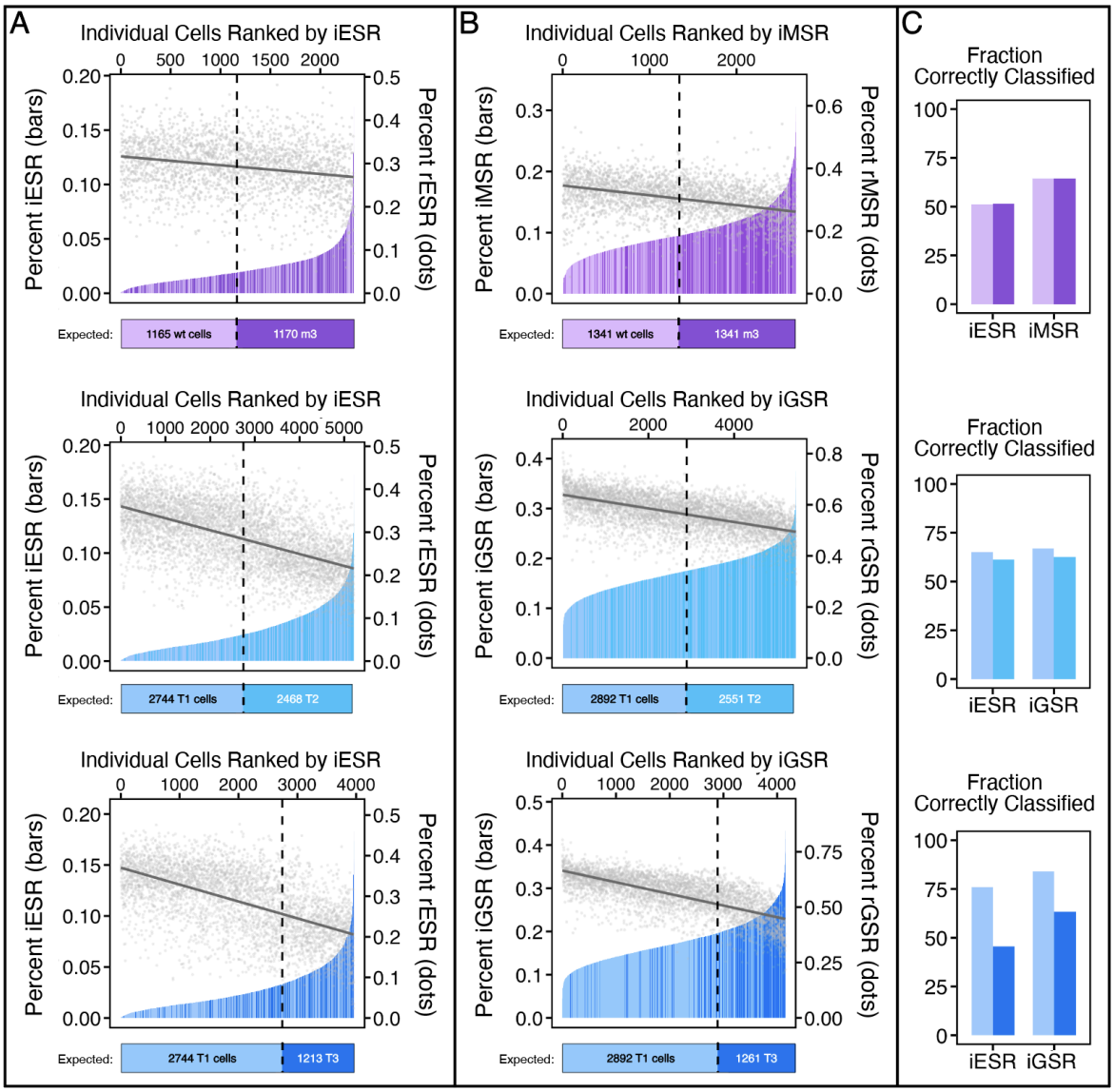

Fig S3: Single-cell heterogeneity across induced stress responses in additional stress comparisons. Rows: (1) YFPwt and YFPm3, (2) T1 and T2 (glucose), and (3) T1 and T3 (glucose). **(A)** Cells are ordered by the percent of their transcriptome dedicated to the iESR (bars). Any cells with no counts detected to the iESR were removed for plotting purposes. An inverse correlation is observed between the (induced) iESR and (repressed) rESR (points). The horizontal dashed line signifies the number of unstressed cells. Horizontal bars indicate the true number of unstressed and stressed cells in each plot. **(B)** Cells are ordered by the percent of their transcriptome dedicated to corresponding induced stress responses (bars). Any cells with no counts detected to the induced stress response for the corresponding graph were removed for plotting purposes. An inverse correlation is observed between the induced stress response (bars) and repressed stress response (points). **(C)** Bars display the number of correctly classified unstressed or stressed cells based on the rank order of their iESR or corresponding induced stress response.

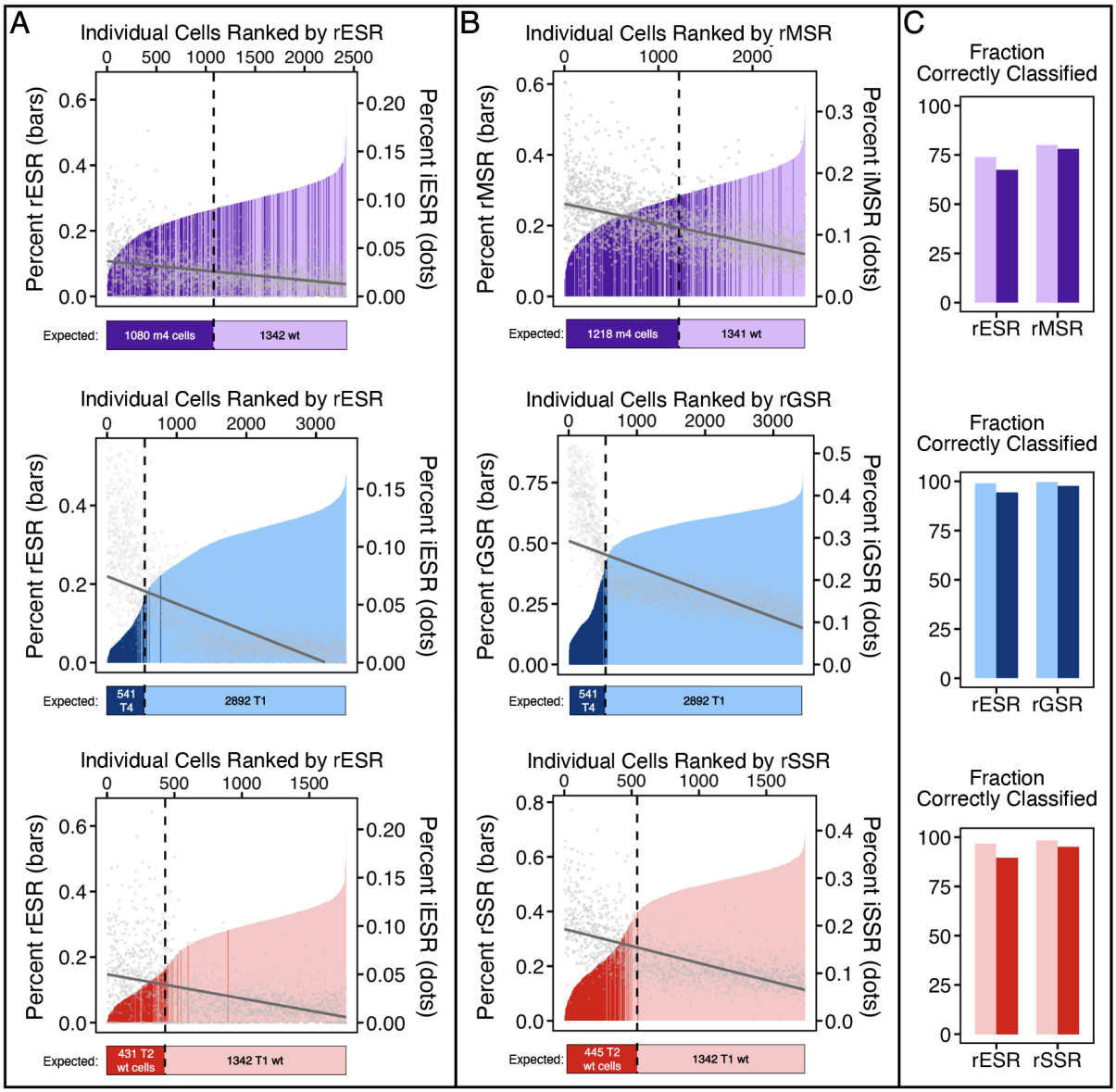

Fig S4: Single-cell heterogeneity across repressed stress responses. Rows: (1) YFPwt and YFPm3, (2) T1 and T2 (glucose), and (3) T1 and T3 (glucose). **(A)** Cells are ordered by the percent of their transcriptome dedicated to the rESR (bars). Any cells with no counts detected to the rESR were removed for plotting purposes. An inverse correlation is observed between the (repressed) rESR and (induced) iESR (points). The horizontal dashed line signifies the number of stressed cells. Horizontal bars indicate the true number of stressed and unstressed cells in each plot. **(B)** Cells are ordered by the percent of their transcriptome dedicated to corresponding repressed stress responses (bars). Any cells with no counts detected to the repressed stress response for the corresponding graph were removed for plotting purposes. An inverse correlation is observed between the repressed stress response (bars) and induced stress response (points). **(C)** Bars display the number of correctly classified unstressed or stressed cells based on the rank order of their rESR or corresponding repressed stress response.

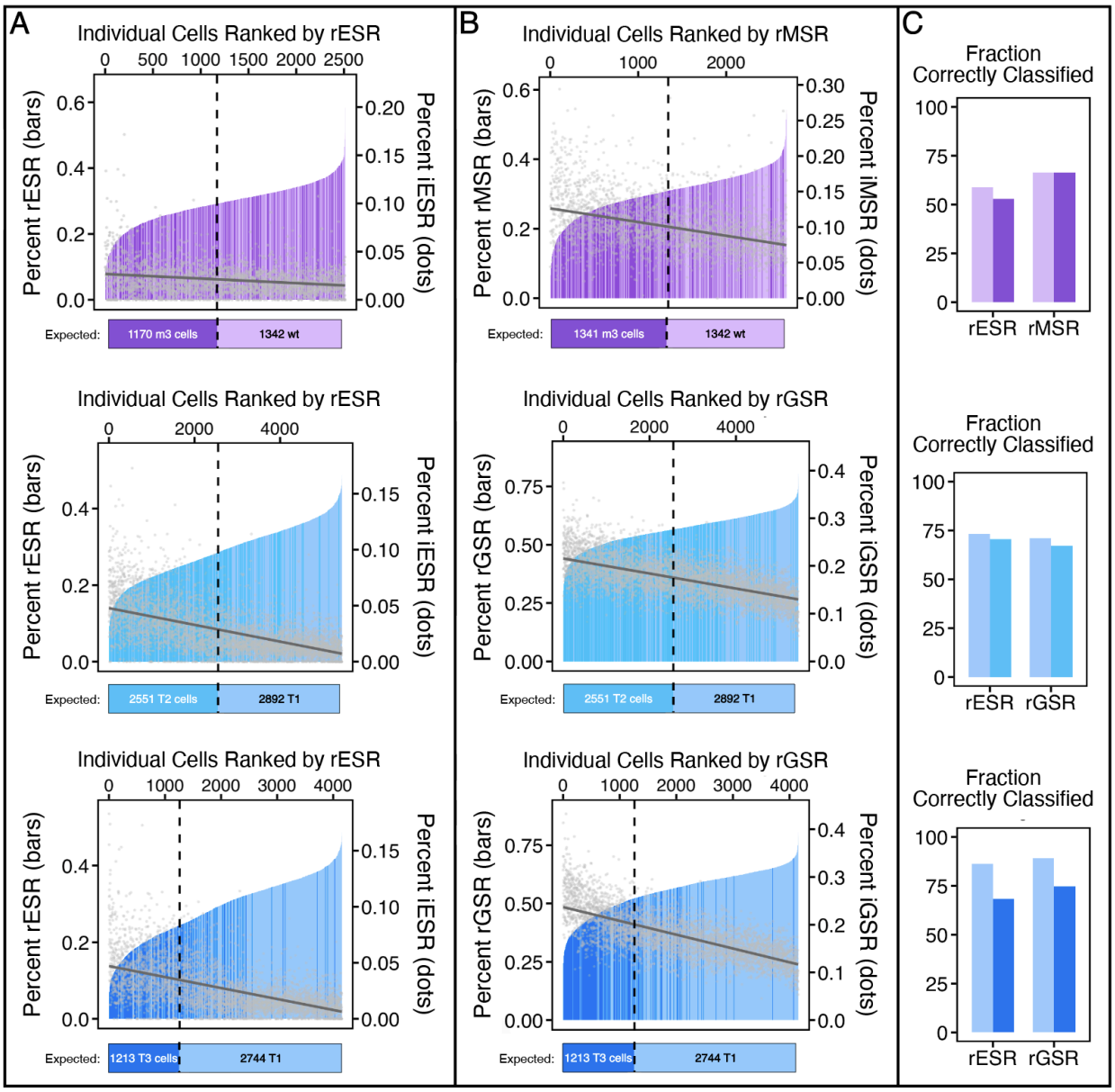

Fig S5: Single-cell heterogeneity across repressed stress responses in additional stress comparisons. Rows: (1) YFPwt and YFPm3, (2) T1 and T2 (glucose), and (3) T1 and T3 (glucose). **(A)** Cells are ordered by the percent of their transcriptome dedicated to the rESR (bars). Any cells with no counts detected to the rESR were removed for plotting purposes. An inverse correlation is observed between the (repressed) rESR and (induced) iESR (points). The horizontal dashed line signifies the number of stressed cells. Horizontal bars indicate the true number of stressed and unstressed cells in each plot. **(B)** Cells are ordered by the percent of their transcriptome dedicated to corresponding repressed stress responses (bars). Any cells with no counts detected to the repressed stress response for the corresponding graph were removed for plotting purposes. An inverse correlation is observed between the repressed stress response (bars) and induced stress response (points). **(C)** Bars display the number of correctly classified unstressed or stressed cells based on the rank order of their rESR or corresponding repressed stress response.

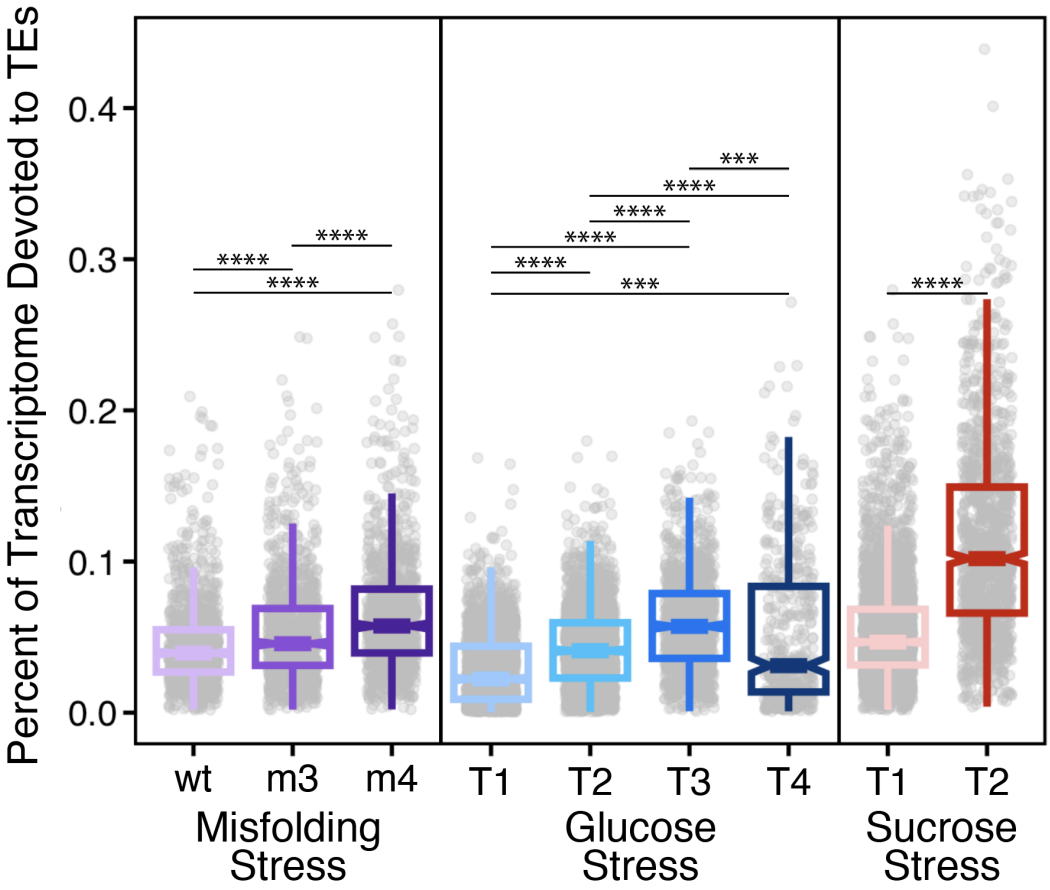

Fig S6: TEs generally increase as stressors increase. The percent of the transcriptome devoted to TEs for each individual cell is represented by a dot. Violin plots show the medians and distribution of cells within each strain (misfolding stress) or sampling time (nutrient stress). Cells with no detected TEs were excluded. *** indicates $p < 0.001$ and **** indicates $p < 0.0001$ with Wilcoxon signed-rank test.

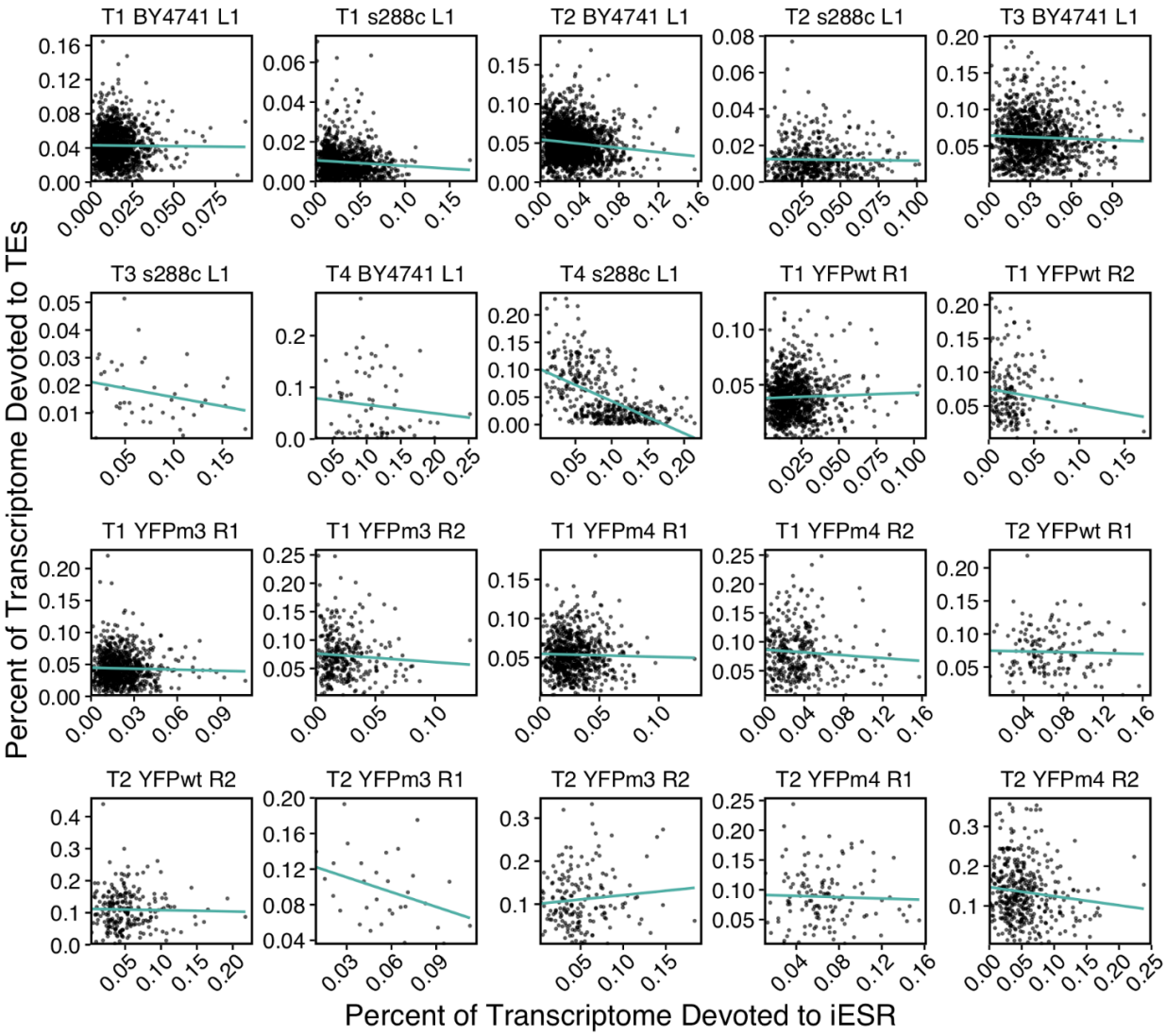

Fig S7: TEs and the iESR anticorrelate on the single-cell level. The percent of the transcriptome devoted to the iESR (x-axis) and TEs (y-axis) are plotted for each individual cell (a point on the graph) for all 20 datasets. As noted in each plot's title, the plots are split by sampling timepoint (T), strain, and experiment ID number. If two replicate flasks were studied on the same day, those cells are included in the same plot. Trend lines indicate an anticorrelation or near-zero correlation for all plots. Cells with no detected iESR or TEs were excluded. **Fig 3C** shows "T4 s288c L1".

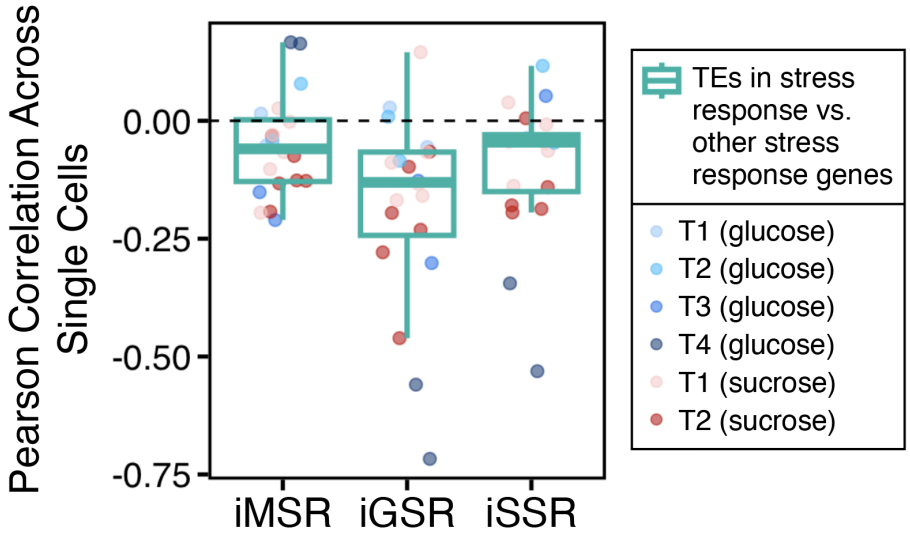

Fig S8: TEs in stress responses anticorrrelate with other stress response genes. For each induced stress response, the TEs within that stress response were correlated with all other genes within that stress response for each individual cell. The Pearson correlations for each condition are plotted. Medians are negative, indicating an anticorrelated trend.

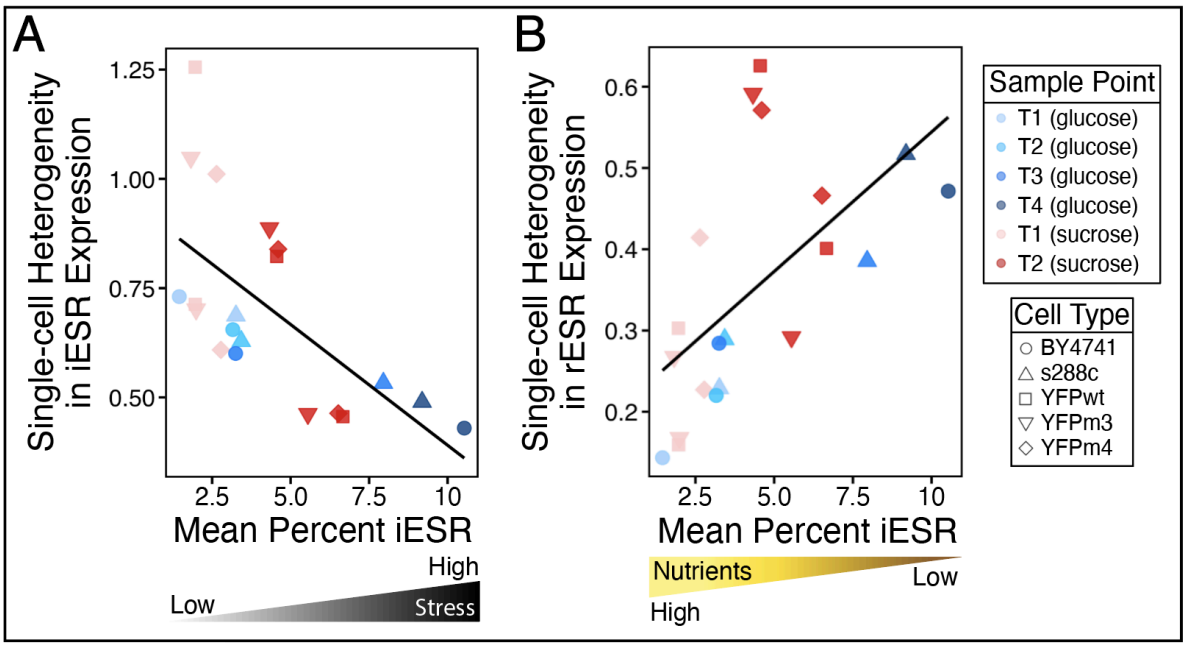

Fig S9: Single-cell variation in transcriptional responses trends with magnitude of stress. (A) iESR expression variation, as measured by the population coefficient of variation (CV), decreases as the population mean iESR expression increases. **(B)** rESR expression variation (CV) increases with population mean iESR expression.

References

1. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, et al. Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell. 2000;11: 4241–4257.

2. Brauer MJ, Huttenhower C, Airoldi EM, Rosenstein R, Matese JC, Gresham D, et al. Coordination of Growth Rate, Cell Cycle, Stress Response, and Metabolic Activity in Yeast. Molecular Biology of the Cell. 2008. pp. 352–367. doi:10.1091/mbc.e07-08-0779

3. Ho Y-H, Gasch AP. Exploiting the yeast stress-activated signaling network to inform on stress biology and disease signaling. Curr Genet. 2015;61: 503–511.

4. Costa-Mattioli M, Walter P. The integrated stress response: From mechanism to disease. Science. 2020;368: eaat5314.

5. Chen X, Shi C, He M, Xiong S, Xia X. Endoplasmic reticulum stress: molecular mechanism and therapeutic targets. Signal Transduct Target Ther. 2023;8: 352.

6. Bergen AC, Kocik RA, Hose J, McClean MN, Gasch AP. Modeling single-cell phenotypes links yeast stress acclimation to transcriptional repression and pre-stress cellular states. Elife. 2022;11. doi:10.7554/eLife.82017

7. Cai L, Dalal CK, Elowitz MB. Frequency-modulated nuclear localization bursts coordinate gene regulation. Nature. 2008;455: 485–490.

8. Li S, Giardina DM, Siegal ML. Control of nongenetic heterogeneity in growth rate and stress tolerance of Saccharomyces cerevisiae by cyclic AMP-regulated transcription factors. PLoS Genet. 2018;14: e1007744.

9. Kocik RA, Gasch AP. Regulated resource reallocation is transcriptionally hard wired into the yeast stress response. bioRxivorg. 2024. doi:10.1101/2024.12.03.626567

10. Sampaio NMV, Blassick CM, Andreani V, Lugagne J-B, Dunlop MJ. Dynamic gene expression and growth underlie cell-to-cell heterogeneity in Escherichia coli stress response. Proc Natl Acad Sci U S A. 2022;119: e2115032119.

11. Munsky B, Li G, Fox ZR, Shepherd DP, Neuert G. Distribution shapes govern the discovery of predictive models for gene regulation. Proc Natl Acad Sci U S A. 2018;115: 7533–7538.

12. Su Y, Xu C, Shea J, DeStephanis D, Su Z. Transcriptomic changes in single yeast cells under various stress conditions. BMC Genomics. 2023;24: 88.

13. Nadal-Ribelles M, Lieb G, Solé C, Matas Y, Szachnowski U, Andjus S, et al. Transcriptional heterogeneity shapes stress-adaptive responses in yeast. Nature Communications. 2025;16: 1–17.

14. Nadal-Ribelles M, Islam S, Wei W, Latorre P, Nguyen M, de Nadal E, et al. Sensitive high-throughput single-cell RNA-seq reveals within-clonal transcript correlations in yeast

populations. Nat Microbiol. 2019;4: 683–692.

15. He Z, Liu Z, Wang Q, Sima X, Zhao W, He C, et al. Single-cell and spatial transcriptome assays reveal heterogeneity in gliomas through stress responses and pathway alterations. Front Immunol. 2024;15: 1452172.

16. Choudhary D, Lagage V, Foster KR, Uphoff S. Phenotypic heterogeneity in the bacterial oxidative stress response is driven by cell-cell interactions. Cell reports. 2023;42. doi:10.1016/j.celrep.2023.112168

17. Wang B, Lin AE, Yuan J, Novak KE, Koch MD, Wingreen NS, et al. Single-cell massively-parallel multiplexed microbial sequencing (M3-seq) identifies rare bacterial populations and profiles phage infection. Nature Microbiology. 2023;8: 1846–1862.

18. Brettner L, Eder R, Schmidlin K, Geiler-Samerotte K. An ultra high-throughput, massively multiplexable, single-cell RNA-seq platform in yeasts. Yeast. 2024. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/yea.3927

19. Horváth V, Merenciano M, González J. Revisiting the Relationship between Transposable Elements and the Eukaryotic Stress Response. Trends Genet. 2017;33: 832–841.

20. Bradshaw VA, McEntee K. DNA damage activates transcription and transposition of yeast Ty retrotransposons. Mol Gen Genet. 1989;218: 465–474.

21. Paquin CE, Williamson VM. Temperature effects on the rate of ty transposition. Science. 1984;226: 53–55.

22. Kuchina A, Brettner LM, Paleologu L, Roco CM, Rosenberg AB, Carignano A, et al. Microbial single-cell RNA sequencing by split-pool barcoding. Science. 2021;371. doi:10.1126/science.aba5257

23. Zaneveld JR, McMinds R, Vega Thurber R. Stress and stability: applying the Anna Karenina principle to animal microbiomes. Nat Microbiol. 2017;2: 17121.

24. Geiler-Samerotte KA, Dion MF, Budnik BA, Wang SM, Hartl DL, Drummond DA. Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. Proceedings of the National Academy of Sciences. 2011;108: 680–685.

25. Trotter EW, Kao CM-F, Berenfeld L, Botstein D, Petsko GA, Gray JV. Misfolded proteins are competent to mediate a subset of the responses to heat shock in Saccharomyces cerevisiae. J Biol Chem. 2002;277: 44817–44825.

26. Travers KJ, Patil CK, Wodicka L, Lockhart DJ, Weissman JS, Walter P. Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. Cell. 2000;101: 249–258.

27. Narayana Rao KB, Pandey P, Sarkar R, Ghosh A, Mansuri S, Ali M, et al. Stress Responses Elicited by Misfolded Proteins Targeted to Mitochondria. J Mol Biol. 2022;434: 167618.

28. Geiler-Samerotte KA, Hashimoto T, Dion MF, Budnik BA, Airoldi EM, Drummond DA. Quantifying condition-dependent intracellular protein levels enables high-precision fitness estimates. PLoS One. 2013;8: e75320.

29. Gasch AP, Yu FB, Hose J, Escalante LE, Place M, Bacher R, et al. Single-cell RNA sequencing reveals intrinsic and extrinsic regulatory heterogeneity in yeast responding to stress. PLoS Biol. 2017;15: e2004050.

30. Causton HC, Ren B, Koh SS, Harbison CT, Kanin E, Jennings EG, et al. Remodeling of yeast genome expression in response to environmental changes. Mol Biol Cell. 2001;12: 323–337.

31. Breiman L. Random Forests. Machine Learning. 2001;45: 5–32.

32. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. arXiv [physics.soc-ph]. 2008. doi:10.48550/ARXIV.0803.0476

33. Stanley D, Fraser S, Stanley GA, Chambers PJ. Retrotransposon expression in ethanol-stressed Saccharomyces cerevisiae. Appl Microbiol Biotechnol. 2010;87: 1447–1454.

34. Servant G, Pinson B, Tchalikian-Cosson A, Coulpier F, Lemoine S, Pennetier C, et al. Tye7 regulates yeast Ty1 retrotransposon sense and antisense transcription in response to adenylic nucleotides stress. Nucleic Acids Res. 2012;40: 5271–5282.

35. Rolfe M, Spanos A, Banks G. Induction of yeast Ty element transcription by ultraviolet light. Nature. 1986;319: 339–340.

36. Lindquist S, Craig EA. THE HEAT-SHOCK PROTEINS. Annu Rev Genet. 1988;22: 631–677.

37. Laskar S, Bhattacharyya MK, Shankar R, Bhattacharyya S. HSP90 controls SIR2 mediated gene silencing. PLoS One. 2011;6: e23406.

38. Khurana N, Bhattacharyya S. Hsp90, the concertmaster: tuning transcription. Front Oncol. 2015;5: 100.

39. Huang L, Sun Q, Qin F, Li C, Zhao Y, Zhou D-X. Down-regulation of a SILENT INFORMATION REGULATOR2-related histone deacetylase gene, OsSRT1, induces DNA fragmentation and cell death in rice. Plant Physiol. 2007;144: 1508–1519.

40. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nat Rev Genet. 2009;10: 94–108.

41. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell. 2007;128: 1089–1103.

42. Wood JG, Helfand SL. Chromatin structure and transposable elements in organismal aging. Front Genet. 2013;4: 274.

43. Feng Y, Cao S-Q, Shi Y, Sun A, Flanagan ME, Leverenz JB, et al. Human herpesvirus-associated transposable element activation in human aging brains with Alzheimer's disease. Alzheimers Dement. 2025;21: e14595.

44. Gong Y, Kakihara Y, Krogan N, Greenblatt J, Emili A, Zhang Z, et al. An atlas of chaperone–protein interactions in Saccharomyces cerevisiae: implications to protein folding pathways in the cell. Molecular Systems Biology. 2009 [cited 23 Sep 2025]. doi:10.1038/msb.2009.26

45. Saint M, Bertaux F, Tang W, Sun X-M, Game L, Köferle A, et al. Single-cell imaging and RNA sequencing reveal patterns of gene expression heterogeneity during fission yeast growth and adaptation. Nat Microbiol. 2019;4: 480–491.

46. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, et al. Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. 2012;40: D700–5.

47. Kaminow B, Yunusov D, Dobin A. STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. bioRxiv. bioRxiv; 2021. doi:10.1101/2021.05.05.442755

48. Wolfe KH. Origin of the Yeast Whole-Genome Duplication. PLoS Biol. 2015;13: e1002221.

49. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2024;42: 293–304.

50. Breiman L, Cutler A, Liaw A, Wiener M. RandomForest: Breiman and cutlers random forests for classification and regression. CRAN: Contributed Packages. The R Foundation; 2002. doi:10.32614/cran.package.randomforest

51. Kuhn M. caret: Classification and Regression Training. CRAN: Contributed Packages. The R Foundation; 2007. doi:10.32614/cran.package.caret

52. Website. Available: http://sgd-archive.yeastgenome.org/sequence/S288C_reference/other_features/