I. Strain Preparation for Cytometry

A. Single Colonies

Each of the 4,159 strains from the GFP library prepared by Huh *et al.*[1] was grown overnight at 30°C in 400 μl SD-His. To simplify strain manipulations, strains were grown in 96-well × 1 ml blocks (Cat. Number 267007, Beckman Instruments, Fullerton, CA, USA). The following day, strains were diluted to approximately 1 cell/μl in sterile water. The entire library was plated for single colonies by transferring 50 μl of diluted cells into unique wells of 48-well SD-His plates (Cat. Number X6027, Genetix, Boston, MA, USA). After three days of growth at 30°C, single colonies were picked and used to inoculate unique wells of 96-well × 1ml blocks containing 400 μl of YEPD/well to create a master stock (subsequently stored at 4°C).

B. Strain Growth

One day prior to an experiment, 20 μl of cells from the master stock were used to inoculate 400 μl/well of medium in 96-well × 1ml blocks. The cells were then grown overnight to saturation without shaking at 30°C.

The following morning, cells were inoculated into 1.4 ml/well of fresh, pre-warmed (30°C) media in 96-well × 2ml blocks (Cat. Number 780271, Greiner America, Lake Mary, FL, USA) to give an optical density at 600 nm (OD_{600}) of 0.1 (YEPD) or 0.15 (SD+Leu+Met+Ura). This was done using a liquid-handling robot (BioMek FX, Beckman-Coulter, Fullerton, CA, USA). Each well of the 2ml block contained a single sterilized glass bead (Z143928, Sigma, St. Louis, MI, USA) to facilitate mixing.

Inoculated cultures were grown to mid-log phase OD_{600} ~0.6 in a Hi-Gro machine [Genomic Solutions (formerly GeneMachines), Ann Arbor, MI, USA] with shaking at 400 RPM, at 30°C and with periodic aeration. Typically, mid-log was reached after 6 hours of growth in YEPD and 6.75 hours in SD+Leu+Met+Ura.

Strains were pre-grown to saturation in the same medium used to grow them to mid-log phase.

C. Strain Harvesting

800 μl from each well of the 2 ml blocks were transferred into empty wells of a 96-well × 1 ml block, and then the cells were spun down at 2400 RPM in a tabletop centrifuge for 3 minutes. The supernatant was removed and the cells were suspended in 100 μl of TE. Next, 80 μl of cell suspension was transferred to a 384-well plate [Cat. Number 07-200-645 (Costar 3702), Fisher Scientific Co., Pittsburgh, PA, USA] for cytometry measurements. The OD_{600} of harvested cells was also determined by transferring 200 μl from each well of the 96-well × 2ml block in which the cells were grown to unique wells of a flat-bottom 96-well polystyrene plate [Cat. Number 12-565-365 (Nunc 260895), Fisher Scientific Co., Pittsburgh, PA, USA]. OD_{600} measurements were made with a SpectroMax340 spectrophotometer (Molecular Devices, Sunnyvale, CA, USA).

To minimize variation due to delays between harvesting and measurements, cells were grown in a temporally staggered manner such that 192 different strains (two 96-well blocks) were harvested and measured by cytometry every 35 minutes. A BioMek FX robot (Beckman-Coulter, Fullerton, CA, USA) was used to facilitate liquid handling.

D. Media

YEPD (per L)

20 g bactopeptone

10 g yeast extract

20 g dextrose

SD-His (per L)

6.7g YNB without amino acids

20 g dextrose

0.77 g CSM -Histidine

20g bacto-agar

SD-Lys-Met (per L)

6.7g YNB without amino acids

20 g dextrose

0.72 g CSM -Lysine-Methionine

20g bacto-agar

SD+Leu+Met+Ura (per L)

6.7 g YNB without amino acids

20 g dextrose

0.1 g leucine

0.02 g methionine

0.02 g uracil

TE (per L)

10 mM TRIS (pH 7.5)

1 mM EDTA

II. Cytometry Measurement Protocol

A. Sample Delivery

Custom-written software (HTS-Pro, see below) was used to control both an autosampler (HTS, Becton Dickinson, San Jose, CA, USA) and the acquisition of data. Each sample was mixed by the autosampler and then 25 μl of the cell suspension was injected into the flow cytometer. After an initial boost phase, the flow rate was maintained at 0.416 μl/second. Typical carryover from one well to another was measured to be at (or below) 0.1%. Sample delivery occurred for approximately 7 seconds. When combined with the time needed to move between wells *etc.*, a 96-well plate was routinely measured in about 14 minutes.

B. HTS-Pro

A manual describing HTS-Pro features is included as part of the Supplementary Material. Briefly, HTS-Pro allows a user to dictate how samples are delivered to a cytometer via a High-Throughput Sampler (Becton Dickinson, San Jose, CA, USA). Additionally, HTS-Pro manages data acquisition on an LSR-II cytometer (Becton Dickinson, San Jose, CA, USA) by controlling the instrument's software (FACSDiva, Becton Dickinson, San Jose, CA, USA). Data described in this paper were obtained by running HTS-Pro in the 'Timed Run' mode with the following settings:

Pump time:	7 seconds
Flow rate:	0.4167 μl/second
Delay time:	1800 milliseconds
Pause length:	0 seconds
Pause after:	96 wells

Not-for-profit organizations can obtain this software directly from the authors. For-profit companies should contact the UCSF Technology Licensing Office.

C. Cytometer Settings

Haploid GFP-tagged yeast cells were illuminated using a 488 nm laser, and fluorescence was collected through a 505 nm long-pass and HQ510/20X band-pass filters (Chroma Technology Corp., Rockingham, VT, USA) on an LSR-II cytometer. Thresholds of 20,000 and 15,000 for the FSC and SSC gates, respectively, were typically used to exclude cellular debris. The data were recorded using the 'Area' option.

Diploid strains containing both GFP and tdTomato were illuminated by 488 nm and 532 nm lasers, respectively, also using an LSR-II cytometer. GFP fluorescence was collected through 505 nm long-pass and HQ515/20 nm band-pass filters. tdTomato fluorescence was collected through a 585/42 nm band-pass filter.

III. Data processing for Bulk Measurements

A. Mean and Median Calculations

Flow cytometry data were exported from the acquisition program (FACSDiva, Becton Dickinson, San Jose, CA, USA) in the FCS3.0 format with a data resolution of 2^{18}. Custom software was written to calculate statistics for each file according to the following rules:

1. The first second and last 0.2 seconds of data were removed to minimize errors due to unstable sample flow through the cytometer
2. All FSC and SSC data with minimal or maximal values (i.e., 0 and 2^{18}-1) were removed, as these are undefined
3. The bottom and top 5% of the FSC and SSC data were excluded to limit the influence of cellular debris and aggregated cells

The remaining data were used to calculate the mean and median GFP fluorescence. To further maintain the quality of the calculations, values were rejected if any of the following conditions were met:

1. Repeat measurements showed a deviation greater than 20% from the average of the fluorescence medians
2. The fluorescence medians calculated using the first and last halves of any given data set (i.e., in the time domain) showed a deviation greater than 20% from the average of the medians
3. The optical density of cells measured at harvest was less than 0.02 or greater than 0.3 (for a ~0.2 cm sample path-length)

B. Standard Errors of the Total Fluorescence

Median fluorescence measurements for cells grown under the same conditions were divided into bins spanning intensities from 100 to 100,000 units. For each bin, the standard error (σ) was calculated as:

$$\sigma \approx \sqrt{\frac{2 \times \sum_{i=1}^{n} \left(M_i - \overline{M_i} \right)^2}{n}} \qquad (1)$$

where M_i is one of the two measurements made on each sample, $\overline{M_i}$ is the average of the repeated measurements, and i is the index of one of n measurements for a given bin. The standard error (per bin) was plotted against the average bin value, and the data were fitted to a straight line passing through the origin. The slope was then used to calculate the expected error for each measurement (in fluorescence units). Standard errors were calculated separately for each growth medium.

C. Autofluorescence and Autofluorescence Standard Errors

Yeast cells fluoresce when illuminated at 488 nm. This fluorescence has been attributed to small-molecule metabolic intermediates[2]. In control studies using TAP-tagged strains (i.e., strains in which the *S. pombe* HIS5 selectable marker is present and proteins are tagged at their C-termini, but that have no GFP fluorescence) we observed heterogeneity in cellular autofluorescence, perhaps due to perturbation of the cells' normal metabolic state. This suggests that autofluorescence cannot be estimated from any single strain. Instead, we calculated autofluorescence from an analysis of the signals of all GFP-tagged strains grown under a given condition. More specifically, first, for a specified condition,

the distribution of fluorescence intensities was calculated (Supplementary Figure S1a). Next, the bin containing the greatest number of strains was identified ($F_{\max}$). Subsequently, the frequency values for all bins to the left of $F_{\max}$ were reflected in the vertical plane bisected by $F_{\max}$ to give a symmetric distribution (red line, Supplementary Figure S1b). This symmetric distribution captures information about autofluorescence variation, since the vast majority of GFP-tagged strains to the left of $F_{\max}$ have GFP fluorescence undetectable by cytometry: it concomitantly eliminates confounding effects due to the long tail to the right of the maximum that corresponds to detectable GFP fluorescence. Finally, the symmetric distribution was fitted to a single Gaussian curve. The offset of the Gaussian represents mean autofluorescence, and the width represents the standard error (dashed line, Supplementary Figure S1c). Note that this estimation also accounts for instrument drift observed across thousands of measurements.

D. Corrected Fluorescence and Corrected Fluorescence Measurement Errors

Corrected fluorescence (i.e., signal attributable to GFP) was calculated by subtracting the estimated autofluorescence from the total fluorescence for a given growth condition. Errors in each measurement were propagated as follows:

$$GFP \pm \sigma_{GFP} = (T - A) \pm \sqrt{\sigma_T^2 + \sigma_A^2} \qquad (2)$$

where *GFP* is signal due to GFP alone; *T* is total fluorescence; *A* is the estimated autofluorescence; and σ_{GFP}, σ_T, σ_A are the standard errors associated with each of these three measurements, respectively.

Corrected signals were averaged, and errors were propagated:

$$\overline{GFP} \pm \sigma_{\overline{GFP}} = \left(\frac{GFP_1 + GFP_2}{2} \right) \pm \sqrt{\frac{\sigma_{GFP1}^2}{4} + \frac{\sigma_{GFP2}^2}{4}} \qquad (3)$$

where $\overline{GFP}$ is the average corrected signal, σ_{GFP1} and σ_{GFP2} are the standard errors for each independent measurement, and $\sigma_{\overline{GFP}}$ is the average standard error (see Supplementary Table S1).

E. Distinguishing Corrected Fluorescence from Autofluorescence

We fitted a single Gaussian curve to a histogram of the averaged, corrected GFP values as described above. Values larger than $1.7\sigma_A$ (note the average corrected fluorescence is now 0) have >95% chance of being due to GFP fluorescence alone.

F. Number of False Negative and False Positives

We compute false positives (*FP*) by dividing the number of autofluorescent strains that are incorrectly identified as GFP-positive by the total number of strains considered to be GFP-positive (red area, Supplementary Figure S2):

$$FP \approx \frac{0.05 \times NA}{NG} \qquad (4)$$

where *NA* is the estimated number of autofluorescent strains, and *NG* is the number of strains defined as GFP positive (i.e., strains that have fluorescence $>1.7\sigma_A$).

The number of false negatives (*FN*) is calculated by dividing the number of strains that have fluorescence $<1.7\sigma_A$, but whose fluorescence is probably due to GFP, by the total number of strains that are likely to be GFP-positive (blue area, Supplementary Figure S2):

$$FN \approx \frac{Total - NG - (0.95 \times NA)}{Total - NA} \qquad (5)$$

where *Total* is the total number of measured strains. The results from these calculations are summarized in Supplementary Table S3.

G. Errors in Ratios

We calculated the ratio of fluorescence measured in SD compared to YEPD as $R_{S/Y}$. The corresponding error was determined using the following equation:

$$R_{S/Y} \pm \sigma_R = \frac{\overline{GFP}_{SD}}{\overline{GFP}_{YEPD}} \pm \sqrt{\frac{\sigma_{SD}^2}{\overline{GFP}_{YEPD}^2} + \frac{\overline{GFP}_{SD}^2 \sigma_{YEPD}^2}{\overline{GFP}_{YEPD}^4}} \qquad (6)$$

where σ_R is the standard error of the ratio, and σ_{SD} and σ_{YEPD} are the standard errors calculated for the average corrected GFP measurements made in SD and YEPD media, respectively.

H. Errors After Log Transformation

Changes in levels of biological molecules are often reported using the intuitive scale of $\log_2$ (i.e., unchanged values are centered about 0 and each unit on the log scale reflects a 2-fold induction or repression). The $\log_2$ data with associated error are:

$$Log_2\left(R_{S/Y}\right) \pm \sqrt{\sigma_{SD}^2\left[\frac{1}{\overline{GFP_{SD}}\ln(2)}\right]^2 + \sigma_{YEPD}^2\left[\frac{1}{\overline{GFP_{YEPD}}\ln(2)}\right]^2} \quad (7)$$

Log_2 ratios for protein and mRNA measurements are listed in Supplementary Tables S1 and S2.

IV. Data Processing for Single-Color, Single-Cell Measurements

A. Calculating Coefficients of Variation

For calculating protein variation, cytometry files were processed as follows:

1. The first second and last 0.2 seconds of data were removed to minimize errors due to unstable sample flow through the cytometer
2. All FSC and SSC data with minimal or maximal values (i.e., 0 and 2^{18}-1) were removed, as these are undefined
3. The bottom and top 5% of the FSC and SSC data were excluded to further limit the influence of cellular debris and aggregated cells
4. For the remaining data, the FSC and SSC medians were calculated and the distance of the i^{th} sample to the medians was determined:

$$\text{Distance}_i = \sqrt{\left(\text{FSC}_i - \text{FSC}_{\text{median}}\right)^2 + \left(\text{SSC}_i - \text{SSC}_{\text{median}}\right)^2} \qquad (8)$$

5. The smallest and largest 5% of the GFP values situated within a radius of 4,096 centered about the FSC and SSC medians were discarded and the remaining GFP data within the radius were used to calculate the (percent) coefficient of variation:

$$\text{CV} = \frac{\text{Standard deviation}}{\text{mean}} \times 100 \qquad (9)$$

Calculations were rejected if any of the following criteria were met:

1. Corrected median fluorescence for the gated values was $<1.7\sigma_A$ (where σ_A is the autofluorescence standard deviation, calculated as described for the bulk measurements)
2. Fewer than 120 measurements were used to calculate CVs
3. Repeat measurements did not agree to within ±20% of the average CV

2213 pairs of CV measurements made in YEPD and 2058 pairs of CV measurements for cells grown in SD pass the above criteria.

B. Choice of Circular Gate Radius

The rationale for choosing a radius of 4,096 can be seen from Supplementary Figure S3a. CVs decrease as the gate radius is reduced until the radius is about 5000 units. From 1000-5000 units the CVs are fairly constant. Thus, a radius of 4,096 represents a balance between choosing a gate yielding the lowest CV possible, while maintaining sufficient numbers of cells to calculate CVs accurately.

Supplementary Figures S3b and S3c show the distributions of the number of cells used to calculate CVs in YEPD and SD, respectively, and confirm that most calculations are based on several hundred independent measurements.

Supplementary Figures S3d and S3e show the reproducibility of CV measurements in YEPD and SD, respectively.

V. Data Processing for Dual-Color, Single-Cell Measurements

A. Calculating Intrinsic, Extrinsic, and Total Noise

Data were processed according to the following rules:

1. The first and last 0.2 seconds of data were removed to minimize errors due to unstable sample flow through the cytometer
2. All FSC and SSC data with minimal or maximal values (i.e., 0 and 2^{18}-1) were removed, as these are undefined
3. The bottom and top 5% of the FSC and SSC data were excluded to further limit the influence of cellular debris and aggregated cells
4. FSC and SSC medians were calculated and the distance of the i^{th} sample to the medians was determined using Equation (8)
5. GFP and tdTomato values situated within a gate having a radius of 4,096 centered about the FSC and SSC medians were normalized with respect to their individual median values
6. A second gate defined by a 24-sided polygon centered about median GFP and median tdTomato values was used to exclude cells that give rise to rare, unrepresentative fluorescence measurements. The gate lies in the plane determined by the GFP and tdTomato fluorescence values. The 12 parallel sides of the gate are calculated such that they individually encompass 98% of the total number of data points situated on the plane

Intrinsic, extrinsic, and total noise are calculated for points inside the polygon (as per the definitions of Elowitz *et al.*[3]):

$$\eta_{in}^2 = \frac{\left\langle (g-t)^2 \right\rangle}{2\langle g \rangle \langle t \rangle} \tag{10}$$

$$\eta_{ex}^2 = \frac{\langle gt \rangle - \langle g \rangle \langle t \rangle}{\langle g \rangle \langle t \rangle} \tag{11}$$

$$\eta_{tot}^2 = \frac{\langle g^2 + t^2 \rangle - 2\langle g \rangle \langle t \rangle}{2\langle g \rangle \langle t \rangle} \quad (12)$$

where g represents the GFP fluorescence of a single cell, and t represents the tdTomato fluorescence. The angled brackets denote a mean value calculated for all cells in a population.

B. Independence

To investigate the extent to which the signal given by one fluorescent protein is influenced by the presence of a second fluorescent protein, we compared strains expressing a single tagged protein to those expressing two tagged proteins. With the exception of Rps25Ap, strains containing a single fluorescent tag are influenced minimally by the presence of a second one.

	GFP Fluorescence		tdTomato Fluorescence		% Different from Ave.	
Strain	**GFP x WT**	**GFP x tdT**	**tdT x WT**	**GFP x tdT**	**GFP**	**tdTomato**
TEF2	24285	24703	22497	22934	2	2
RPS25A	6333	8948	5813	8826	34	41
ATP4	1079	1260	1218	1333	15	9
HIS4	2224	2314	2314	2277	4	2

The fluorescence intensities are uncorrected and reported in arbitrary units.

Abbreviations: GFP, green fluorescent protein; tdT, tandem dimer Tomato fluorescent protein; WT, wildtype. GFP bleed-through into the tdTomato channel is 0.29%; tdTomato bleed-through into the GFP channel is 0.17%.

C. Equivalence

To investigate the equivalence of the two fluorophores (GFP and tdTomato), we used the Kolmogorov-Smirnov test to calculate the chance that the two datasets composed of the normalized fluorescence values for each fluorophore are drawn from the same distribution. Upon an initial inspection, we found that there were statistically significant differences between the two datasets:

Strain	Number Cells	KS Value
TEF2	5461	0.0223
RPS25A	5562	2.05E-06
ATP4	5221	0.0946
HIS4	5095	0.00224

These statistical results were somewhat surprising given the large degree of overlap of the histograms of the normalized fluorescence intensities for the four proteins studied (Supplementary Figure S4). Further investigations revealed that there was a strong dependence of the KS values on the total numbers of cells used to calculate the statistic. Measurement of fewer cells (e.g., in the range of the numbers used in 2-color experiments performed by microscopy[4]) gave substantially higher KS values, with no change in the calculated levels of intrinsic or extrinsic noise:

Strain	Number Cells	KS Value
TEF2	159	0.228
RPS25A	182	0.664
ATP4	170	0.871
HIS4	147	0.426

These values are similar to those reported in the previously cited study (cf. 0.27-1.00[4]). Finally, we investigated whether the fluorescent distributions of GFP and tdTomato could be made more similar by stopping protein production (by addition of cycloheximide, CHX) and waiting several hours to let incompletely folded and oxidized proteins attain their final fluorescent state[5]. We observed no difference between the distributions obtained before or after CHX addition (data not shown).

In summary, the formal tests for independence and equivalence indicate that the combination of the GFP and tdTomato tags can be used to extract information about intrinsic and extrinsic noise. However, they remain imperfect pairs: tagging of certain classes of proteins (e.g., membrane proteins) does not result in collection of meaningful data (data not shown). Nevertheless, at least some of the statistical differences seen between the fluorophores are due to the inherent precision of cytometry and its ability to measure thousands of cells.

VI. Experimental Controls

A. Measurement Reproducibility (Fluorescence)

To assess the influence of the delay between harvesting cells and performing cytometry on cellular fluorescence, 96 strains were measured twice sequentially. As seen in Supplementary Figure S5, the fluorescence intensities for the first and second measurements are highly similar, arguing that experimental delays do not substantially change protein levels. Approximately 15 minutes elapses between cell harvesting and the first measurement. Each set of 96 measurements requires 14 minutes.

B. Measurement Reproducibility (Coefficients of Variation)

As shown in Supplementary Figures S3d and S3e, CVs calculated for full biological replicates are highly reproducible.

C. Instrument-Independence of Fluorescence Measurements

To assess possible instrument-related biases in measuring fluorescence, we compared the bulk fluorescence signals for a series of strains quantified by cytometry to the integrated fluorescence intensities determined by microscopy. As seen in Supplementary Figure S6a there is excellent agreement between the two methods (R^2=0.997).

D. Instrument-Independence of CV Measurements

To investigate the instrument-dependence of CV values, we compared CVs determined using cytometry to those obtained using microscopy. For these studies, seven GFP-tagged strains were selected to encompass a broad range of CVs. As can be seen in

Supplementary Figure S6b, there is good overall agreement between the two different methods used to measure CVs. Note that for CV calculations based on cytometry data we did not gate the cells.

E. Instrument Contribution to CV Measurements

We performed the following experiment to measure the instrument response noise. The fluorescence from a yeast strain expressing Gal1p-YFP was measured by two distinct PMTs. Subsequently, the intensity of the fluorescence light was reduced by sequential introduction of neutral-density filters designed to block increasing amounts of light. For each neutral-density filter, the resulting fluorescence was recorded for 100,000 cells. The data were then processed to calculate the intrinsic and extrinsic noise for a given fluorescence intensity/neutral density filter. The intrinsic noise quantifies the signal-dependence of the instrument response.

F. Proportionality of Fluorescence and Western Blotting Measurements

One potential concern is that the green fluorescent protein yields different amounts of fluorescence depending on the protein to which it has been fused. This situation might arise due to incomplete folding or oxidation of GFP, or to unanticipated effects due to differences in cellular environments. Using western blotting to measure the total amount of tagged protein in the cell, and cytometry to measure the fluorescence signal produced by GFP, we find a good overall agreement between total protein and fluorescence (R^2=0.80) (Supplementary Figure S7a). Indeed, this correlation is fully consistent with errors from western blotting alone (R^2=0.77) (Supplementary Figure S7b). Note also that

the western blotting used TAP-tagged strains[6], thus providing evidence that protein levels are not differentially affected by the tag used in preparing the yeast library (i.e., TAP or GFP).

A second concern is that the measured fluorescence results from free GFP, rather than from GFP-protein fusions. To investigate this systematically, we compared the actual and expected molecular weights for more than 150 different GFP-tagged proteins, chosen to encompass more than 20 cellular localizations. As noted in the main text, with the exception of vacuolar proteins, most GFP-fusions remain intact *in vivo* (Supplementary Figure S7c and legend).

G. Degradation of Proteins Regulated by Ubiquitin-mediated Proteolysis

To assess the ability of GFP-tagged strains to be degraded by ubiquitin-mediated proteolysis, strains containing Sic1-GFP and Clb2-GFP were studied by western blotting. Logarithmically growing strains were arrested by addition of alpha factor, and then released into fresh medium. As shown in Supplementary Figure S8, after release from alpha-factor arrest, both strains show the characteristic oscillations in protein levels associated with cell-cycle regulated proteins. Importantly, destruction of the tagged proteins is not accompanied by the liberation and persistence of free GFP. This observation has been extended to a range of other cell-cycle regulated targets (David Toczyski, personal communication).

H. Confirmation of Examples of Discordant mRNA and Protein Changes

To confirm potential instances of post-transcriptional regulation we found it necessary to first establish confidence in the mRNA measurements. We did this by independently repeating microarray experiments four times, considering only those instances where messages were unchanged when cells were grown in SD or YEPD. Next, protein changes determined by cytometry were confirmed by western blotting experiments using tagged strains. Microarray experiments were confirmed by qPCR, using primers hybridizing to the sequence encoding the tag (Supplementary Figure S9). Western blots and qPCR experiments were performed on the same samples, were each internally controlled (using hexokinase for the western blots and actin for qPCR), were in the linear range of the measurement, and the signals were above background levels.

I. Confirmation of Reduction in Cell-Cycle Variation by Gating

Spellman *et al*. proposed a metric (termed the CDC score) to capture information about the amplitude of cell-cycle oscillations in mRNA levels[7]. We find that, for abundant proteins, there is a strong correlation between CDC score and the appearance of cell-cycle variation at the protein level (Supplementary Figure S10a). However, after gating (see Section IV), we find little correlation between CDC scores and DM values (where the later is an abundance-independent measure of protein variation; Supplementary Figure S10b; see also Figure 2c).

J. Independence of Environmental Effects on SAGA Targets

SAGA-targeted genes often respond to environmental perturbations[8]. Therefore, one potential concern is that by using 96-well × 2-ml blocks to grow cells, we are exposing cells to varying environmental conditions, and thus detecting intercellular rather than intracellular variation. Two independent lines of evidence argue against this. First, the reproducibility of CV measurements is insensitive to whether a gene is a target of SAGA or not:

Protein Class	Mean |% Diff.| DM	Median |% Diff.| DM
SAGA/TFIID	3.84	3.13
SAGA	3.86	3.22
TFIID	4.07	3.22
All	4.02	3.21

where |% Diff.| is the absolute value of the percent difference of one DM value from the average of both DM values.

Second, proteins encoded by SAGA-targeted genes still have large CVs even when cells are grown in a single-tube format (data not shown). The single-tube format presumably generates more uniform growth conditions than the 96-well × 2-ml blocks due to better aeration and mixing, and the absence of any mechanical stress arising from the presence of a glass bead in the 2-ml wells.

K. Dependence of Noise on Cellular Localization

To explore the effect of cellular localization on protein variation in more detail two GFP derivatives were constructed. One consisted of GFP fused to a *bona fide* peroxisomal-targeting sequence (Pro-Leu-His-Ser-Lys-Leu, PTS1), the second of GFP fused to this sequence in reverse (Leu-Lys-Ser-His-Leu-Pro, pts1*). Both proteins were then expressed from identical promoters, and the variation measured. Localization to peroxisomes of GFP-PTS1 was verified by microscopy. As shown in Supplementary Figure S11a, targeting GFP to peroxisomes results in increased protein noise.

To further explore the role of correct organelle partitioning on protein variation we introduced a mutation (Δ*inp1*) that has been shown to disrupt correct peroxisome segregation[9]. Proteins localized to peroxisomes (GFP-PTS1, Pex11p-GFP, Pex25p-GFP and Mdh2p-GFP) show an increased level of noise in Δ*inp1* strains. This increase in noise is not seen for proteins localized to other cellular compartments (e.g., GFP-pts1*, Sec28p-GFP, Mrp8p-GFP and Tpo4p-GFP) or for the parental strain (BY4741) upon introduction of the same mutation (Supplementary Figure S11b).

VII. Statistical Calculations

A. Correlations between quantities

To assess the statistical significance of correlations between quantities, P-values were calculated using the Spearman rank correlation test. Quantities investigated are listed below, and the data sources are listed in Supplementary Table S6:

1. mRNA copy number
2. mRNA half-life
3. CAI score
4. Ribosome density
5. Number of proteins per mRNA
6. Protein copy number
7. Protein half-life

B. Noisy/quiet gene classifications

To assess whether genes assigned to a given classification have larger or smaller protein variation than expected by chance, we used the Kolmogorov-Smirnov (KS) test and the Wilcoxon rank-sum test. Both tests yielded similar results over a range of P-values (Supplementary Figure S12). For larger P-values the Wilcoxon test gives increasingly smaller P-values than the KS-test. Cellular parameters investigated are listed below, and the data sources are listed in Supplementary Table S6:

1. Proximity to nuclear pore components
2. Presence or absence of a TATA box in a gene's promoter
3. Histone modification
4. Target of chromatin remodeling complex
5. Target of chromatin modifying complex
6. GO-categories
7. Transcription modules

C. Gene Proximity

Chromosomal distances between genes were downloaded from SGD. For each gene, the neighboring gene with the smallest chromosomal distance was identified. To test for effects of chromosomal location, we calculated two metrics:

$$d_{ij} = \left| v_i - v_j \right| \tag{13}$$

and

$$d_{ij} = \left| v_i - v_j \right| \Big/ \left| v_i + v_j \right| \tag{14}$$

where v_i is the noise (DM or CV) associated with gene i. The resulting distribution of d_{ij} for neighboring genes was compared to the distribution of random gene pairs. No significant difference was observed for either metric (independent of whether DM or CV values were used; Supplementary Figure S13).

VIII. Supplementary Discussion of Global Structure of Noise

Abundance is the major factor governing protein variation. Moreover, the shape and magnitude of the global relationship between noise and abundance strongly argues that variation most likely originates from the stochastic production and destruction of mRNA molecules. We arrive at this conclusion by noting that, for a typical protein, noise can be divided into two components. One component arises from cell-to-cell heterogeneity ($CV_{\text{Het.}}$), due to factors affecting cells' ability to express proteins both in a global and a protein-specific manner (e.g., due to differences in cell size that are not negated by cytometry and differences in the concentration of a particular transcription factor, respectively). A second component consists of intrinsic sources of noise that should scale as $1/\langle n_{\text{protein}} \rangle$ and $1/\langle n_{\text{mRNA}} \rangle$, assuming that protein and mRNA production for a typical protein occurs in a manner following Poisson statistics. However, fluctuations in message numbers (n_{mRNA}) are reduced by a noise-filter that arises because proteins are almost always longer-lived than mRNAs. Thus, the noise from a typical protein may be approximated by Equation 15 (see Ref. 10):

$$CV^2_{\text{protein}} \approx \overbrace{CV^2_{\text{Het.}}}^{\text{expression noise}} + \overbrace{\frac{1}{\langle n_{\text{protein}} \rangle}}^{\text{stochastic protein noise}} + \overbrace{\frac{1}{\langle n_{\text{mRNA}} \rangle} \underbrace{\frac{1/\tau_2}{1/\tau_1 + 1/\tau_2}}_{\text{time-averaging}}}^{\text{stochastic mRNA noise}} \qquad (15)$$

where τ_1 and τ_2 are mRNA and protein half-lives, respectively. Of the two stochastic terms, mRNA copy number will contribute dominantly, as mRNA molecules are generally at least 1000-fold less abundant than proteins (indeed, 80% of mRNAs are present at ≤ 2 copies/cell[11]). Yet, there will still be an apparent proportionality between noise and $1/\langle n_{\text{protein}} \rangle$ as there generally is a constant relationship between protein and mRNA molecule numbers over the full range of protein abundances[6].

These arguments now permit us to understand the overall dependence of protein noise on protein abundance. For abundant proteins, $1/\langle n_{\mathrm{mRNA}} \rangle$ will be negligible and therefore $CV_{\mathrm{Het.}}$ will dominate, thus explaining the lack of concentration dependence in this regime. In contrast, at lower abundances, the concentration-dependent term will make a more significant contribution to total noise, leading to the observed slope of -1 in the plot of $\log(CV^2)$ vs. $\log\langle n_{\mathrm{protein}} \rangle$ (Figure 2g, Main Text). Indeed, the magnitude of variation observed here is consistent with known half-lives for proteins and mRNA copy numbers of low and medium abundance proteins. That is, given that there are 1-2 messages/cell[11] for proteins at the mid-to-lower range of observed abundances, and assuming noise filtering occurs due to the fact that proteins have 5- to 10-fold longer half-lives than mRNAs ($t_{1/2}$ for mRNAs ~9 minutes[12], whereas the effective lifetime of a protein will be limited to the 90 minute doubling time of yeast), one would predict $CV^2 \approx 1/10$ or $CV \approx 30\%$ (Figure 2g).

IX. Correlations Between Noise and Biological Parameters

A. Rationale

In our analyses of how different cellular properties are correlated with noise, we attempted to move methodically from the physical gene environment, through the production of mRNAs and proteins, to the organization and use of proteins in the cell. Our results are far from comprehensive, and they depend heavily on data availability and quality. We cannot rule out that our own measurements, by reporting noise both intrinsic and extrinsic to protein production and folding, obscure important biological findings, and this will be the subject of future studies.

A brief synopsis of our results are presented in Figure 4 (Main Text). A more comprehensive list of properties and calculated P-values is presented in Supplementary Table S5, with references for the data sources listed in Supplementary Table S6.

B. Organization of GO-Term, Transcription Factor, and Transcription Module Findings

There are three sources of redundancy in the gene sets that we considered. First, different definitions can give rise to similar or identical gene sets (e.g., transcription factor target genes may overlap with transcription modules of co-expressed genes). Second, GO-term definitions were modified to include not only the most specific association of a gene with a GO term, but also all of its parent terms in the directed acyclic graph. This ensured that highly subcategorized processes can still be identified as significant, but gives rise to redundant definitions in many cases. Third, transcription modules were identified for a range of resolutions, giving rise to sequences of modules differing slightly in their size but otherwise very similar.

To account for these instances of redundancy, we first identified all gene sets with a significance of $P < 10^{-4}$. For each pair of sets, an overlap score was defined:

$$overlap_{ij} = \left| s_i \cap s_j \right| \Big/ \sqrt{\left(\left| s_i \right| \cdot \left| s_j \right| \right)} \qquad (16)$$

and the resulting square matrix of pairwise overlap scores was clustered using a hierarchical clustering algorithm (Supplementary Figure S14)[13]. For each cluster of highly similar overlaps, we selected a representative gene set with the lowest P-value when preparing Figure 4. The list of all clusters and their constituent gene sets is given in Supplementary Table S5.

C. Relationship Between Variation in mRNA Expression Levels and DM

Our data suggest that transcription plays an important role in defining protein variation. To better understand how transcription and protein variation are linked, we examined the relationship between the flexibility of transcription (defined as the extent to which the expression levels of genes can change) and protein noise (measured using the DM parameter described in the main text).

To calculate expression flexibility we determined the standard deviation in mRNA ratios obtained across many different conditions for a particular gene. Next, proteins were binned according to the standard deviations of their mRNA ratios, and the median DM for each bin was calculated. As shown in Supplementary Figure S15, genes that are capable of large changes in their mRNA levels encode proteins that exhibit large amounts of variation.

D. R^2 Values for Major Factors Contributing to Biological Noise

The R^2 values for the relationships described in Figure 4b are listed below:

Test	$-\log_{10}(p)$	#components	R^2
mRNA/cell vs. CV	109.0	1836	0.23
mRNA $T_{1/2}$ vs. CV	3.8	1429	0.01
mRNA σ vs. DM	61.8	2196	0.12
CAI vs. DM	3.0	2188	0.0005
Rib. dens. Vs. DM	6.5	2065	0.012
Prot./mRNA vs. DM	5.9	1893	0.012
1/(prot./cell) vs. CV	321.0	1429	0.64
Number PPI vs. DM	0.3	1523	7.5×10^{-4}

Note that gene proximity is based on a rank-sum test of neighbors versus non-neighbors and therefore does not have an associated R^2 value.

R^2 values can obscure strong relationships (such as that existing between mRNA variation and DM; Supplementary Figure S15) where the distribution is highly skewed.

X. Strain Genotypes

Strain genotypes are listed below. The GFP library was constructed using the S288C derivative BY4741[1, 14]. For the two-color experiments, the GFP tag was integrated into BY4742 and the tdTomato tag into BY4741. Both fluorescent proteins were selected using the HIS5 gene of *Schizosaccharomyces pombe*. Diploid strains were obtained by crossing BY4741 and BY4742 derivatives, and selecting on SD-Lys-Met.

Strain	ATCC Number	Mating Type	HIS	LEU	MET	LYS	URA
BY4741	201388	*MAT***a**	*his3Δ1*	*leu2Δ0*	*met15Δ0*	+	*ura3Δ0*
BY4742	201389	*MATα*	*his3Δ1*	*leu2Δ0*	+	*lys2Δ0*	*ura3Δ0*

XI. Molecular Biology Techniques and Plasmids

A. Microarray Analysis

200 ml of cells were grown to OD_{600} ~0.7. The cells were harvested by vacuum filtration onto nitrocellulose and frozen at –80°C. Total RNA was prepared following the protocol available at:

http://derisilab.ucsf.edu/pdfs/TotalRNAIsolation.pdf

Poly-$(A)^+$ mRNA was prepared following the protocol available at:

http://derisilab.ucsf.edu/pdfs/polyARNAIsolation.pdf

mRNA was reverse transcribed and labeled as described previously using the protocol available at:

http://derisilab.ucsf.edu/pdfs/amino-allyl-protocol.pdf

except that DNA Clean & Concentrator-5™ columns (Zymo Research Corporation, Orange, CA, USA) were used to remove free dye, and sample was loaded onto these columns and spun through twice.

Microarray ratios are listed in Supplementary Table S2.

B. Tagging ORFs with Fluorescent Proteins

C-terminal fusions at chromosomal loci were prepared by transforming yeast strains with a PCR product containing DNA encoding a fluorescent protein and a selectable marker. The PCR products were flanked by DNA with homology to the final 50 nucleotides of an ORF, and to the 50 nucleotides immediately following the ORF stop codon. Homologous integration results in the exact fusion of the last amino acid of the endogenous protein with the first amino acid of the fluorescent protein tag.

Confirmation of integration was based on the observation of a PCR product generated by amplification of a genomic DNA template using a forward primer within the targeted ORF and a reverse primer within the integrated sequence. Typically, multiple transformants confirmed by check-PCR were used in cytometry studies to minimize the impact of artifacts associated with errors during PCR or integration.

Primer sequences are listed in Supplementary Table S4. Information about plasmid templates used to generate GFP or tdTomato fusion proteins is provided below.

C. ORF Disruption

INP1 was deleted by using homologous recombination to replace the entire coding sequence with a selectable marker encoding either the Kan or Nat resistance genes (see below). INP1 was deleted in the strains indicated in the main text. Asymmetric segregation of peroxisomes in Δ*inp1* was confirmed by microscopy in strains where the localization of the GFP-tagged protein made this feasible.

D. PCR Protocol

PCR products for transformation were amplified using the iProof polymerase (Bio-Rad Laboratories, Hercules, CA, USA), and primers and plasmids specified in the Supplementary Materials. PCR conditions were:

98°C	30"	
98°C	10"	} x 30
55°C	30"	
72°C	1:30'	
72°C	7'	

E. Transformation Protocol

PCR products for ORF tagging or disruption were transformed into yeast according to the method described by Gietz *et al.*[15]

F. Construction of GFP-PTS1/GFP-pts1* Sequences

GFP (S65T) was fused with a C-terminal peroxisomal targeting sequence (PTS1) consisting of the six amino acids Pro-Leu-His-Ser-Lys-Leu[16], or with a reversed version of this sequence (pts1*) that causes cytosolic localization (Leu-Lys-Ser-His-Leu-Pro) using the polymerase chain reaction. These two GFP-derivatives were cloned downstream of a 237 basepair sequence from the *S. cerevisiae* ILS1 promoter and upstream of a 213 basepair sequence from the *A. gossypii* ADH1 terminator into a pFA6a-KANMX4 plasmid. We find that the ILS1 promoter induces moderate levels of expression and that the *A. gossypii* ADH1 terminator is effective in *S. cerevisiae*. The ILS1-GFP-PTS1-ADH1 construct (and the control reversed version) were then integrated into the yeast strain BY4741 either at the AGA2 locus (causing deletion of AGA2) or at

the *his3Δ0* locus. Integration at either locus gave essentially identical results in our assays. Correct localization of the GFP construct targeted to the peroxisome (and the cytosolic control) was confirmed by microscopy (data not shown).

G. Plasmids

pFA6a-KANMX4

The GenBank accession number for pFA6a-KANMX4 is AJ002680.

pFA6a-GFP(S65T)-HIS3MX6

The GenBank accession number for pFA6a-GFP(S65T)-HIS3MX6 is AJ002683.

pFA6a-tdTomato-HIS3MX6

The sequence of pFA6a-tdTomato-HIS3MX6 is listed below. Researchers interested in obtaining this plasmid must sign an MTA from the laboratory of Dr. Roger Tsien (University of California, San Diego, California, USA) before we can distribute it. Primer binding sites are highlighted in yellow. Repeat copies of dTomato are shown in red. Joining sequences are shown in orange. The HIS5 selectable marker is shown in blue.

GAACGCGGCCGCCAGCTGAAGCTTCGTACGCTGCAGGTCGACGGATCCCCGGGTTAATTAACgtgagcaag
ggcgaggaggtcatcaaagagttcatgcgcttcaaggtgcgcatggagggctccatgaacggccacgagtt
cgagatcgagggcgagggcgagggccgcccctacgagggcacccagaccgccaagctgaaggtgaccaagg
gcggccccctgcccttcgcctgggacatcctgtccccccagttcatgtacggctccaaggcgtacgtgaag
caccccgccgacatccccgattacaagaagctgtccttccccgagggcttcaagtgggagcgcgtgatgaa
cttcgaggacggcggtctggtgaccgtgacccaggactcctccctgcaggacggcacgctgatctacaagg
tgaagatgcgcggcaccaacttcccccccgacggccccgtaatgcagaagaagaccatgggctgggaggcc
tccaccgagcgcctgtaccccgcgacggcgtgctgaagggcgagatccaccaggccctgaagctgaagga
cggcggccactacctggtggagttcaagaccatctacatggccaagaagcccgtgcaactgcccggctact
actacgtggacaccaagctggacatcacctcccacaacgaggactacaccatcgtggaacagtacgagcgc

tccgagggccgccaccacctgttc**ctg**gggcatggcaccggcagcaccggcagcggcagctccggcaccgc
ctcctccgaggacaacaacatggcc**gtc**atcaaagagttcatgcgcttcaaggtgcgcatggagggctcca
tgaacggccacgagttcgagatcgagggcgagggcgagggccgcccctacgagggcacccagaccgccaag
ctgaaggtgaccaagggcggccccctgcccttcgcctgggacatcctgtccccccagttcatgtacggctc
caaggcgtacgtgaagcaccccgccgacatccccgattacaagaagctgtccttccccgagggcttcaagt
gggagcgcgtgatgaacttcgaggacggcggtctggtgaccgtgacccaggactcctccctgcaggacggc
acgctgatctacaaggtgaagatgcgcggcaccaacttcccccccgacggccccgtaatgcagaagaagac
catgggctgggaggcctccaccgagcgcctgtaccccgcgacggcgtgctgaagggcgagatccaccagg
ccctgaagctgaaggacggcggccactacctggtggagttcaagaccatctacatggccaagaagcccgtg
caactgcccggctactactacgtggacaccaagctggacatcacctcccacaacgaggactacaccatcgt
ggaacagtacgagcgctccgagggccgccaccacctgttc**ctg**tacggcatggacgagctgtacaag**taa**G
GCGCGCCACTTCTAAATAAGCGAATTTCTTATGATTTATGATTTTTATTATTAAATAAGTTATAAAAAAAA
TAAGTGTATACAAATTTTAAAGTGACTCTTAGGTTTTAAAACGAAAATTCTTATTCTTGAGTAACTCTTTC
CTGTAGGTCAGGTTGCTTTCTCAGGTATAGTATGAGGTCGCTCTTATTGACCACACCTCTACCGGCAGATC
CGCTAGGGATAACAGGGTAATATAGATCTGTTTAGCTTGCCTCGTCCCCGCCGGGTCACCCGGCCAGCGAC
ATGGAGGCCCAGAATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGTGACTGTCGCCC
GTACATTTAGCCCATACATCCCCATGTATAATCATTTGCATCCATACATTTTGATGGCCGCACGGCGCGAA
GCAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGGGAAACGCTCCCCTCACAGACGCGTTGAATTGT
CCCCACGCCGCGCCCCTGTAGAGAAATATAAAAGGTTAGGATTTGCCACTGAGGTTCTTCTTTCATATACT
TCCTTTTAAAATCTTGCTAGGATACAGTTCTCACATCACATCCGAACATAAACAACCATGGGTAGGAGGGC
TTTTGTAGAAAGAAATACGAACGAAACGAAAATCAGCGTTGCCATCGCTTTGGACAAAGCTCCCTTACCTG
AAGAGTCGAATTTTATTGATGAACTTATAACTTCCAAGCATGCAAACCAAAAGGGAGAACAAGTAATCCAA
GTAGACACGGGAATTGGATTCTTGGATCACATGTATCATGCACTGGCTAAACATGCAGGCTGGAGCTTACG
ACTTTACTCAAGAGGTGATTTAATCATCGATGATCATCACACTGCAGAAGATACTGCTATTGCACTTGGTA
TTGCATTCAAGCAGGCTATGGGTAACTTTGCCGGCGTTAAAAGATTTGGACATGCTTATTGTCCACTTGAC
GAAGCTCTTTCTAGAAGCGTAGTTGACTTGTCGGGACGGCCCTATGCTGTTATCGATTTGGGATTAAAGCG
TGAAAAGGTTGGGGAATTGTCCTGTGAAATGATCCCTCACTTACTATATTCCTTTTCGGTAGCAGCTGGAA
TTACTTTGCATGTTACCTGCTTATATGGTAGTAATGACCATCATCGTGCTGAAAGCGCTTTTAAATCTCTG
GCTGTTGCCATGCGCGCGGCTACTAGTCTTACTGGAAGTTCTGAAGTCCCAAGCACGAAGGGAGTGTTGTA
AAGAGTACTGACAATAAAAAGATTCTTGTTTTCAAGAACTTGTCATTTGTATAGTTTTTTTATATTGTAGT
TGTTCTATTTTAATCAAATGTTAGCGTGATTTATATTTTTTTTCGCCTCGACATCATCTGCCCAGATGCGA
AGTTAAGTGCGCAGAAAGTAATATCATGCGTCAATCGTATGTGAATGCTGGTCGCTATACTGCTGTCGATT
CGATACTAACGCCGCCATCCAGTTTAAACGAGCTCGAATTCATCGATGATATCAGATCCACTAGTGGCCTA
TGCGGCCGCGGATCTGCCGGTCTCCCTATAGTGAGTCGTATTAATTTCGATAAGCCAGGTTAACCTGCATT
AATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGAC
TCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCAC
AGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGG
CCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAG
AGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCC
TGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAAT
GCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCC
GTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATC
GCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGA
AGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACC
TTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTG
CAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACG
CTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATC
CTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCA
ATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCG
TCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCA
CGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGC
AACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATA
GTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTC
AGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTT
CGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATA
ATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGA
GAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAG

AACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGA
GATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCT
GGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACT
CATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTG
AATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAA
GAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTT
CGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATG
CCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCG
GCATCAGAGCAGATTGTACTGAGAGTGCACCATATGGACATATTGTCGTTAGAACGCGGCTACAATTAATA
CATAACCTTATGTATCATACACATACGATTTAGGTGACACTATA

pFA6a-ILSPr-eGFP-PTS1-ADH1-KANMX6

The sequence of the plasmid encoding EGFP fused to an intact PTS1 sequence driven by the ILS1 promoter is listed below. The ILS1 promoter is highlighted in blue, the start codon of the GFP is shown in green, the sequence encoding the PTS1 tag is highlighted in yellow, the stop codon of the protein is shown in red, and the ADH1 terminator from *A. gossypii* is highlighted in grey. The 20-bp sequences shown in red can be used to amplify the construct prior to integration into the yeast genome. The bold sequence denotes the KAN resistance gene.

GAACGCGGCCGCCAGCTGAAGCTTCGTACGCTGCAGGTCGA**CGGATCCCCGGGTTAATTAA**GGCGCGCCAG
ATCTGTTTAGCTTGCCTCGTCCCCGCCGGGTCACCCGGCCAGC*GACATGGAGGCCCAGAATACCCTCCTTG*
ACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGTGACTGTCGCCCGTACATTTAGCCCATACATCCCCAT
GTATAATCATTTGCATCCATACATTTTGATGGCCGCACGGCGCGAAGCAAAAATTACGGCTCCTCGCTGCA
GACCTGCGAGCAGGGAAACGCTCCCCTCACAGACGCGTTGAATTGTCCCCACGCCGCGCCCCTGTAGAGAA
ATATAAAAGGTTAGGATTTGCCACTGAGGTTCTTCTTTCATATACTTCCTTTTAAAATCTTGCTAGGATAC
*AGTTCTCACATCACATCCGAACATAAACAACC***ATGGGTAAGGAAAAGACTCACGTTTCGAGGCCGCGATTA**
AATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGAC
AATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCA
ATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCAT
TTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGCAAAACAGCATTCCAGGTATT
AGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGA
TTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAAT
AACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGA
AATGCATAAGCTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTA
TTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGAT
CTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGG
TATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGTACTGA
CAATAAAAAGATTCTTGTTTTCAAGAACTTGTCATTTGTATAGTTTTTTTATATTGTAGTTGTTCTATTTT
AATCAAATGTTAGCGTGATTTATATTTTTTTTCGCCTCGACATCATCTGCCCAGATGCGAAGTTAAGTGCG
CAGAAAGTAATATCATGCGTCAATCGTATGTGAATGC*TGGTCGCTATACTGCTGTCG*ATTCGATACTAACG
CCGCCATCCAGTTTAAACGAGCTCTTATTCTTCTATAGTGTTCTATTTAGCTACTTTTTATGTTTAACCTT
TTATACGATGGCGGGTAATCTATCCATGATGACGAAAAATTTTTTTTTTTTTTGTTTCCGCAGCACGCAAG
AAATCTCGAAACAATGATGACTCTTAAGCATGAAAAATATCATTTTGCGCTTTAAACTAGAGTGATGTTAC

TCGACTTCCTACAACCTTTAGCCAAAAGCTTCAAAAAACCAAGGAAATACTAGTACCATGGTGAGCAAGGG
CGAGGAGGTCATCAAAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCATGAACGGCCACGAGTTCG
AGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGC
GGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAGTTCATGTACGGCTCCAAGGCGTACGTGAAGCA
CCCCGCCGACATCCCCGATTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACT
TCGAGGACGGCGGTCTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCACGCTGATCTACAAGGTG
AAGATGCGCGGCACCAACTTCCCCCCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTC
CACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACCAGGCCCTGAAGCTGAAGGACG
GCGGCCACTACCTGGTGGAGTTCAAGACCATCTACATGGCCAAGAAGCCCGTGCAACTGCCCGGCTACTAC
TACGTGGACACCAAGCTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCTC
CGAGGGCCGCCACCACCTGTTCCTGGGGCATGGCACCGGCAGCACCGGCAGCGGCAGCTCCGGCACCGCCT
CCTCCGAGGACAACAACATGGCCGTCATCAAAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCATG
AACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCT
GAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAGTTCATGTACGGCTCCA
AGGCGTACGTGAAGCACCCCGCCGACATCCCCGATTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGG
GAGCGCGTGATGAACTTCGAGGACGGCGGTCTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCAC
GCTGATCTACAAGGTGAAGATGCGCGGCACCAACTTCCCCCCCGACGGCCCCGTAATGCAGAAGAAGACCA
TGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACCAGGCC
CTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGACCATCTACATGGCCAAGAAGCCCGTGCA
ACTGCCCGGCTACTACTACGTGGACACCAAGCTGGACATCACCTCCCACAACGAGGACTACACCATCGTGG
AACAGTACGAGCGCTCCGAGGGCCGCCACCACCTGTTCCTGTACGGCATGGACGAGCTGTACAAGCCACTG
CACTCAAAACTCTAACTCGAGgcccgctattaacgctttgtaatgtatagcttttaatgtgtgatcgcctg
acttttgcacgccggccggcggcagcgcgctgcgtccaggcagcgtcacggcgcaggtggcgggctgtgcg
agctggcgcgactggcgcgactgtcgctgtcacgtgacggcgaccacgggtggagaaaatttttggccaac
ggcgcgcagagcagtcctcgaCCGCGGATCTGCCGGTCTCCCTATAGTGAGTCGTATTAATTTCGATAAGC
CAGGTTAACCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGC
TTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG
TAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCC
AGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAA
TCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCT
CCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGC
GTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTG
TGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGG
TAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGT
GCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCT
GCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCG
GTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTT
TCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAG
GATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTT
GGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATA
GTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAAT
GATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGC
GCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGT
AGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTT
TGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAA
AAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTT
ATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTC
AACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATA
CCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGG
ATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTAg
CTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACA
CGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCAT
GAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAG
TGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCC
TTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGC
TTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGG

GCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGGACATATTGTCGTTAGAA
CGCGGCTACAATTAATACATAACCTTATGTATCATACACATACGATTTAGGTGACACTATA

pFA6a-ILSPr-eGFP-*pts1**-ADH1-KANMX6

The sequence of the plasmid encoding EGFP fused to a reversed PTS1 sequence driven by the ILS1 promoter is listed below. The ILS1 promoter is highlighted in blue, the start codon of the GFP is shown in green, the sequence encoding the pts1* tag is highlighted in yellow, the stop codon of the protein is shown in red, and the ADH1 terminator from *A. gossypii* is highlighted in grey. The 20-bp sequences shown in red can be used to amplify the construct prior to integration into the yeast genome. The bold sequence denotes the KAN resistance gene.

GAACGCGGCCGCCAGCTGAAGCTTCGTACGCTGCAGGTCGA**CGGATCCCCGGGTTAATTAA**GGCGCGCCAG
ATCTGTTTAGCTTGCCTCGTCCCCGCCGGGTCACCCGGCCAGC*GACATGGAGGCCCAGAATACCCTCCTTG*
ACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGTGACTGTCGCCCGTACATTTAGCCCATACATCCCCAT
GTATAATCATTTGCATCCATACATTTTGATGGCCGCACGGCGCGAAGCAAAAATTACGGCTCCTCGCTGCA
GACCTGCGAGCAGGGAAACGCTCCCCTCACAGACGCGTTGAATTGTCCCCACGCCGCGCCCCTGTAGAGAA
ATATAAAAGGTTAGGATTTGCCACTGAGGTTCTTCTTTCATATACTTCCTTTTAAAATCTTGCTAGGATAC
*AGTTCTCACATCACATCCGAACATAAACAACC***ATGGGTAAGGAAAAGACTCACGTTTCGAGGCCGCGATTA**
AATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGAC
AATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCA
ATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCAT
TTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGCAAAACAGCATTCCAGGTATT
AGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGA
TTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAAT
AACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGA
AATGCATAAGCTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTA
TTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGAT
CTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGG
TATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGTACTGA
CAATAAAAAGATTCTTGTTTTCAAGAACTTGTCATTTGTATAGTTTTTTTATATTGTAGTTGTTCTATTTT
AATCAAATGTTAGCGTGATTTATATTTTTTTTCGCCTCGACATCATCTGCCCAGATGCGAAGTTAAGTGCG
CAGAAAGTAATATCATGCGTCAATCGTATGTGAATGC*TGGTCGCTATACTGCTGTCG*ATTCGATACTAACG
CCGCCATCCAGTTTAAACGAGCTCTTATTCTTCTATAGTGTTCTATTTAGCTACTTTTTATGTTTAACCTT
TTATACGATGGCGGGTAATCTATCCATGATGACGAAAAATTTTTTTTTTTTTTGTTTCCGCAGCACGCAAG
AAATCTCGAAACAATGATGACTCTTAAGCATGAAAAATATCATTTTGCGCTTTAAACTAGAGTGATGTTAC
TCGACTTCCTACAACCTTTAGCCAAAAGCTTCAAAAAACCAAGGAAATACTAGTACCATGGTGAGCAAGGG
CGAGGAGGTCATCAAAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCATGAACGGCCACGAGTTCG
AGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGC
GGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAGTTCATGTACGGCTCCAAGGCGTACGTGAAGCA
CCCCGCCGACATCCCCGATTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACT
TCGAGGACGGCGGTCTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCACGCTGATCTACAAGGTG
AAGATGCGCGGCACCAACTTCCCCCCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTC

CACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACCAGGCCCTGAAGCTGAAGGACG
GCGGCCACTACCTGGTGGAGTTCAAGACCATCTACATGGCCAAGAAGCCCGTGCAACTGCCCGGCTACTAC
TACGTGGACACCAAGCTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCTC
CGAGGGCCGCCACCACCTGTTCCTGGGGCATGGCACCGGCAGCACCGGCAGCGGCAGCTCCGGCACCGCCT
CCTCCGAGGACAACAACATGGCCGTCATCAAAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCATG
AACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCT
GAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAGTTCATGTACGGCTCCA
AGGCGTACGTGAAGCACCCCGCCGACATCCCCGATTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGG
GAGCGCGTGATGAACTTCGAGGACGGCGGTCTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCAC
GCTGATCTACAAGGTGAAGATGCGCGGCACCAACTTCCCCCCCGACGGCCCCGTAATGCAGAAGAAGACCA
TGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACCAGGCC
CTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGACCATCTACATGGCCAAGAAGCCCGTGCA
ACTGCCCGGCTACTACTACGTGGACACCAAGCTGGACATCACCTCCCACAACGAGGACTACACCATCGTGG
AACAGTACGAGCGCTCCGAGGGCCGCCACCACCTGTTCCTGTACGGCATGGACGAGCTGTACAAGCTCAAA
TCACACCTGCCATAACTCGAGgcccgctattaacgctttgtaatgtatagcttttaatgtgtgatcgcctg
acttttgcacgccggccggcggcagcgcgctgcgtccaggcagcgtcacggcgcaggtggcgggctgtgcg
agctggcgcgactggcgcgactgtcgctgtcacgtgacggcgaccacgggtggagaaaatttttggccaac
ggcgcgcagagcagtcctcgaCCGCGGATCTGCCGGTCTCCCTATAGTGAGTCGTATTAATTTCGATAAGC
CAGGTTAACCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGC
TTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG
TAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCC
AGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAA
TCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCT
CCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGC
GTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTG
TGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGG
TAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGT
GCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCT
GCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCG
GTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTT
TCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAG
GATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTT
GGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATA
GTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAAT
GATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGC
GCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGT
AGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTT
TGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAA
AAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTT
ATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTC
AACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATA
CCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGG
ATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTAg
CTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACA
CGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCAT
GAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAG
TGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCC
TTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGC
TTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGG
GCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGGACATATTGTCGTTAGAA
CGCGGCTACAATTAATACATAACCTTATGTATCATACACATACGATTTAGGTGACACTATA

XII. Supplementary Figure Legends

Supplementary Figure S1
Calculating Autofluorescence

a. A histogram of all uncorrected fluorescence values for strains grown in YEPD.
b. The number of autofluorescence strains (red) is estimated by reflecting the left-hand side of the histogram representing all strains (blue) about a vertical line bisecting the bin containing the largest number of values.
c. The estimated distribution of autofluorescent strains (red) is fitted to a Gaussian curve (black dashed line). The offset of the curve is the estimated mean autofluorescence, the width approximates the standard deviation.

Supplementary Figure S2
Calculating Number of False Positives and False Negatives

A graphical illustration of the origins and magnitudes of false positive and false negative measurements. False positives (red area) correspond to strains that appear to show fluorescence due to GFP, but whose fluorescence is actually due to endogenous cellular factors. False negatives (blue area) correspond to strains whose GFP fluorescence, although greater than cellular autofluorescence, cannot be distinguished with confidence from this signal. Autofluorescent strains (red line) are estimated as described in Supplementary Figure S1. The blue line represents the distribution of fluorescent values for GFP-tagged strains grown in YEPD.

Supplementary Figure S3
Calculating Coefficients of Variation

a. Choice of circular gate radius. Gate radii (based on forward and side scatter) were decreased and the coefficients of variation for GFP fluorescence values were calculated. CVs decrease markedly as the radii are lowered from 65,541 to 5,793, and then either decrease slightly or are relatively flat as the radii are reduced further. CV values reported in this paper are calculated using a radius of 4,096 (dotted red line). The strains shown here were grown in YEPD using single-tube 5 ml cultures.
b. A histogram showing the number of data points used to calculate gated CV values for strains grown in YEPD. Un-gated files generally contained between 50,000 and 75,000 individual data points.
c. As in (b), but for cells grown in SD.
d. The reproducibility of CV calculations is illustrated by plotting values obtained from repeat measurements (R^2=0.85).
e. As in (d), but for cells grown in SD (R^2=0.81).

Supplementary Figure S4
Histograms of Normalized Fluorescence Intensities for Two-Color Studies

Frequency distributions of normalized GFP intensities (blue) are overlaid with frequency distributions of normalized tdTomato intensities (red) for four diploid strains expressing GFP- and tdTomato-tagged proteins. The distributions arising from each fluorophore in

the diploid strains are very similar. The tagged proteins are identified in the inset of each figure.

Supplementary Figure S5
Reproducibility of Measurements

The reproducibility of fluorescence measurements obtained ~14 minutes apart is illustrated. Two groups of 96 proteins (in red and blue) were grown, harvested and then measured by flow cytometry. After completing the measurements on all 96 strains (~14 minutes), the strains were immediately re-measured. The correlation coefficients (R^2) for the first and second sets of measurements are 0.999 and 1.000, respectively.

Supplementary Figure S6
Instrument-Independent Measurement of Fluorescence and CVs

a. Fluorescence intensities are independent of the instrument used to make the measurements. Fluorescence measurements from seven GFP-tagged proteins (Mck1p, Cnb1p, Met6p, Tef2p, Gpm1p, Rpn7p, Ara1p) were obtained either using cytometry or by calculating the background-corrected integrated intensities from microscopy images. Results obtained by the two techniques agree very well (R^2=0.997).

b. CV measurements obtained using cytometry or microscopy are proportional to each other (R^2=0.87). CV values were calculated from cytometry measurements after excluding cells in the bottom and top 5th percentile of fluorescence values (a circular gate was not used). CV values were calculated from microscopy data using background-corrected integrated fluorescence intensities of at least 120 cells/strain. The seven strains described in (**a**) were also used for these calculations.

Supplementary Figure S7
Proportionality of GFP Abundance Measured by Cytometry and Western Blotting

a. Protein abundance measured for 146 GFP-tagged strains by flow cytometry is proportional to protein abundance determined by western blotting studies using TAP-tagged strains (R^2=0.80).

b. Biological replicates of 201 TAP-tagged strains quantified by western blotting illustrate the reproducibility of this approach (R^2=0.77)[6].

c. GFP-fusions are generally not cleaved *in vivo* to yield free GFP. 165 tagged proteins from more than 20 different cellular compartments were assayed for GFP cleavage by western blotting. Most fusion proteins remain intact (blue), with the notable exception of those localized to the vacuole (red circles) and, to a lesser extent, the vacuolar membrane (red squares). Three additional proteins were found to be cleaved, including [in order of increasing molecular weight (black open squares)]: Sed1p (ER/cell wall); Thr1p (ambiguous); Rnr1p (cytoplasm). The horizontal dotted line shows the expected molecular weight of free GFP (27,756 Da, including the linker).

Supplementary Figure S8
Proper Degradation of GFP-tagged Targets of Ubiquitin-mediated Proteolysis

GFP-tagged Sic1p and Clb2p are degraded without the concomitant release of free GFP. Strains were grown to logarithmic phase (U) in YEPD at 30°C. Alpha factor was added to a concentration of 3 μM and cells were arrested for 2 hours at 30°C (0'). Cells were removed from the media, washed twice and suspended in fresh YEPD. Cells were grown for an additional 2 hours and aliquots were removed at the times indicated. Lysis and western blotting were performed as described by Ghaemmaghami *et al.* [6] using a commercial GFP antibody and an antibody against PGK as a loading control. A Li-Cor Odyssey® instrument was used to quantify both fluorescent 2° antibodies. The single and double asterisks denotes the expected molecular weight of cleaved GFP (27,756 Da) and Pgk1p (44,738 Da), respectively.

Supplementary Figure S9
Independent confirmation of instances where steady state protein and mRNA levels disagree.

Protein (blue) and mRNA (red) abundance changes are corroborated by western blotting and qPCR, respectively. Interestingly, some proteins participate in common biological functions such as ergosterol biosynthesis and iron transport[17].

Supplementary Figure S10
Cell-Cycle Regulated Proteins Do Not Have Large DM Scores

a. The CDC score proposed by Spellman *et al.*[7] is a good predictor of protein variation for abundant proteins. Hhf2-GFP (top panel) has a high CDC score and has a bimodal distribution. Cwp2-GFP (middle panel) has a moderate CDC score and a shoulder can be seen in the GFP histogram. Ssb2-GFP (bottom panel) is not cell-cycle regulated and shows a single peak in the histogram.
b. After gating, there is little correlation between CDC score and DM value. Gates were based on a radius of 4,096 centered about the FSC and SSC median values (see Section IV).

Supplementary Figure S11
Organelle localization and protein variation

a. Targeting GFP to peroxisomes results in increased protein noise. Two strains were constructed expressing either GFP containing an intact peroxisome-targeting (GFP-PTS1) or GFP fused to a reversed sequence (GFP- pts1*). Strains where GFP is localized to peroxisomes exhibit higher variation than strains were GFP exhibits diffuse cytoplasmic staining.
b. Variation due to sub-cellular localization is sensitive to correct organelle partitioning. Proteins localized to peroxisomes (GFP-PTS1, Pex11p-GFP, Pex25p-GFP, and Mdh2p-GFP) show an increased level of noise when a mutation (Δinp1) is introduced that causes peroxisomes to partition unevenly. This increase in noise is not seen for proteins localized to other cellular compartments (e.g., GFP-pts1*, Sec28p-GFP,

Mrp8p-GFP and Tpo4p-GFP) or for the parental strain (BY4741) upon introduction of the same mutation.

Supplementary Figure S12
Relationship Between P-values Calculated Using the Kolmogorov-Smirnov and Wilcoxon Rank-Sum Tests
P-value calculations using the Kolmogorov-Smirnov and Wilcoxon rank-sum tests generally give comparable results. However, for larger P-values the Wilcoxon test gives increasingly smaller P-values than the KS-test.

Supplementary Figure S13
Relationship Between Gene Proximity and Variation
Two metrics were used to determine whether gene proximity is related to levels of variation in proteins. The distributions obtained using neighboring genes (blue) are not statistically different from the distributions obtained using random gene pairs (red). Distributions that were not normalized were calculated using Equation (13). Normalized distributions were calculated using Equation (14) (see Supplementary Text).

Supplementary Figure S14
Clustering of Protein Groups to Identify Minimally Redundant Subsets
Protein groups identified using GO-terms, transcription factor binding data or transcription modules were tested in a pairwise manner to determine the number of redundant terms within each group. The groups were then clustered according to their degree of overlap, and representative examples were identified that have the most significant statistical correlation with low or high protein variation.

Supplementary Figure S15
Relationship Between Variation in mRNA Expression Levels and DM
mRNA expression ratios were obtained across many different experiments for each gene encoding a protein for which we had a CV value. The standard deviation in expression ratios was then calculated, and proteins were grouped according to these standard deviations. The median DM for the set of proteins in each bin was then calculated. As shown in the figure, genes with dramatic transcript level alterations from one condition to another are associated with large DM values.

XIII. List of Supplementary Tables

Supplementary Table S1
Abundance and Variation Measurements for Strains Grown in YEPD and SD

Supplementary Table S2
Changes in Protein and mRNA Levels for Strains Grown in YEPD and SD

Supplementary Table S3
Statistics for Calculating the Number of False Positive and False Negative Strains

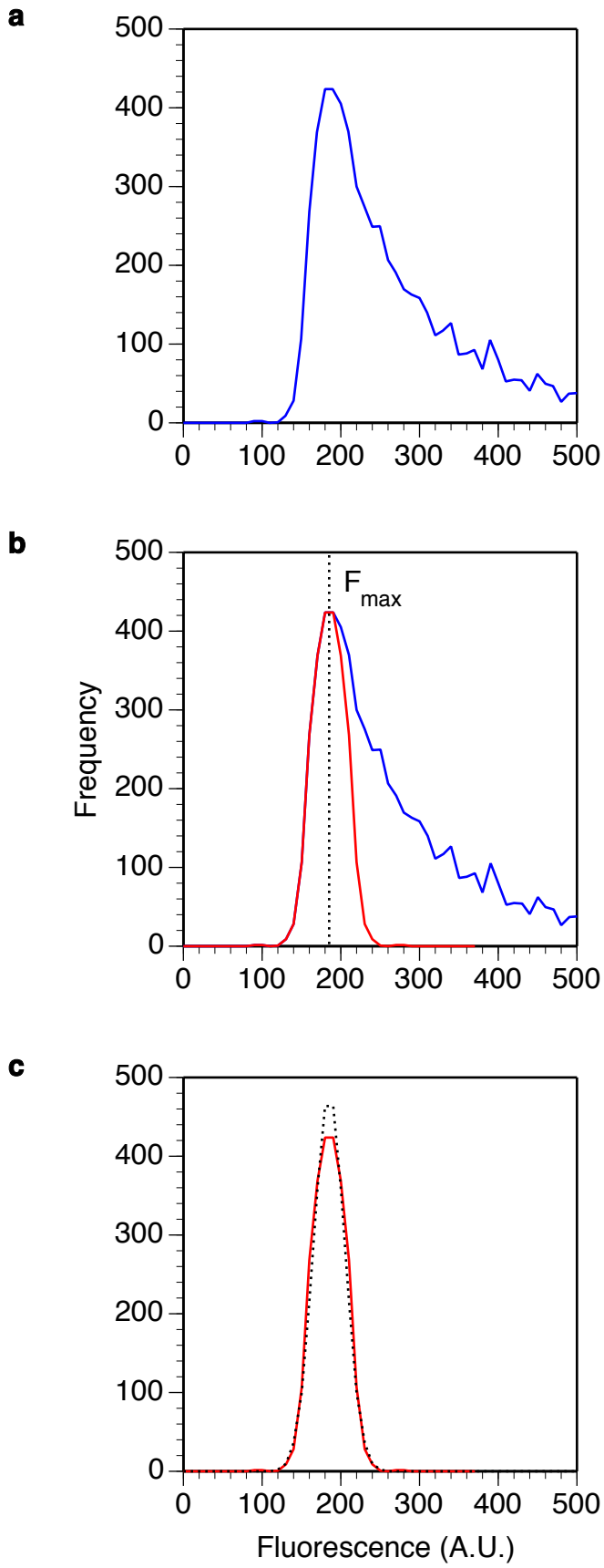
Supplementary Table S4
Primers Used for Tagging and Deletion of ORFs

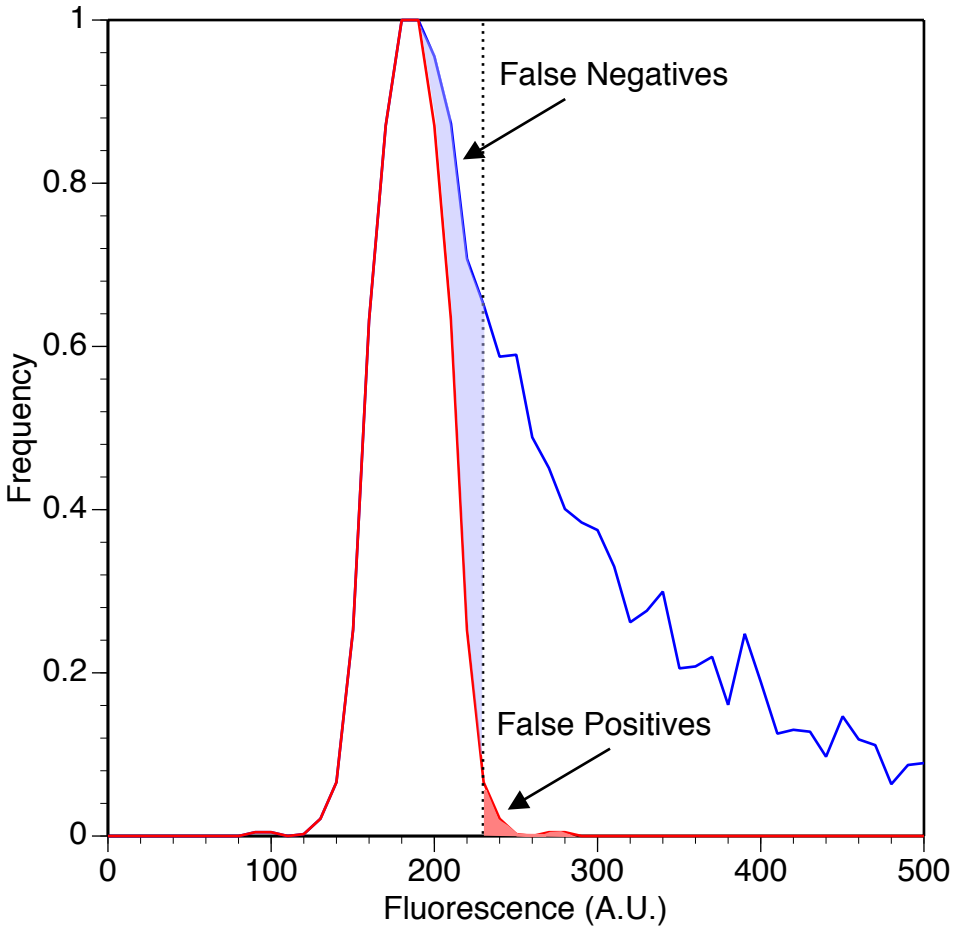
Supplementary Table S5
Organization of GO-Term, Transcription Factor and Transcription Module Correlations Associated with Low or High Variation

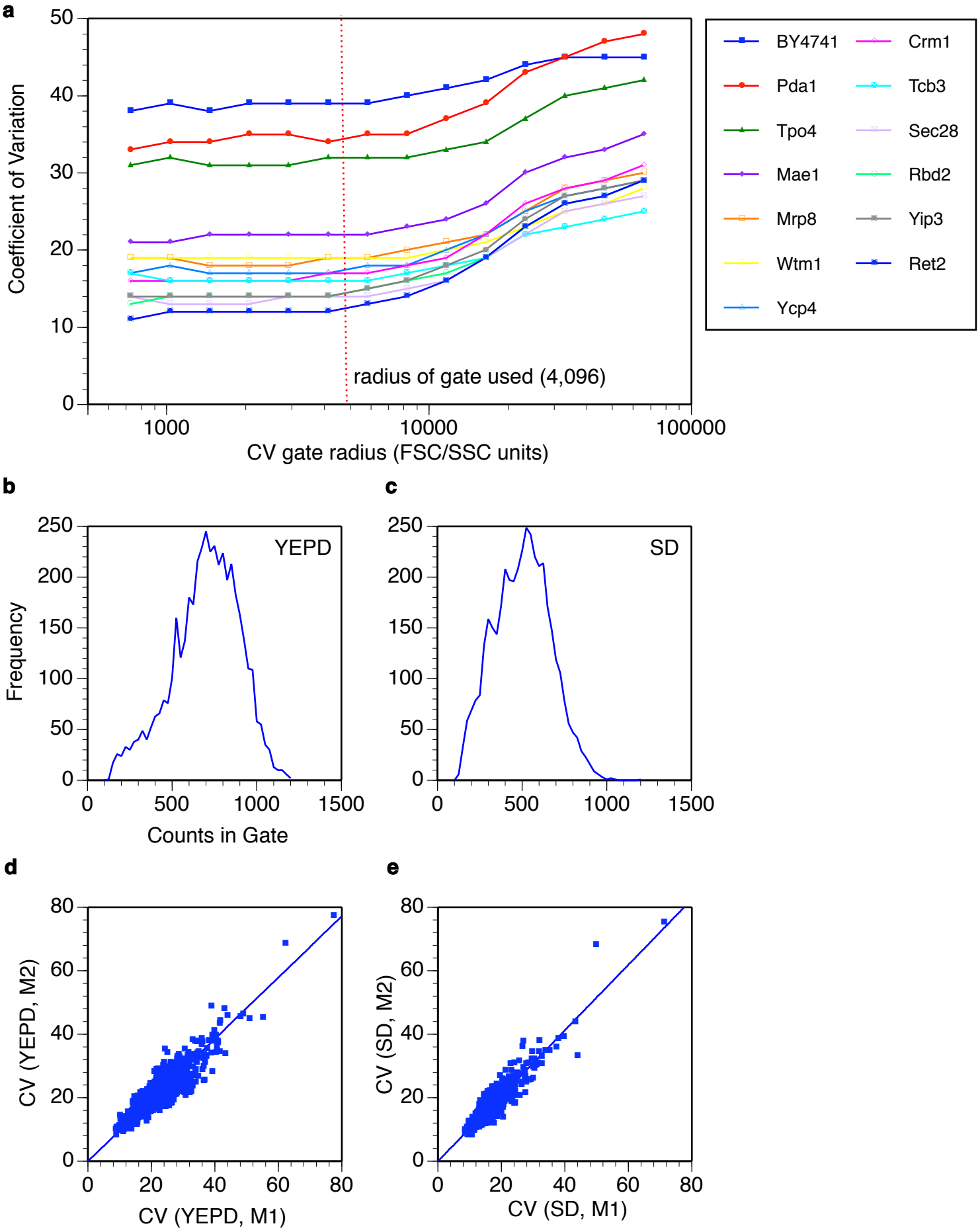
Supplementary Table S6
References for Data Used for P-value Calculations

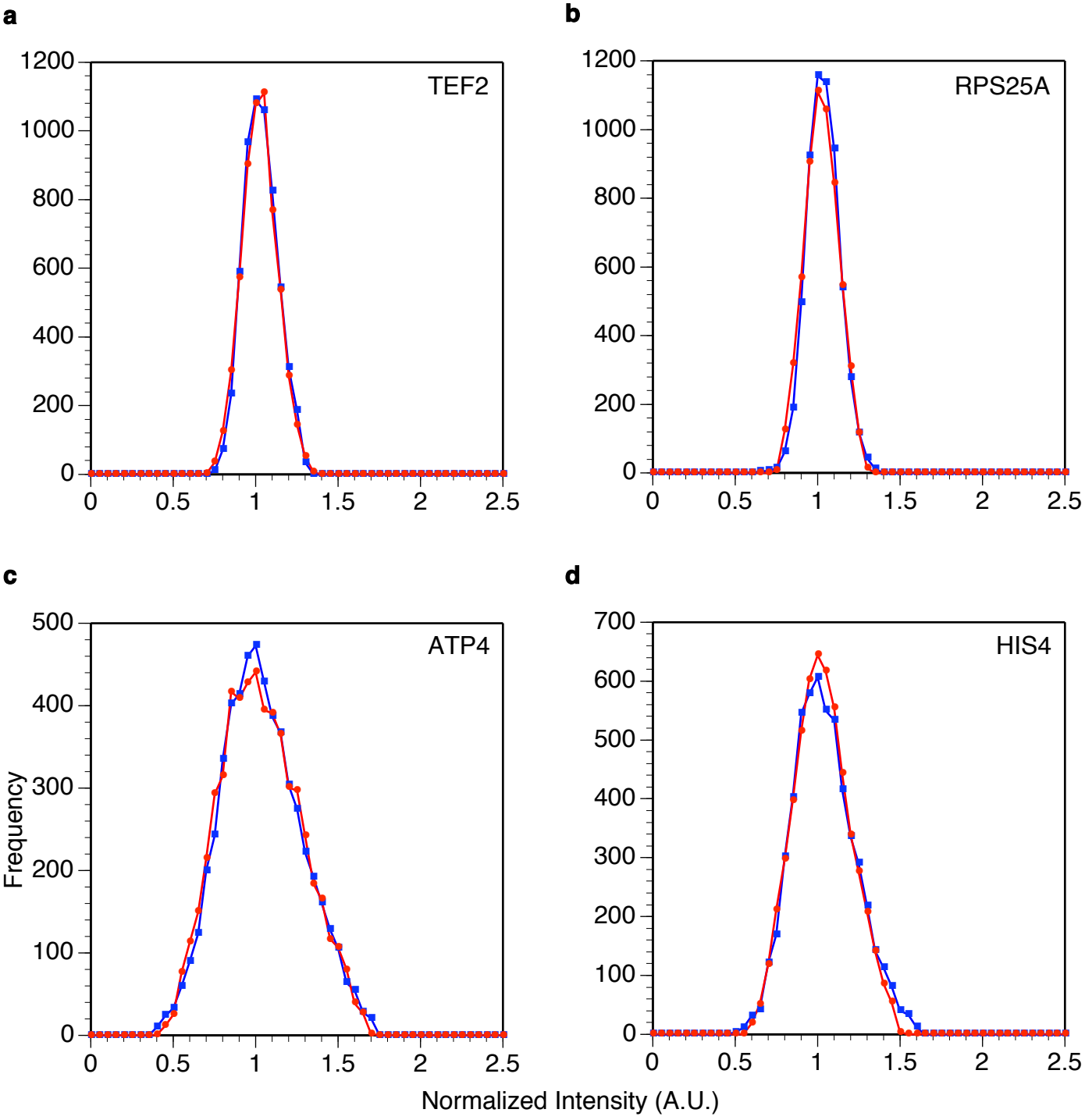
XIV. References

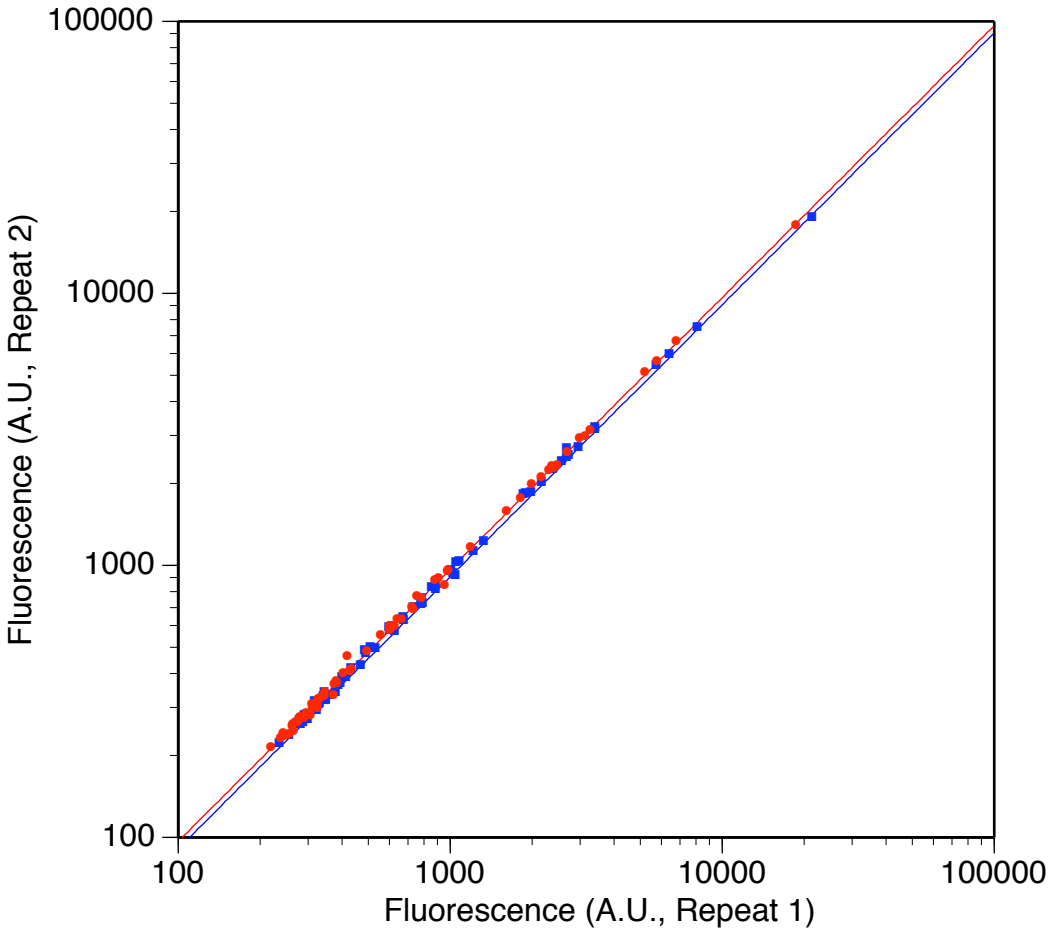
1. Huh, W. K. et al. Global analysis of protein localization in budding yeast. *Nature* **425**, 686-91 (2003).
2. Billinton, N. & Knight, A. W. Seeing the wood through the trees: a review of techniques for distinguishing green fluorescent protein from endogenous autofluorescence. *Anal. Biochem.* **291**, 175-97 (2001).
3. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183-6 (2002).
4. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811-4 (2004).
5. Colman-Lerner, A. et al. Regulated cell-to-cell variation in a cell-fate decision system. *Nature* **437**, 699-706 (2005).
6. Ghaemmaghami, S. et al. Global analysis of protein expression in yeast. *Nature* **425**, 737-41 (2003).
7. Spellman, P. T. et al. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273-97 (1998).
8. Huisinga, K. L. & Pugh, B. F. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in *Saccharomyces cerevisiae*. *Mol. Cell* **13**, 573-85 (2004).
9. Fagarasanu, M., Fagarasanu, A., Tam, Y. Y., Aitchison, J. D. & Rachubinski, R. A. Inp1p is a peroxisomal membrane protein required for peroxisome inheritance in *Saccharomyces cerevisiae*. *J. Cell. Biol.* **169**, 765-75 (2005).
10. Paulsson, J. Summing up the noise in gene networks. *Nature* **427**, 415-8 (2004).
11. Holstege, F. C. et al. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717-28 (1998).
12. Wang, Y. et al. Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci. USA* **99**, 5860-5 (2002).
13. Alon, U. et al. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci. USA* **96**, 6745-50 (1999).
14. Brachmann, C. B. et al. Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**, 115-32 (1998).
15. Gietz, R. D. & Woods, R. A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* **350**, 87-96 (2002).
16. Gould, S. J., Keller, G. A., Hosken, N., Wilkinson, J. & Subramani, S. A conserved tripeptide sorts proteins to peroxisomes. *J. Cell Biol.* **108**, 1657-64 (1989).
17. Felice, M. R. et al. Post-transcriptional regulation of the yeast high affinity iron transport system. *J. Biol. Chem.* **280**, 22181-90 (2005).


a
500
400
300
200
100
0
0 100 200 300 400 500
b
500
400
300
200
100
0
Frequency
F max
0 100 200 300 400 500
c
500
400
300
200
100
0
0 100 200 300 400 500
Fluorescence (A.U.)

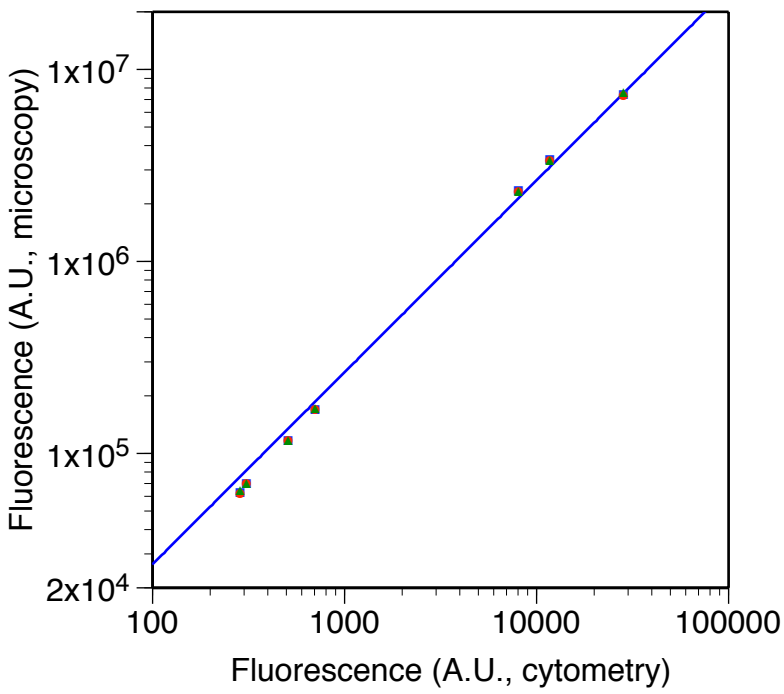

False Negatives
False Positives
Frequency
Fluorescence (A.U.)
0
0.2
0.4
0.6
0.8
1
0
100
200
300
400
500


a
Coefficient of Variation
CV gate radius (FSC/SSC units)
radius of gate used (4,096)
BY4741
Pda1
Tpo4
Mae1
Mrp8
Wtm1
Ycp4
Crm1
Tcb3
Sec28
Rbd2
Yip3
Ret2
b
YEPD
Frequency
Counts in Gate
c
SD
d
CV (YEPD, M2)
CV (YEPD, M1)
e
CV (SD, M2)
CV (SD, M1)

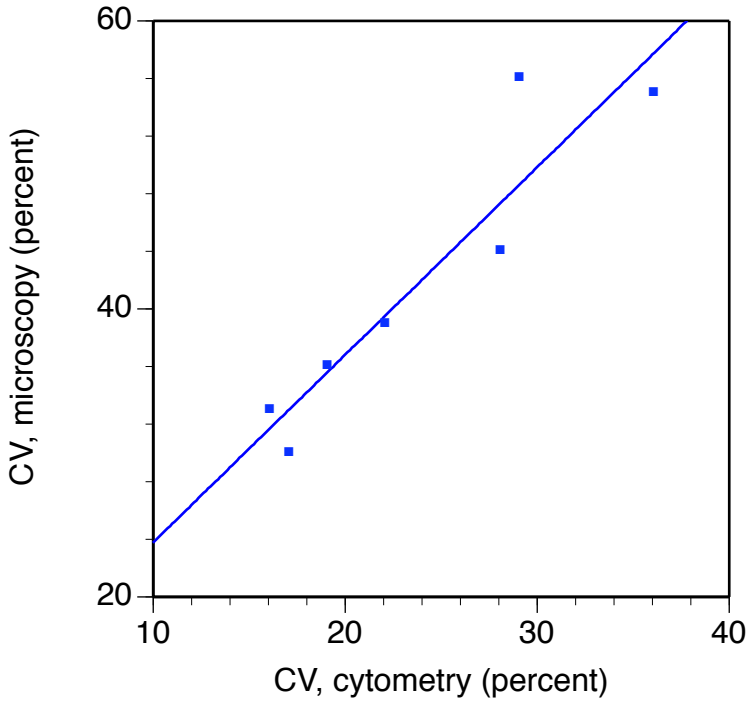

a
TEF2
b
RPS25A
c
ATP4
d
HIS4
Frequency
Normalized Intensity (A.U.)


100000
10000
1000
100
Fluorescence (A.U., Repeat 2)
100
1000
10000
100000
Fluorescence (A.U., Repeat 1)

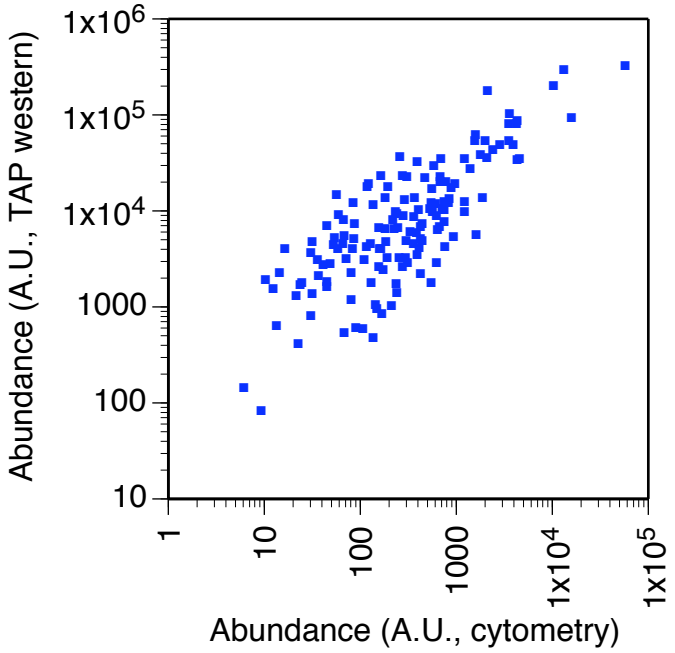
a


Fluorescence (A.U., microscopy)
1x10^7
1x10^6
1x10^5
2x10^4
100
1000
10000
100000
Fluorescence (A.U., cytometry)

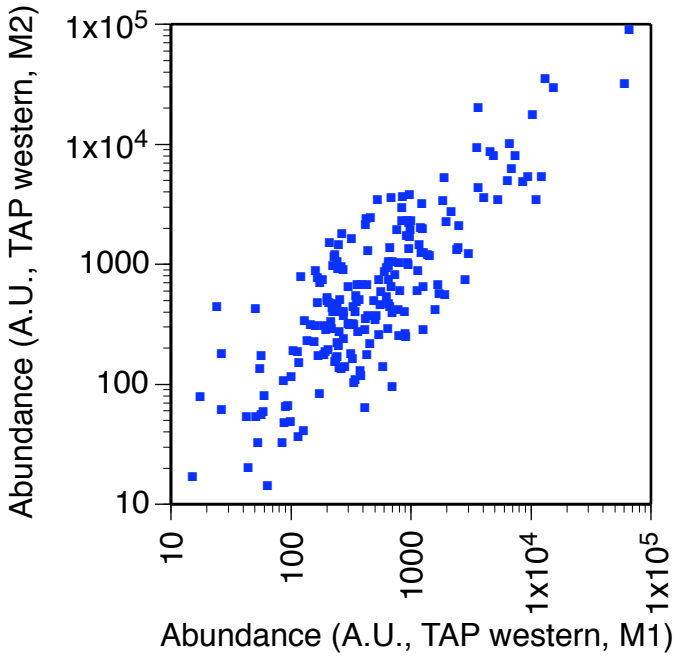
b


CV, microscopy (percent)
60
40
20
10
20
30
40
CV, cytometry (percent)

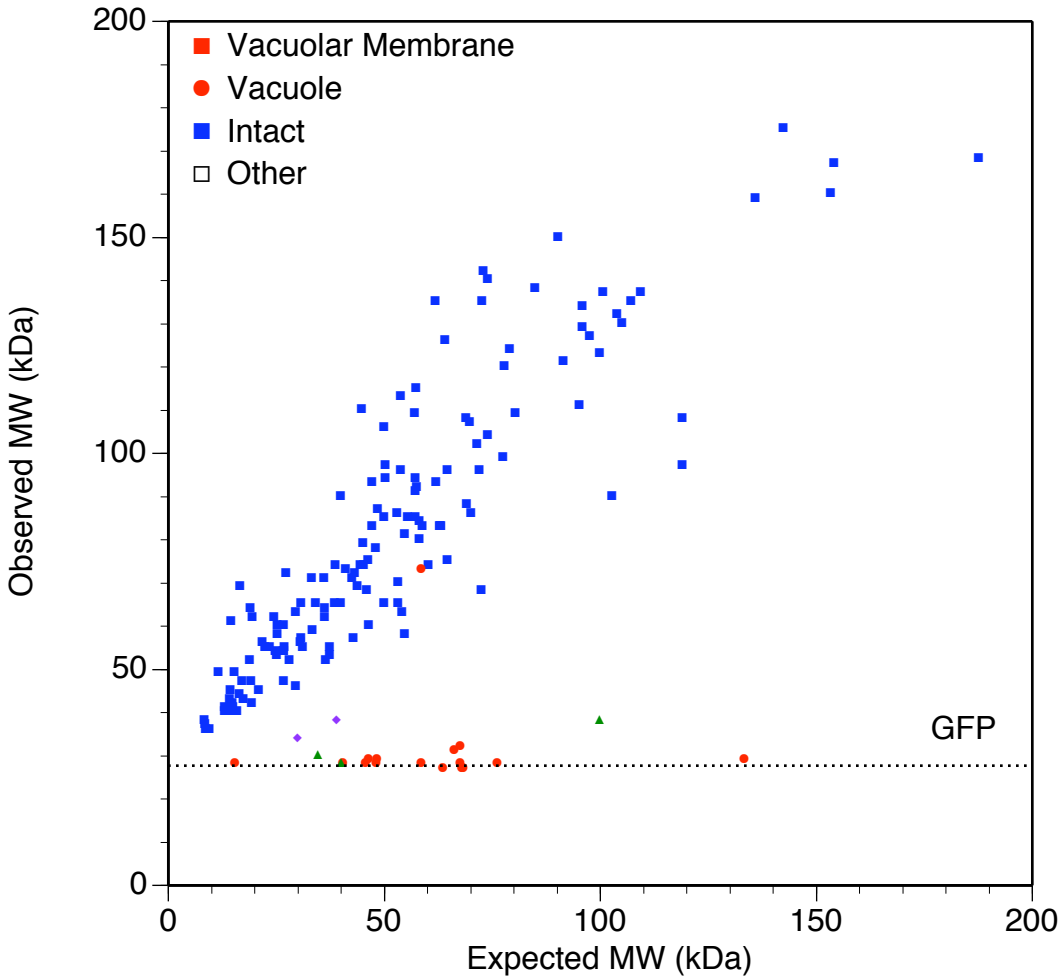
a

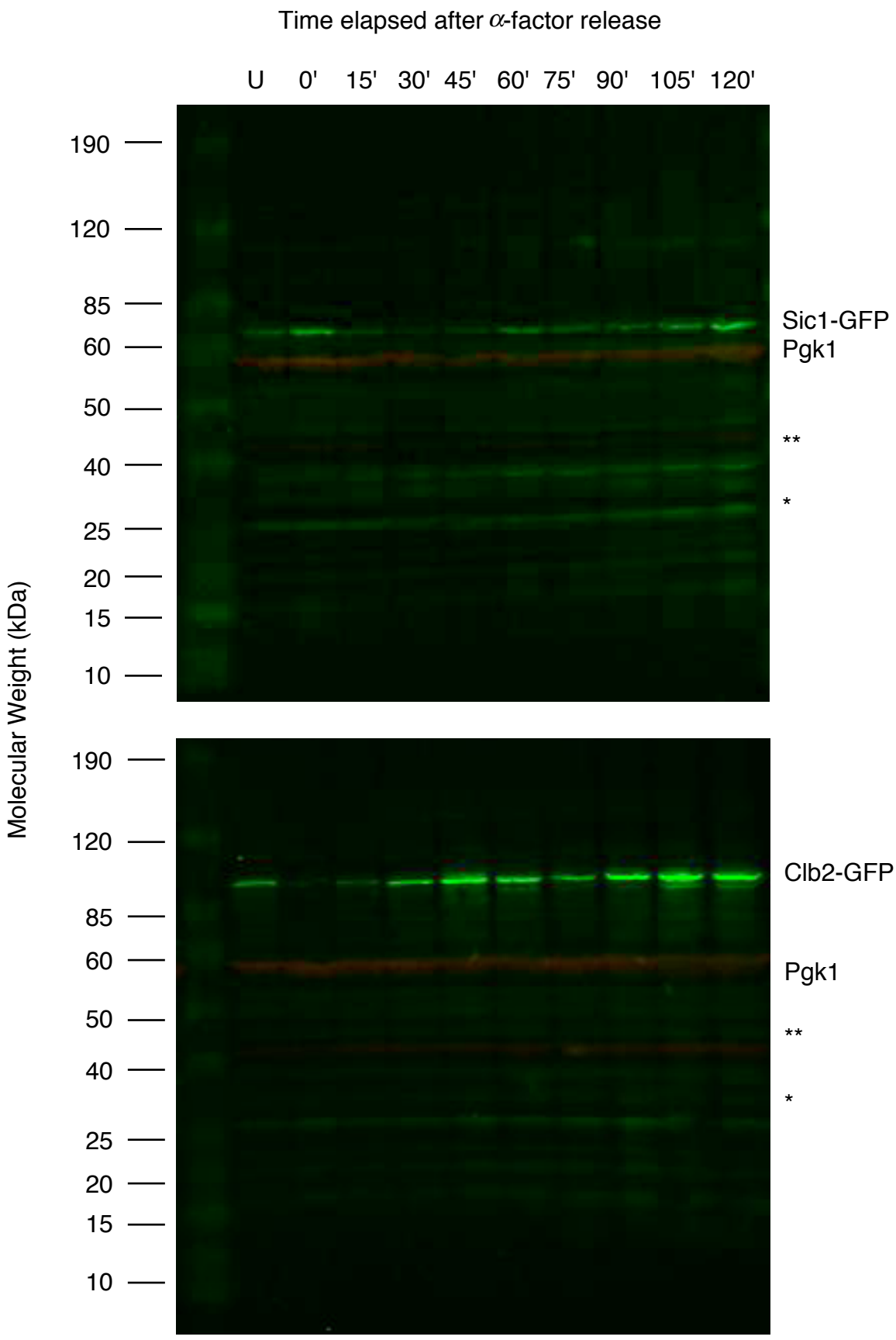

Abundance (A.U., TAP western)
1x10^6
1x10^5
1x10^4
1000
100
10
1
10
100
1000
1x10^4
1x10^5
Abundance (A.U., cytometry)

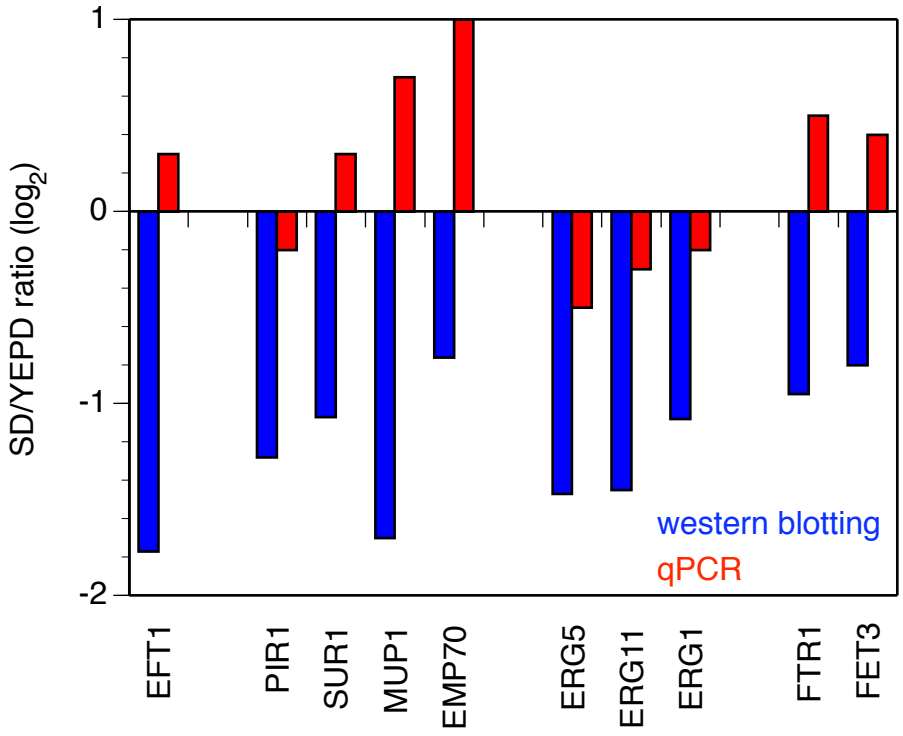
b


Abundance (A.U., TAP western, M2)
1x10^5
1x10^4
1000
100
10
10
100
1000
1x10^4
1x10^5
Abundance (A.U., TAP western, M1)

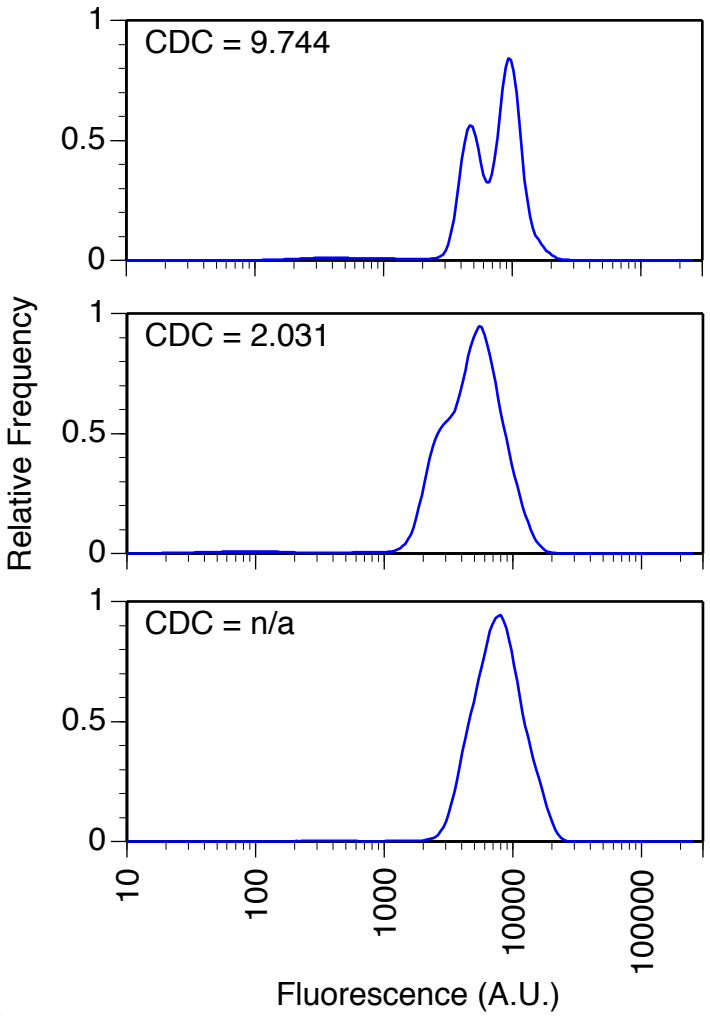
c


Vacuolar Membrane
Vacuole
Intact
Other
GFP
Observed MW (kDa)
200
150
100
50
0
0
50
100
150
200
Expected MW (kDa)

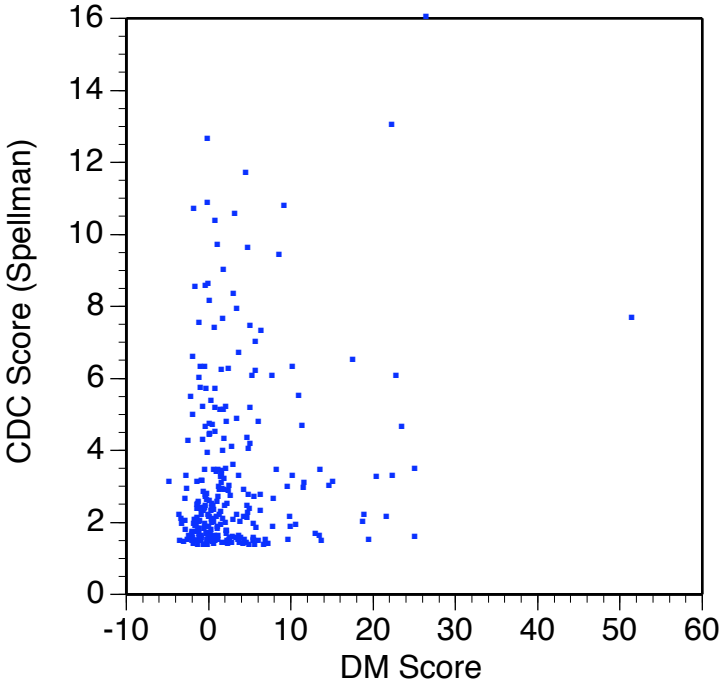

Time elapsed after α-factor release
U
0'
15'
30'
45'
60'
75'
90'
105'
120'
Molecular Weight (kDa)
190
120
85
60
50
40
25
20
15
10
Sic1-GFP
Pgk1
**
*
Clb2-GFP
Pgk1
**
*

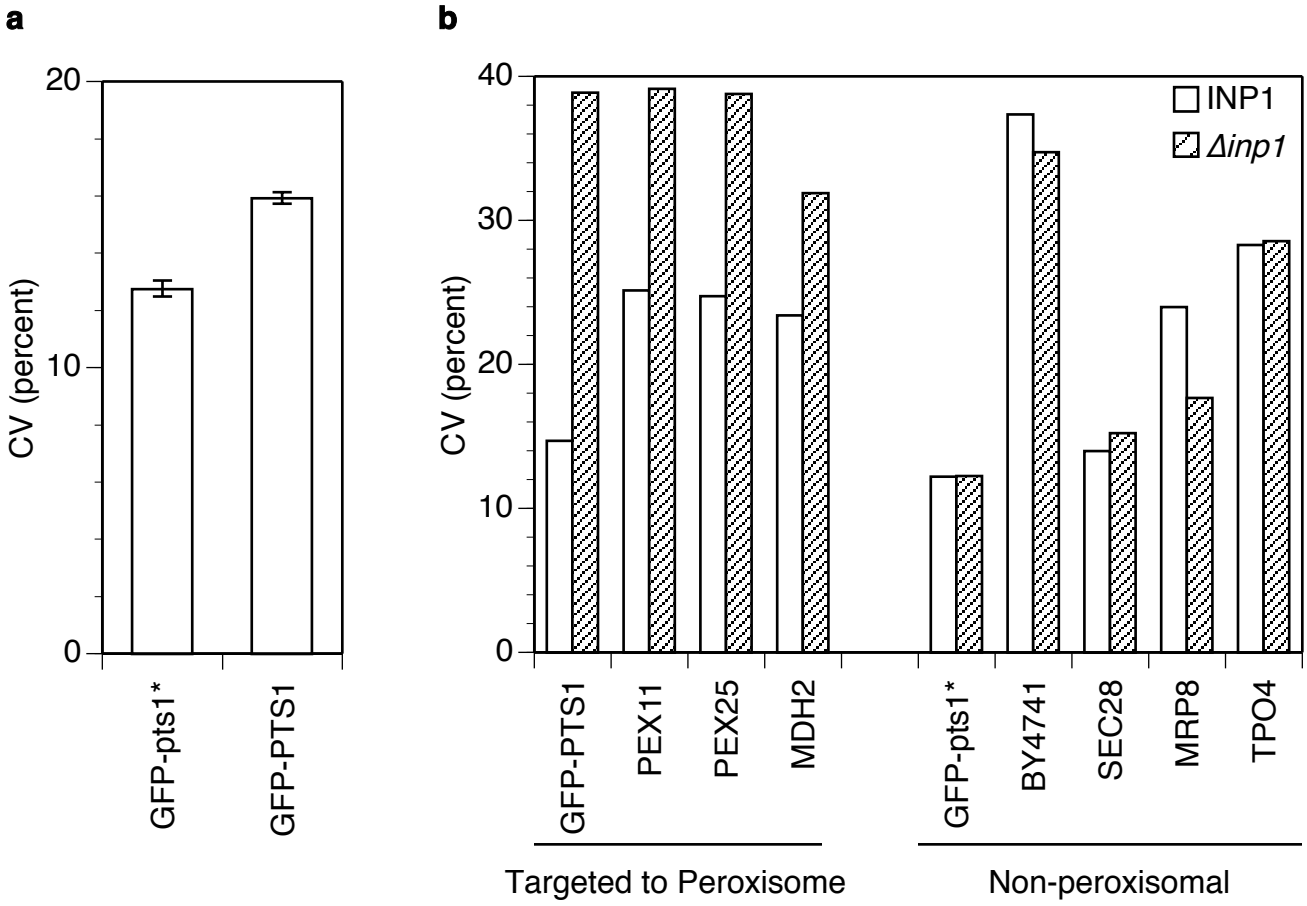

1
0
-1
-2
SD/YEPD ratio (log2)
western blotting
qPCR
EFT1
PIR1
SUR1
MUP1
EMP70
ERG5
ERG11
ERG1
FTR1
FET3

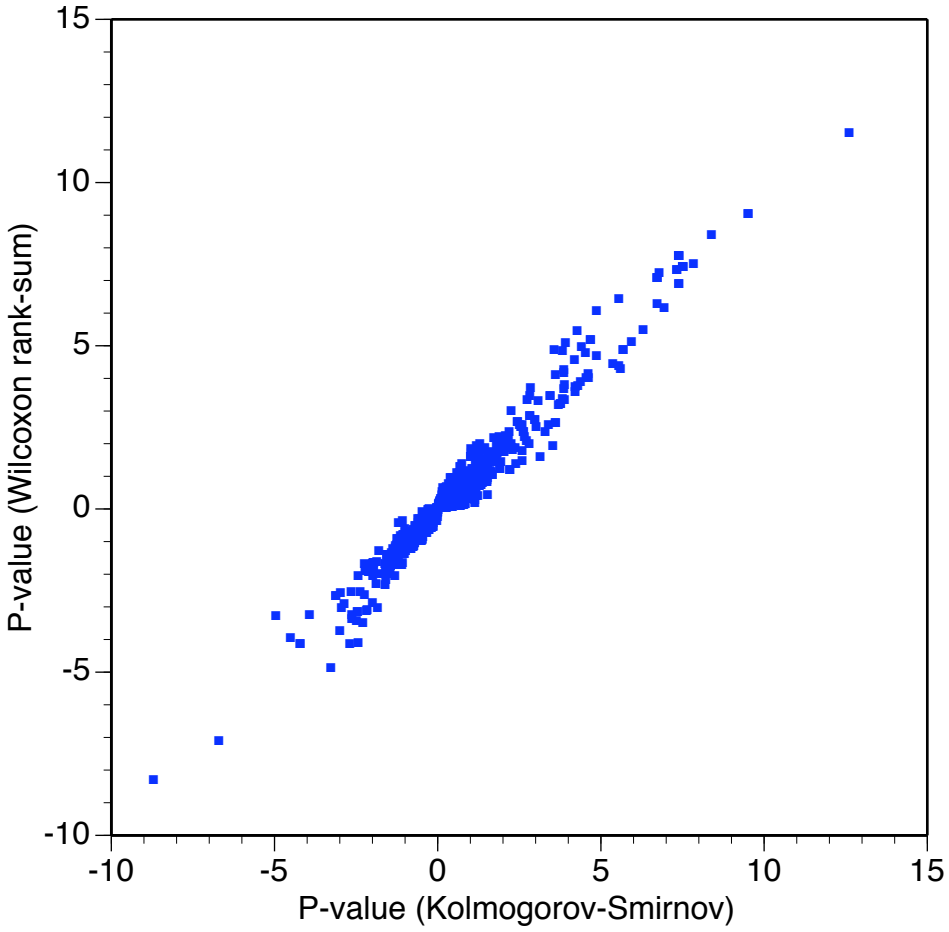
a

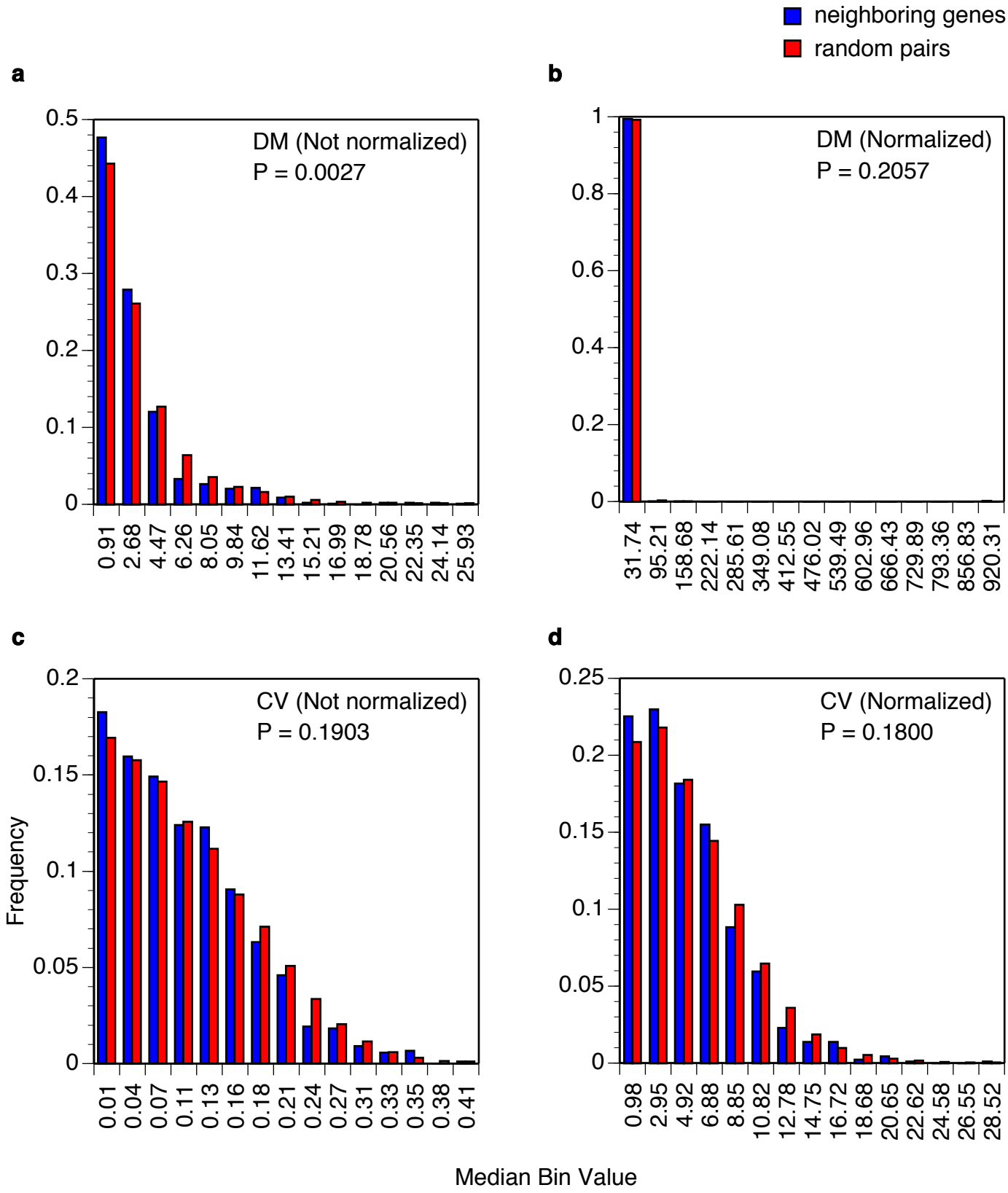

CDC = 9.744
CDC = 2.031
CDC = n/a
Relative Frequency
0
0.5
1
10
100
1000
10000
100000
Fluorescence (A.U.)

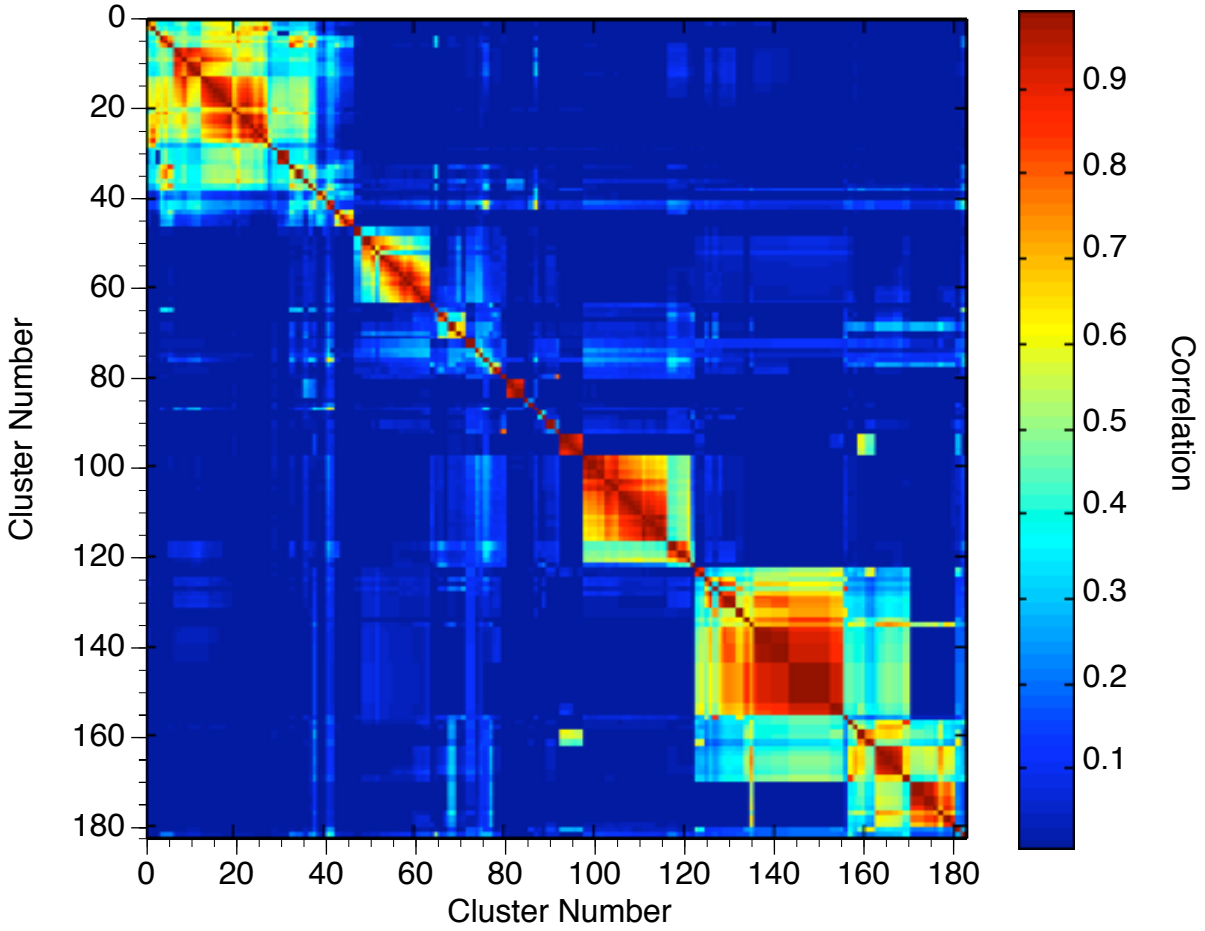
b

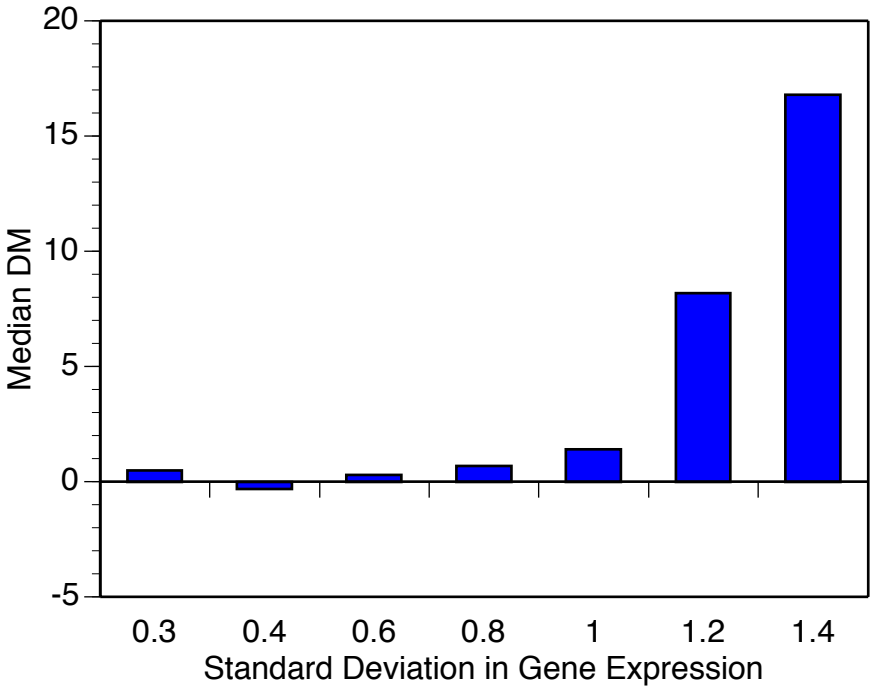

CDC Score (Spellman)
0
2
4
6
8
10
12
14
16
-10
0
10
20
30
40
50
60
DM Score


a
b
CV (percent)
20
10
0
GFP-pts1*
GFP-PTS1
40
30
20
10
0
INP1
Δinp1
GFP-PTS1
PEX11
PEX25
MDH2
GFP-pts1*
BY4741
SEC28
MRP8
TPO4
Targeted to Peroxisome
Non-peroxisomal


15
10
5
0
-5
-10
-10
-5
0
5
10
15
P-value (Wilcoxon rank-sum)
P-value (Kolmogorov-Smirnov)


neighboring genes
random pairs
a
DM (Not normalized)
P = 0.0027
0.5
0.4
0.3
0.2
0.1
0
0.91
2.68
4.47
6.26
8.05
9.84
11.62
13.41
15.21
16.99
18.78
20.56
22.35
24.14
25.93
b
DM (Normalized)
P = 0.2057
1
0.8
0.6
0.4
0.2
0
31.74
95.21
158.68
222.14
285.61
349.08
412.55
476.02
539.49
602.96
666.43
729.89
793.36
856.83
920.31
c
CV (Not normalized)
P = 0.1903
Frequency
0.2
0.15
0.1
0.05
0
0.01
0.04
0.07
0.11
0.13
0.16
0.18
0.21
0.24
0.27
0.31
0.33
0.35
0.38
0.41
d
CV (Normalized)
P = 0.1800
0.25
0.2
0.15
0.1
0.05
0
0.98
2.95
4.92
6.88
8.85
10.82
12.78
14.75
16.72
18.68
20.65
22.62
24.58
26.55
28.52
Median Bin Value


Cluster Number
0
20
40
60
80
100
120
140
160
180
Cluster Number
0
20
40
60
80
100
120
140
160
180
Correlation
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1


20
15
10
5
0
-5
Median DM
0.3
0.4
0.6
0.8
1
1.2
1.4
Standard Deviation in Gene Expression